Cultural transmission of move choice in chess

Egor Lappo[1,†], Noah A. Rosenberg[1], and Marcus W. Feldman[1]

[1]Department of Biology, Stanford University, Stanford, CA 94305 USA
[†]Email: elappo@stanford.edu

September 7, 2023

Abstract. The study of cultural evolution benefits from detailed analysis of cultural transmission in specific human domains. Chess provides a platform for understanding the transmission of knowledge due to its active community of players, precise behaviors, and long-term records of high-quality data. In this paper, we perform an analysis of chess in the context of cultural evolution, describing multiple cultural factors that affect move choice. We then build a population-level statistical model of move choice in chess, based on the Dirichlet-multinomial likelihood, to analyze cultural transmission over decades of recorded games played by leading players. For moves made in specific positions, we evaluate the relative effects of frequency-dependent bias, success bias, and prestige bias on the dynamics of move frequencies. We observe that negative frequency-dependent bias plays a role in the dynamics of certain moves, and that other moves are compatible with transmission under prestige bias or success bias. These apparent biases may reflect recent changes, namely the introduction of computer chess engines and online tournament broadcasts. Our analysis of chess provides insights into broader questions concerning how social learning biases affect cultural evolution.

Keywords. Chess, cultural evolution, Dirichlet-multinomial, social learning, transmission biases.

1 Introduction

Chess has existed in its current form for hundreds of years; it is beloved as an established sport, a hobby, and also as a source of inspiration for scientists across disciplines. Since the 1950s, playing chess well has served as a goal in the development of artificial intelligence, as a task that a "thinking agent" would be able to accomplish (Shannon 1950). This goal was realized in the victory of a chess algorithm over a top human player (Deep Blue vs. Garry Kasparov in 1997). In physics and signal processing, researchers study time series in databases of chess games to extract information regarding long-term correlations, dynamics of position evaluation, invention of new openings, and other game features (see e.g. Schaigorodsky, Perotti, and Billoni 2016; Blasius and Tönjes 2009; Ribeiro et al. 2013; Perotti et al. 2013). Statisticians have been interested in chess as a case study in the development of human performance measurement (Regan, Macieja, and Haworth 2011; Di Fatta, Haworth, and Regan 2009) and modeling of human choice (Regan, Biswas, and Zhou 2014; Regan, Biswas, and Zhou 2014).

As a *cultural* dataset, a compendium of chess games has great potential to help cultural evolution researchers understand patterns of cultural transmission and social learning. A large body of well-annotated chess games is available online, and, unlike linguistic or textual data, for example, these data contain a precise record of players' behavior. As chess positions and moves are discrete, they can be recorded with complete information. Yet the space of potential game sequences is extremely large, so that there can be great variation in move choices. In addition, the large amount of canonical literature on chess allows for thorough qualitative interpretation of patterns in move choice.

Focusing on the game of Go, a game that also features discrete moves and complete information, Beheim, Thigpen, and McElreath (2014) analyzed the choice of the first move by Go players in a dataset of ∼31,000 games. They concluded that the choice of the first move is driven by a mix of social and individual factors, and the strength of these influences depends on the player's age. Many issues concerning cultural transmission in board games remain to be studied. For example, what are the mechanisms behind social learning: are players choosing to use "successful" moves or, instead, moves played by successful players? What defines success of a move? Answering these questions contributes to understanding both general processes of the spread of innovations and mechanisms that govern dynamics of the evolution of cultural traits.

In this paper, we perform a quantitative study of chess in the context of cultural evolution using a database of 3.45 million chess games from 1971 to 2019. In Section 2, we introduce chess vocabulary and several aspects of the game important for our analysis. In Section 3, we describe cultural factors involved in the game and position them within the context of existing literature on cultural transmission. Section 4 describes the dataset used in this study. In Section 5, we motivate and define a statistical model for frequencies of opening strategies in the dataset. Unlike individual-based analysis of a binary choice of the first move in Go by Beheim, Thigpen, and McElreath (2014), our model incorporates counts for all possible moves in a position, taking a population-level approach. In Section 6, we discuss the fit of the model to data for three positions at different depths in the game tree.

2 The game of chess

In this section, we briefly review chess vocabulary, assuming readers have some basic knowledge of the rules of the game (for a concise summary, see Capablanca 1935).

First, a game of chess consists of two players taking turns moving one of their pieces on the board, starting with the player who is assigned the white pieces. We will call these discrete actions *plys*: the first ply is a move by the white player, the second ply is a move by the black player, and so on. The average length of a chess game at a professional level is around 80 plys (see Section 4 below). We will use the word "ply" when describing specific positions, but otherwise we will use the words "move," "strategy," and "response" interchangeably with "ply."

Moves are typically recorded using *algebraic notation* (Hooper and Whyld 1992, p. 389), in which each ply is represented by a letter for a piece — **K** for king, **Q** for queen, **R** for rook, **B** for bishop, **N** for knight, no letter for a pawn — followed by the coordinates of the square on which the piece ends. The coordinates on the board are recorded using letters from **a** to **h** from left to right for the *ranks* (the x-axis coordinates), and numbers from 1 to 8 for the *files* (the y-axis coordinates). For example, the first few moves of the game could be recorded as **1. e4 e5 2. Nf3 Nc6 3. Bc4 Nf6...** Other special symbols are used for captures (**x**), checks (+), and castling (**O-O** or **O-O-O** for king- and queen-side castling, respectively).

The initial stage of the game is called the *opening*. In the opening, players try to achieve a favorable arrangement of the pieces that gives them the most freedom for further actions while keeping their kings safe. Openings are highly standardized, with many having names, e.g. the Sicilian Defense, or the London Opening. Because the number of possible positions is not that large at the beginning of the game, openings are extensively analyzed by players and then memorized for use in tournaments. Example chess positions in the opening are presented in Figure 1.

The collective body of knowledge about how to play chess from various positions is called *chess theory*. For the opening, theory consists of extensive analyses of many positions by human players as well as by computers. One of the manifestations of chess theory is the existence of fixed sequences of moves called "lines," from which deviations are rare. A *mainline* is a sequence of moves that has proven to be the most challenging for both opponents, such that neither of them is able to claim an advantage. A *sideline* is a sequence of moves that deviates from the established optimal sequence.

Each professional chess player has a numerical rating, usually assigned by the national or international federation. FIDE (The International Chess Federation) uses the *Elo rating system* (Elo 1978). The rating is relative, meaning that it is calculated based on a player's past performance, and is intended to represent a measure of the player's ability. The typical rating of a strong intermediate player is $\sim$1500, and a rating of 2500 is required to qualify for a Grandmaster (GM) title. Most elite tournaments involve ratings above 2700.

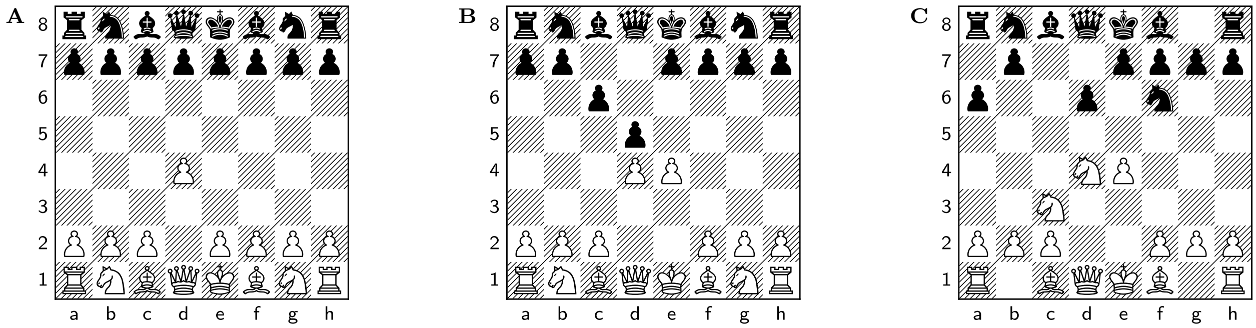

Figure 1: Example chess opening positions. (A) Queen's Pawn opening, **1. d4**. (B) Caro-Kann opening, **1. e4 c6 2. d4 d5**. (C) Najdorf Sicilian opening, **1. e4 c5 2. Nf3 d6 3. d4 cxd4 4. Nxd4 Nf6 5. Nc3 a6**.

3 Culture and chess

Chess is a cultural practice that is actively shaped by the people who participate in it. Individual players enter the practice, altering their performance and behaviors depending on the games they and others have played. Many cultural processes are involved in players' decision-making. To analyze these processes, we will concentrate on decisions made in the opening stage, because the relatively small number of positions allows players to reason about concrete moves and lines in their analyses and preparation. The factors affecting move choice that we discuss below are well-known to the chess community (Euwe and Nunn 1997; Desjarlais 2011; Gobet 2018; Sousa 2002). Our goal here is to place them in the language of cultural evolution.

(a) **Objective strength.** One factor in move choice is the objective strength of the move, which reflects the potential for victory from resulting positions. An evaluation of a move's strength can be made by human analysis or with a chess computer. Many early moves have been extensively analyzed, so the best choice in those positions is well-known to most professional players.

(b) **Social context of the move.** Players are aware of how often a given move has been played in the past. This frequency evaluation can even be automated using websites such as OpeningTree.com. Developed theory often exists for more frequent moves, which can be the default choice for many players. Conversely, rare moves or *novelties* (previously unseen moves) can create problems for opponents who most likely have not prepared a response.

It is important to observe that the frequency with which a move is played is not directly proportional to the objective strength discussed in (a); there are moves that are objectively weak, but only conditional on the opponent finding a *single* good response. If this response is not played by the opponent, then the weak move may give an advantage. In some conditions, e.g. an unprepared opponent or lack of time, such a "weak" move can be highly advantageous. There have also been cases in which a historically frequent move was later "refuted" by deep computer analysis.

Beyond the move frequency, information on the success of strategies in leading to a win can play a role in move choice. In many positions, actually applying information about objective move strength is a complex problem. It is not enough to make a single strong move: a player must then *prove* an advantage by continuing to play further strong moves and executing plans that would lead to victory. The success rate of a move is an indicator of how hard it is to gain a long-term advantage leading to checkmate after choosing it.

The influence of elite players may also be important in move choice. Top players participate in invitational tournaments followed by the wider community. Players, presented with a choice of approximately similar moves, may choose the one that was played by a "superstar" player. This phenomenon is exemplified by strategies named after famous players, such as "Alekhine's Defense" (De Firmian 2008, p. 159) or "Najdorf Sicilian" (De Firmian 2008, p. 246). Leading players can create trends; for example, the Berlin defense was popularized after grandmaster Vladimir Kramnik employed it to win the World Championship in 2000 (De Firmian 2008, p. 43).

(c) **Metastrategy.** Beyond trends in move choice, the "metastrategy" of chess is also evolving. Conceptions of what a game of chess "should" look like have been changing through the years, and so has the repertoire of openings used by professional players (Hooper and Whyld 1992, p. 359). In the 18th century, the swashbuckling Romantic style of chess emphasized winning with "style": declining gambits, or offers of an opponent's piece, could be viewed as ungentlemanly, and Queen's Pawn openings were rarely played (Shenk 2011, Ch. 5). However, by the World Championship of 1927, trends in chess had shifted to long-term positional play (see Shenk 2011, Ch. 8). Queen's Pawn openings were the cutting edge of chess theory, and almost all games at that tournament began with the Queen's Gambit Declined (Chessgames.com 2023a). Following World War I, hypermodern chess emphasized control of the board's center from a distance, and its influence is evident in top-level games of the mid-20th century (Shenk 2011, Ch. 10). Hypermodern players refused to commit their pawns forward, preferring a position where pieces are placed on safe squares from which they could target the opponent's weaknesses. Recently, a style of chess mimicking computer play has emerged, in which players memorize long computer-supported opening lines and play risky pawn advances.

Chess is as much a social phenomenon as it is individual. Some players exhibit personal preferences for certain game features, such as early attacks or long and complicated endgames, and some aspects of play are determined by a player's upbringing. For example, the Soviet school of chess formed around a certain energetic, daring, and yet "level-headed" style (Kotov and Yudovich 1961).

(d) **Psychological aspects.** Finally, psychological aspects and circumstances of the game contribute to move choice (Gobet, Retschitzki, and Voogt 2004). There are lines that are known to lead to a quick draw, and a player might elect to follow one of them, depending on the relevance of the outcome at a particular stage of the tournament. Openings may also be chosen to take opponents out of their comfort zone: in a game against a much weaker opponent, a dynamic and "pushy" line might give a player an advantage. Similarly, a master of attacking play might make mistakes when forced into a long positional game.

The complexities of move choice suggest that chess could serve as a model example for the quantitative study of culture. Players' knowledge is continually altered by their own preparation, the games they play, and other players' actions. In this sense, chess knowledge is "transmitted" over time, in part by players observing and imitating their own past actions and those of other players, or *transmission by random copying* (Bentley et al. 2007). The large historical database of chess games provides an opportunity to study deviations from random copying dynamics known as *transmission biases* or *social learning strategies* (Boyd and Richerson 1985; Kendal et al. 2018; Laland 2004; Henrich and McElreath 2003). In our analysis of the transmission of chess knowledge, we will investigate *success bias* (players paying attention to win rates of different strategies), *prestige bias* (players imitating the world's best grandmasters), and *frequency-dependent bias* (e.g. players choosing rare or unknown strategies).

4 Data

The dataset that serves as the foundation for this project is *Caissabase* – a compendium of ∼5.6 million chess games, available for download at `caissabase.co.uk`. Games in the dataset involve players with Elo rating 2000 or above, and correspond to master-level play, allowing us to focus on the dynamics of high-level chess without the influence of players who are just learning the game.

In filtering the dataset, we have excluded games with errors that did not correspond to a valid sequence of moves as determined by a chess notation parser. We also filtered the dataset to keep only the games that record the result of the game, players' names, and their Elo ratings, and we selected only the games played from 1971 to 2019. This filtering produced a table of 3,448,853 games.

In Figure 2, we highlight the main aspects of the dataset. Figure 2A shows that the number of games per year has been growing steadily since the 1970s, stabilizing at approximately 100,000. In total, there are 77,956 chess players in the dataset, with the number of players per year increasing in recent decades (Figure 2B).

It is widely accepted in the chess community that white has a slight advantage, as the side that starts the game. This view is reflected in Figure 2C, which plots the fractions of outcomes of games in each year.

Finally, Figure 2D shows the average length of games over time; games have become longer since the mid-1980s, which could mean that players are getting better at the game and no longer lose early. To explore the dynamics in the dataset further, we examine the frequencies of individual moves.

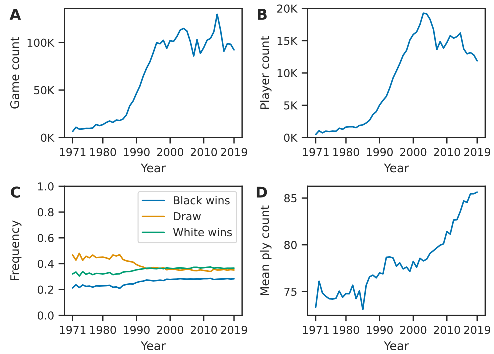

Figure 2: Features of the dataset. (A) Number of games per year. (B) Number of unique players per year. (C) Outcome proportions in each year. (D) Average game length per year, measured in the number of plys (half-moves).

5 Modeling move choice

5.1 Move frequencies

Here, we discuss the dynamics of move frequencies over time for several game positions. Given a position on the board, the player whose turn it is has a choice of which move to play. In positions where their king is in check, players would only have few choices, since they are forced to get out of check. In some other cases, several equally attractive moves could be available, and any of the factors in Section 3 has the potential to affect the choice. Depending on the position, the move frequency trajectories look drastically different, as shown in Figure 3.

Starting Position, ply 1. Figure 3A shows the fractions of games in which different starting moves were played in each year from 1971 to 2019. The frequencies of the moves are mostly constant over time, suggesting that the starting move is a well-understood and well-developed idea.

Sicilian Defense, ply 3. Figure 3B shows move frequencies in response to **1. e4 c5** — the Sicilian Defense. In this position, there is a mainline move — **Nf3** — which an overwhelming majority of players prefer to play, while other moves are rarely played. Move distributions in which one specific move dominates are common, possibly because some sequences of moves are perceived as a single coherent unit.

Queen's Gambit Declined, ply 7. Figure 3C presents an example of a gradual change, which might have happened either due to a change in the metastrategy of play or because of the gradual development of chess theory.

Najdorf Sicilian, ply 11. A game starting with a Sicilian Defense can follow a sequence known as the Najdorf Sicilian. This sequence consists of 10 plys, and the moves at ply 11 that have been played in the resulting position are presented in Figure 3D. Qualitatively, the picture is dramatically different from the early positions considered above. Among the responses to the Najdorf Sicilian, some moves are consistently popular choices (**Be2**, **Be3**, **Bg5**), some became "obsolete" in recent years (**f4**), and some rapidly gained popularity (**h3**).

The qualitative picture of move frequency changes can be summarized as follows. On one hand, very early opening moves do not show large fluctuations in frequencies, most likely because a significant change in frequency necessitates some kind of "innovation," which is impossible to produce at such an early stage. On the other hand, moves beyond the standardized opening frequencies (after the 16th-20th ply) involve positions that do not repeat often enough for humans to memorize and analyze during preparation. This property makes quantitative analysis of specific late-game moves nearly impossible. Somewhere between these two extremes are positions at which chess theory is actively developed and tested. Positions such as the Najdorf Sicilian occur early enough in the game to be reached often, but are advanced enough to provide many continuation possibilities that are approximately equal in terms of objective strength. In such positions, all factors, including engine analysis, move frequency, social context, stylistic trends, and personal preferences could play a role in move choice.

5.2 Population-level modeling of move choice

We develop a statistical model that can help to understand the data described above. A complete model of move choice would involve parameters associated with the whole population, with subgroups of players (e.g. top 50 players), or with each individual. Such a model would be very complex, so our model is restricted to population-level features of dynamics; we analyze frequency-dependent, success, and prestige biases. Features concerning match-level dynamics, personal development, and preferences of individual players are outside of the scope of our analysis, and are present in the form of residual variance, not explained by our population-level treatment.

5.2.1 Unbiased model

First, we consider a null model that generates the simplest dynamics, reflecting unbiased transmission of move choice preferences from one year to the next. Conceptually, the model assumes that each year, players “sample” a move randomly from games that were played in the last year. More precisely, fix an arbitrary chess position and suppose that in each year t, exactly N_t games having this position were played. The data for the model are the counts of k different response moves, denoted by $\boldsymbol{x}_t = (x_t^1, \dots, x_t^k)$. We do not attempt to model appearance of novel strategies, so we will assume that all counts are positive, $x_t^i > 0$. The vector of response strategy counts in the next year, $\boldsymbol{x}_{t+1}$, is multinomially distributed,

$$\boldsymbol{x}_{t+1} \sim \text{Multinomial}(N_{t+1}, \boldsymbol{\theta}_t). \tag{1}$$

The probability vector $\boldsymbol{\theta}_t$ has the Dirichlet distribution with counts in the current year, $\boldsymbol{x}_t$, as Dirichlet allocation parameters,

$$\boldsymbol{\theta}_t \sim \text{Dirichlet}(\boldsymbol{x}_t). \tag{2}$$

The multinomial likelihood depends on a positive integer parameter n and a vector of probabilities $\boldsymbol{\theta}$ that sum to one,

$$f_M(\boldsymbol{y}; n, \boldsymbol{\theta}) = \frac{n!}{y_1! \cdots y_k!} \theta_1^{y_1} \cdots \theta_k^{y_k}; \sum_{i=1}^{k} y_i = n. \tag{3}$$

The Dirichlet likelihood depends on a vector of positive real numbers $\boldsymbol{\alpha}$:

$$f_D(\boldsymbol{\theta}; \boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}. \tag{4}$$

These two likelihoods can be combined into the compound Dirichlet-multinomial likelihood by integrating over $\boldsymbol{\theta}$ (Johnson, Kotz, and Balakrishnan 1997, pp. 80-83),

$$f_{DM}(\boldsymbol{y}; n, \boldsymbol{\alpha}) = \frac{n!\,\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\Gamma\left(n + \sum_{i=1}^k \alpha_i\right)} \prod_{i=1}^{k} \frac{\Gamma(y_i + \alpha_i)}{y_i!\,\Gamma(\alpha_i)}, \tag{5}$$

which will be the likelihood for the model. In other words, under our unbiased model, the counts $\boldsymbol{x}_{t+1}$ of moves in year $t+1$ are distributed with probability density function

$$p(\boldsymbol{x}_{t+1} \mid N_{t+1}, \boldsymbol{x}_t) = f_{DM}(\boldsymbol{x}_{t+1}; N_{t+1}, \boldsymbol{x}_t), \tag{6}$$

so that the counts in the previous year $\boldsymbol{x}_t$ take the roles of the Dirichlet parameters $\boldsymbol{\alpha}$. As a shorthand, we write

$$\boldsymbol{x}_{t+1} \sim \text{Dirichlet-multinomial}(N_{t+1}, \boldsymbol{x}_t). \tag{7}$$

For a vector $\boldsymbol{y}$ having a Dirichlet-multinomial distribution with parameters n and $\boldsymbol{\alpha}$, the expectation is

$$\mathbb{E}\left[\boldsymbol{y}\right] = \frac{n}{\sum_{j=1}^{k} \alpha_j} \boldsymbol{\alpha}. \tag{8}$$

For our model, this formula yields

$$\mathbb{E}\left[\boldsymbol{x}_{t+1}\right] = \frac{N_{t+1}}{\sum_{j=1}^{k} x_t^j} \boldsymbol{x}_t = \frac{N_{t+1}}{N_t} \boldsymbol{x}_t, \tag{9}$$

meaning that no changes are expected to happen in this unbiased model, except possibly for the change in the number of games played. The strategies are "transmitted" from one year to the next proportionally to their current frequencies in the population.

The null model is analogous to a neutral many-allele Wright–Fisher model in population genetics (Ewens 2004). The multinomial distribution arises as a representation of a biological process in Wright–Fisher models, where individuals in the next generation "choose" a parent from the previous generation. In our model of move choice, such sampling is a metaphor that does not correspond exactly to an observed physical process. As we discuss below, working with counts directly via the Dirichlet distribution allows us to account for a potentially higher variance in the strategy counts relative to the multinomial distribution (Corsini and Viroli 2022). Use of the Dirichlet-multinomial likelihood is a common way of dealing with overdispersion in count data in many fields, including ecology (Harrison et al. 2020) and microbiome studies (Wadsworth et al. 2017; Osborne, Peterson, and Vannucci 2022).

It should be noted that chess players pay attention to games further back in the past than just the last year. Our null model is still a reasonable representation of the process for several reasons. First, there is a high degree of autocorrelation in the move count data (Schaigorodsky, Perotti, and Billoni 2016), meaning that it is likely that the most recent data point is representative of counts in the last several years. Second, players tend to look only at *select* famous games of the past, whereas the more recent games can be more easily perceived in their totality.

5.2.2 Fitness and frequency-dependence

A strategy transmitted at a rate greater than expected from the null model can be said to have higher *cultural fitness* (Cavalli-Sforza and Feldman 1981). Conversely, a strategy having a lower transmission rate than expected has lower cultural fitness. Selection on strategies is carried out by players when they decide which move to play based on any of the factors discussed in Section 3. We can account for cultural fitness by associating a *fitness coefficient* f_i to each strategy i. For now, assume that fitness values are constant, $0 < f_i < \infty$. The distribution of moves in the next year can then be described as

$$\boldsymbol{x}_{t+1} \sim \text{Dirichlet-multinomial}(N_{t+1}, f_1 x_t^1, \ldots, f_k x_t^k), \tag{10}$$

with the expression for expected counts in the next year becoming

$$\mathbb{E}\left[x_{t+1}^i\right] = N_{t+1} \frac{f_i x_t^i}{\sum_{j=1}^{k} f_j x_t^j}. \tag{11}$$

As the coefficients f_i are constrained only in that they must be positive, this way of encoding the parameters is useful for inference purposes, especially in the Bayesian framework we employ below. It is straightforward to find reasonable prior distributions on $(0, \infty)$, and absence of "sum to one" constraints makes it easy for an MCMC sampler to efficiently explore the posterior distribution (Gelman, Carlin, et al. 2020, Ch. 12).

However, interpretation of the model is more convenient with a different parameterization: instead of considering values of f_i, we let

$$\bar{f}_t = \frac{1}{N_t} \sum_{j=1}^{k} f_j \, x_t^j \tag{12}$$

be the *mean fitness* at time t, and define

$$f_i' = f_i / \bar{f}_t \tag{13}$$

to be normalized fitness coefficients, such that $\sum_{i=1}^{k} f_i' = 1$. Rewriting eq. (11) as

$$\mathbb{E}\left[x_{t+1}^i\right] = \frac{N_{t+1}}{N_t} \frac{f_i}{\bar{f}_t} x_t^i = \frac{N_{t+1}}{N_t} f_i' x_t^i, \tag{14}$$

we see that $f_i' = 1$ implies no expected change in the frequency of strategy i from time t to $t+1$. Therefore, this choice of parameterization allows us to view f_i' as growth rates, with $f_i' = 1$ corresponding to no selective advantage, i.e. the neutral case. The value of $\bar{f}_t$, in turn, adjusts the variance of the counts in the next year.

To summarize, in our Dirichlet-multinomial model, the f_i's measure two phenomena at once; their *relative* values represent selection, while the *mean* value of the f_i's measures overdispersion with respect to the multinomial model. Mathematically, the expectation of a Dirichlet($\boldsymbol{\alpha}$)-distributed random variable is invariant with respect to multiplying $\boldsymbol{\alpha}$ by a positive constant, but its variance is determined by the magnitudes of the parameters. Although the f_i are convenient to use in inference, we will interpret the results in terms of a parameterization that involves f_i' and $\bar{f}_t$ (eqs. (12) and (13)).

We now allow f_i to depend on the frequency of the strategy, such that

$$\boldsymbol{x}_{t+1} \sim \text{Dirichlet-multinomial}(N_{t+1}, f_1(x_t^1/N_t)\, x_t^1, \ldots, f_k(x_t^k/N_t)\, x_t^k). \tag{15}$$

In this way, we are able to incorporate *frequency-dependent selection* phenomena, which have previously been shown to be present in models of cultural data (e.g. Newberry and Plotkin 2022). Hence, we will refer to f_i as *frequency-dependent fitness functions*. The expression for the mean fitness now becomes

$$\bar{f}_t = \sum_{j=1}^{k} f_j(p_t^j)\, p_t^j, \tag{16}$$

where $p_t^j = x_t^j / N_t$, and k is the number of distinct moves played from a position.

We choose a piecewise-constant form for the functions f_i, as this form introduces minimal assumptions about their shape while keeping the number of parameters low. That is, for $i = 1, \ldots, k$, we have

$$f_i(x) = \begin{cases} c_1^i & \text{if } x \in [0, b_1^i), \\ c_j^i & \text{if } x \in [b_{j-1}^i, b_j^i), \\ c_\ell^i & \text{if } x \in [b_{\ell-1}^i, 1], \end{cases} \tag{17}$$

where c_j^i are values of f_i and b_j^i are breakpoints that determine the boundaries of constant segments. For ℓ segments, $\ell - 1$ breakpoints $b_j^i \in (0, 1)$ must be specified. We choose quartiles of move frequencies as the values for b_j^i, so that each function f_i has three breakpoints and $\ell = 4$ constant segments. This choice does not uniformly cover the domain of f_i, but allows for the same amount of data to be used in estimating each segment.

5.2.3 Full model

We complete our model by accounting for additional features that could affect move choice dynamics. In the final model, the vector of strategy counts in the year $t+1$ again has the Dirichlet-multinomial distribution with parameters N_{t+1} and $\boldsymbol{\alpha}$:

$$\boldsymbol{x}_{t+1} \sim \text{Dirichlet-multinomial}(N_{t+1}, \boldsymbol{\alpha}). \tag{18}$$

However, vector $\boldsymbol{\alpha}$ is now defined as

$$\alpha_i = \exp(\boldsymbol{\beta}_i \cdot \boldsymbol{y}_t^i)\, f_i(x_t^i/N_t)\, x_t^i. \tag{19}$$

Here, x_t^i is the count of games with the ith strategy in year t, f_i is a piecewise constant function of the strategy frequency described in Section 5.2.2 above, and $\boldsymbol{\beta}_i$ is a vector of constant coefficients.

Additional features beyond just the move count or frequency are denoted $\boldsymbol{y}_t^i$ in eq. (19). There are three of these features:

1. The average outcome of the strategy in the whole population for games in year t, with a win for the side making the move encoded as 1, a win for the opposing side encoded as -1, and a draw encoded as 0. We denote the corresponding coefficient by $\beta_{\text{win},i}$.

2. The average outcome of the strategy among the top 50 players in the dataset in year t, encoded in the same way as the population win rate. The list of top 50 players was compiled separately for each year using the average Elo rating of the players in that year. We denote the corresponding coefficient by $\beta_{\text{top50-win},i}$.

3. The frequency of the strategy among the top 50 players in year t. We denote the corresponding coefficient by $\beta_{\text{top50-freq},i}$.

These features represent biases different from frequency dependence that could also contribute to cultural fitness of moves; if the average outcome significantly affects move choice, success bias is present in transmission, as represented by coefficients $\beta_{\text{win},i}$ and $\beta_{\text{top50-win},i}$. Similarly, prestige bias could be important for transmission if players imitate the top 50 players as represented by coefficients $\beta_{\text{top50-win},i}$ and $\beta_{\text{top50-freq},i}$.

The extra features are included in the model as an exponential factor $\exp(\boldsymbol{\beta}_i \cdot \boldsymbol{y}_t^i)$. This choice of factor has two purposes: first, it ensures that the variables α_i stay positive for all parameter values and data points; second, it represents *multiplicative* effects of several types of transmission biases, a common approach both in theoretical models of cultural evolution (see e.g. Denton et al. 2020; Lappo, Denton, and Feldman 2023) and in analyses of experimental data (Barrett, McElreath, and Perry 2017; Deffner, Kleinow, and McElreath 2020; Canteloup et al. 2021).

5.2.4 Inference

In total, the parameter vector $\boldsymbol{\theta} = (c_j^i, \boldsymbol{\beta}_i)$ has length $7k$, where k is the number of different moves played in a given position. For each move, there are three coefficients $\beta_{\text{win},i}$, $\beta_{\text{top50-win},i}$, $\beta_{\text{top50-freq},i}$, as well as four values $c_1^i, c_2^i, c_3^i, c_4^i$ characterizing the function f_i in eq. (17).

We choose to fit the model in a Bayesian framework using Markov Chain Monte Carlo sampling, as this choice makes implementation of the model straightforward and allows us to obtain both point estimates and uncertainty quantification from the same analysis. To conduct Bayesian inference, we need to specify a prior distribution for $\boldsymbol{\theta}$. Following Gelman, Carlin, et al. (2020), we specify non-informative priors for each parameter. Each constant segment c_j^i of each function f_i was assigned an Exp(1) prior, such that f_i is always non-negative, and the prior mean of f_i is equal to one, corresponding to neutrality. We assigned each parameter $\boldsymbol{\beta}_i$ a normal $\mathcal{N}(0, 1)$ prior and standardized the corresponding features $\boldsymbol{y}_t^i$ to have zero mean and unit variance. Given these priors and the model likelihood (defined in eqs. (18) and (19)), samples were generated from the posterior distribution using the Hamiltonian Markov Chain Monte-Carlo sampler provided by the Stan software package (Gelman, Lee, and Guo 2015; Stan Development Team 2023). For this procedure, we only consider the data from 1980 to 2019, since earlier years have significantly less data available.

Many moves were played only a few times in the whole dataset. To prevent extremely rare moves from inflating the number of parameters, we have combined moves that individually have average frequency less than 2% into a single category called “other.” In addition, it is commonly accepted by professional players that rare moves serve the same purpose: to take the opponent “out of theory” into positions where neither player had spent significant time preparing, leading to more chaotic and tense games.

There are also years in which some move counts are equal to zero, and in this case, our assumption that move counts are nonzero is violated. To remedy this situation, in computational inference we replace the parameter $\boldsymbol{\alpha}$ from eq. (19) by $\boldsymbol{\alpha} + 1$, such that for all strategies,

$$\alpha_i = 1 + \exp(\boldsymbol{\beta}_i \cdot \boldsymbol{y}_t^i) f_i(x_t^i / N_t) x_t^i. \tag{20}$$

This approach is commonly used to deal with the potential for zero counts of rare categories in models involving multinomial likelihoods. For example, it is used in Dirichlet-multinomial modeling of ecological data (Harrison et al. 2020) and in multinomial "assignment tests" of individuals to populations in genetics (Paetkau et al. 1995; Rosenberg 2005). For moves with non-zero counts, this correction biases expectations from x_t^i/N_t to $(x_t^i+1)/(N_t+K)$, where K is the number of strategies. The bias is negligible when move counts are in the hundreds or above.

6 Modeling results

We discuss model fits for three positions at three different depths in the game tree: the Queen's Pawn opening at ply 2 (**1. d4**), the Caro-Kann opening at ply 5 (**1. e4 c6 2. d4 d5**), and the Najdorf Sicilian at ply 11 (**1. e4 c5 2. Nf3 d6 3. d4 cxd4 4. Nxd4 Nf6 5. Nc3 a6**). The parameters of the Stan HMC sampler and convergence diagnostics for each position are reported in Supplementary Information S1. In total, there are $N = 1,083,146$ games with the Queen's Pawn opening, $N = 80,890$ games with the Caro-Kann opening, and $N = 82,557$ games with the Najdorf Sicilian opening. Input data such as raw strategy counts and win rates in each year appear in Supplementary Information S2.

Figure 4 shows the original frequency data, the move choice probabilities as estimated by the model, and estimates of frequency-dependent fitness ${f_i}'(x_t^i/N_t) = f_i(x_t^i/N_t)/\bar{f}_t$ of moves over time, as defined in eqs. (12) and (13). Comparing the first and second rows of panels in Figure 4, our model fits the data well, with estimated move choice probabilities (panels B, E, H) matching the actual move frequencies (panels A, D, G). The estimates of the parameters f_i and $\boldsymbol{\beta}_i$ are presented in Figures 5 and 6, respectively. For point estimates, the posterior median is used, and for quantifying uncertainty, we report posterior 1% and 99% quantiles for each estimate. In our analysis, we focus on effects $\boldsymbol{\beta}$ for which the middle 98% of the distribution does not contain zero and on significant effects that have reasonable justifications in chess literature or history. Finally, Figure 7 illustrates frequency dependence in the choice of strategies using posterior predictive sampling. We discuss Figure 7 in detail below.

6.1 Frequency dependence: Queen's Pawn opening

Considering the responses to the Queen's Pawn opening in Figure 4A, from 1980 to 2005, the move **d5** is, on average, increasing in popularity, with this trend reversing after 2005. The move **Nf6** shows the opposite dynamics. In fact, in World Championship matches of 2016, 2018, and 2021, players responded with **Nf6** in all but one game in this position (see e.g. Chessgames.com 2023b). Gradual changes can be caused by cultural drift (Bentley et al. 2007) or changes in metastrategy. However, our model suggests that transmission biases may play a role as well. In particular, the values of the fitness functions for **d5** and **Nf6** observed in Figure 4C are higher when they are are at lower frequencies. The plots of frequency-dependent function functions $f_i(x)$ for x from 0 to 1 are shown in Figure 5A, and there is a downward slope in the values of $f_i(x)$ characteristic of negative *frequency-dependent bias*, or anti-conformity. Win rates or features related to top-50 players appear to have no effect on the choice of **d5** or **Nf6** (Figure 6A). The other strategies are played in only a small proportion of games, and for those strategies, it may be hard to distinguish meaningful effects from statistical artifacts.

To further understand the nature of frequency dependence, we plot expected deviations of move choice probability $\mathbb{E}[p_{t+1}^i]$ from random choice ($\mathbb{E}[p_{t+1}^i] = p_t^i$) for initial frequencies $x_t^i/N_t = 0, 0.02, \ldots, 0.98, 1$, keeping other variables constant (see Supplementary Information S3 for a detailed description of the calculation). For the Queen's Pawn opening, this plot appears in Figure 7A. The choice of move **d5** clearly has negative frequency dependence, as it is chosen with probability higher than what is expected under random choice when its frequency is low and with lower probability when its frequency is high, with deviations from random choice as large as 1.9%. Similar behavior can be seen for the move **Nf6**.

6.2 Success bias: Caro-Kann

In the Caro-Kann opening, the move **exd5** is used less and less in more recent years (Figure 4D). The plot of move fitnesses in Figure 4F and the choice probability plot in Figure 7B suggest that negative frequency-dependent dynamics play a role in determining this behavior. However, the functions f_i are not the only determinants of move frequencies in our model; the coefficients $\boldsymbol{\beta}_i$ shown in Figure 6B suggest that the choice to play the move **exd5** is affected by the win rate in the population, indicating *success bias*. The decrease in the frequency of **exd5** then comes from many players losing after playing this move (see Figure S2K). Indeed, computer engines have shown that the move **e5** provides the strongest winning probability for the player, while after **exd5** the opponent can "equalize" the position and take over the game (Schandorff 2021).

6.3 Prestige bias: Najdorf Sicilian

In the case of the Najdorf Sicilian, in Section 5.1, we highlighted **h3** as a recent strong trend. The frequency-dependent fitness function $f_{\mathbf{h3}}$ shows that there is no negative frequency-dependent bias for a choice of **h3** (Figure 5C); in fact, Figure 7C shows that **h3** is, on average, chosen with probability greater than random choice at every value of the frequency in the previous year. This result suggests that the move is a genuine innovation, becoming more popular "on its own merit" and not because of frequency-dependent trends. The coefficient for the win rate among the top 50 players, $\beta_{\mathbf{h3},\text{top50-win}}$ is large (Figure 6C), meaning that the increase in the frequency of **h3** could possibly be due to a trend started by elite players, which then led to wider adoption and development of theory. We conclude that the choice to play **h3** is subject to *prestige bias*. In chess literature, side pawn pushes such as **h3**, **h4**, **a3**, and **a4** in various positions are ideas introduced by strong chess engines (Sadler and Regan 2019, Ch. 9) in the most recent decade. This trend may explain why top players, who often have teams analyzing engine suggestions for them, have been adopting the move **h3**, subsequently influencing the general population.

6.4 Game sample size N_s

Finally, we address the way our model characterizes the variance of move counts in the data. As we have discussed in Section 5.2.2, the mean fitness $\bar{f}_t$ controls the variance of x^i_{t+1} conditional on model parameters and x^i_t. Mathematically, this influence can be seen as follows. As a shorthand, let

$$p_i = \frac{f_i(x^i_t/N_t)\, x^i_t}{\sum_{j=1}^{K} f_i(x^j_t/N_t)\, x^j_t} \tag{21}$$

be the "frequencies" of strategies assuming no effect of prestige or success biases. Then the variance of x^i_{t+1} is (Johnson, Kotz, and Balakrishnan 1997, p. 81):

$$\mathrm{Var}_{\mathrm{DM}}(x^i_{t+1} \mid x^i_t) = N_{t+1} p_i (1-p_i) \left(\frac{N_{t+1} + N_t \bar{f}_t}{1 + N_t \bar{f}_t} \right). \tag{22}$$

The last term of eq. (22) is a decreasing rational function of $\bar{f}_t$, so $\mathrm{Var}_{\mathrm{DM}}(x^i_{t+1} \mid x^i_t)$ decreases as $\bar{f}_t$ grows.

In the fitted models, the mean fitness $\bar{f}_t$ is consistently below 1 for all three positions considered, equal to ~ 0.22 for the Queen's Pawn, ply 2 position (approximately constant over time), ~ 0.3 for Caro-Kann, ply 5, and ~ 0.45 for Najdorf Sicilian, ply 11 (Figure S3A). That we have observed $\bar{f}_t < 1$ can be interpreted in relation to players' behavior. Mechanistically, our model describes players observing move counts in a previous year, adjusting their preferences because of transmission biases, and then selecting a move with *higher variance* than what is expected if $\bar{f}_t = 1$, corresponding to multinomial choice. We define *game sample size* $N_s(t) = \bar{f}_t N_t$ to be the number of games in the population at time t that achieves the same value for the variance $\mathrm{Var}_{\mathrm{DM}}(x^i_{t+1} \mid x^i_t)$ as in eq. (22) under the condition $\bar{f}_t = 1$. Indeed, with game sample size defined as $N_s(t) = \bar{f}_t N_t$, eq. (22) becomes

$$\mathrm{Var}_{\mathrm{DM}}(x^i_{t+1} \mid x^i_t) = N_{t+1} p_i (1-p_i) \left(\frac{N_{t+1} + N_s(t)}{1 + N_s(t)} \right), \tag{23}$$

so that now a mechanistic interpretation of our model consists of players observing move counts in a population of size $N_s(t)$, adjusting their preferences according to transmission biases, and then choosing the strategy according to a multinomial distribution.

As the game progresses from early to later positions, the players sample a higher fraction of all games in their decision-making process (Figure S3). Possibly, the fraction of games sampled by players is low for early positions because tens of thousands of professional games each year start with a move **d4** (Figure S2A), and it is likely that players cannot monitor all of these games. However, a player who specializes in playing the Najdorf Sicilian may pay attention to a larger proportion of games involving this opening, because the total number of games to analyze is much smaller for ply 11 in the Najdorf Sicilian (Figure S2C) than in the Queen's Pawn at ply 2 (Figure S2A).

7 Discussion

We have developed a population-level model for the influence of transmission biases on move choice in chess. We have shown that many of the moves analyzed are under negative frequency-dependent cultural selection, having higher fitness and being favorably selected with probability greater than random choice at lower frequencies (Figures 4 and 7). This result suggests that anti-conformity is important in the transmission of chess opening strategies. In addition, our model is able to identify moves for which other factors play a role: the dynamics of **h3** in the Najdorf Sicilian are affected by the win rate among the top 50 players (Figure 6C), indicating the presence of prestige bias, and the choice of **exd4** in the Caro-Kann suggests success bias (Figure 6B).

We have also inferred absence of significant success bias for many strategies, consistent with our discussion in Section 3: a win in chess is conditional on strong performance at every move, so making decisions about the opening based on the average eventual outcome may not be the best choice from many positions. Similarly, following choices of top players would be effective only if a strong continuation were found. Support for our findings of strong success bias in the Caro-Kann and prestige bias in the Najdorf Sicilian comes from information commonly known to professional chess players, such as new insights from extensive computer analysis, or new styles of play introduced by computer players.

In addition to measuring transmission biases, we have introduced a concept of "game sample size" N_s that appears naturally from the analysis of game counts (Section 6.4). N_s can be interpreted as the number of games that players observe when making use of social information. We have shown that later positions have a greater ratio N_s/N, which could mean that players use more complete information when positions become more complicated and less standardized (Figure S3).

The estimated game sample size relates to several theoretical concepts. First, from the perspective of population genetics, N_s is equivalent to variance effective population size $N_e(t) = \bar{f}_t N_t$ used to account for overdispersed allele counts relative to a standard Wright-Fisher model (Ewens 2004; Caballero 2020, Ch. 3). Second, theoretical studies of conformity typically involve individuals sampling role models from the population and making a choice based on this sample (Boyd and Richerson 1985; Denton et al. 2020; Lappo, Denton, and Feldman 2023). The number of role models is usually taken to be equal to a small number such as 3, which is much smaller than the population size. The value of N_s can be seen as relating to these theoretical models, measuring how many role models are sampled from the population.

7.1 Related and complementary work

Our model complements other recent work on measuring the strength of transmission biases in cultural datasets of competitive activities, such as the studies by Beheim, Thigpen, and McElreath (2014) on Go, Miu et al. (2018) on programming contests, and Mesoudi (2020) on football strategy. Beheim, Thigpen, and McElreath (2014) employed multilevel logistic regression to study social and individual learning involved in the board game Go. They observed strong success bias and *positive* frequency-dependence for the choice of one of the opening moves. Positive frequency-dependence in Go and negative frequency-dependence in chess could be connected to the differences in the communities around each game. Among board games, chess is unique in its use of computer engines. Computer chess engines became widely available to elite players starting from the late 1990s, revolutionizing tournament preparation. Finding the best response in a position or solving a chess puzzle became possible in a matter of seconds rather than hours or days. Post-game analysis now helps players quickly identify and address their weaknesses, which means that players can no longer "catch" many opponents with the same "trick." Playing into popular lines can also lead to positions in which the opponent has the most preparation. In contrast, the space of possible moves in the opening is much larger in Go, and computers have reached human level only in the most recent decade. Hence, the effectiveness of studying a *particular* position in Go is diminished, and players may choose to conform to a popular strategy for their first move and hope to outplay the opponent later in the game.

Transmission biases and social learning strategies in various games have also been measured in field observations and experiments (e.g. Aplin, Sheldon, and McElreath 2017; Barrett, McElreath, and Perry 2017; Deffner, Kleinow, and McElreath 2020; Vale et al. 2017; Canteloup et al. 2021). Studies using experimental data typically involve models that estimate parameters for each observed individual or category of individuals, whereas we focus on analysis of large-scale population-level data. Still, some aspects of such models are similar to our Dirichlet-multinomial approach. For example, in the experience-weighted attraction (EWA) model employed by Barrett, McElreath, and Perry (2017) to analyze social learning in Capuchin monkeys (also used in Canteloup et al. (2021) and Deffner, Kleinow, and McElreath (2020)), decisions are influenced by a convex combination of functions representing individual and social learning, and different social biases are encoded in a multiplicative way similar to eq. (19) in our model. This similarity suggests that our model could potentially be modified to model move choice of each player via a Dirichlet-multinomial likelihood, enabling comparisons of learning modalities between individual players.

Frequency-dependent selection has previously been measured by Newberry and Plotkin (2022) in other large datasets such as baby name statistics and dog breed popularity data. These authors focused on modeling "exchangeable" entities, for which selection acts on every variant in exactly the same way. They estimate a single fitness function that is shared by every cultural variant and that characterizes average frequency-dependence in the population. Chess differs from such contexts in that it contains the concept of a "win." Each chess move leads to a different position, altering the winning chances of each player, so that strategies at different stages of the game are dependent. Thus, our model assigns a separate fitness function to each individual strategy, treating strategies as *nonexchangeable*.

Our model also extends the multinomial model of Newberry and Plotkin (2022) by incorporating the Dirichlet distribution into the model likelihood. This approach has a clear mechanistic interpretation in terms of players' behaviors and allows us to perform efficient Bayesian inference of model parameters. Statistical models of count data based on the Dirichlet-multinomial likelihood are known in many related areas, including linguistics (e.g. Madsen, Kauchak, and Elkan 2005), human genetics (Wang et al. 2023), molecular ecology (Harrison et al. 2020), and microbiome data analysis (e.g. Osborne, Peterson, and Vannucci 2022; Wadsworth et al. 2017).

7.2 Caveats

The parameters of our model can be represented in two different ways. One uses k fitness functions f_i that are only constrained to be non-negative and are naturally suited to Bayesian inference. Another uses k functions f_i' (eq. (13)) that are required to sum to one, together with the mean fitness $\bar{f}_t$ (eq. (16)). While estimates of f_i (Figure 5) show presence of frequency-dependent dynamics, it is hard to characterize strength and significance of frequency-dependence using the values of f_i. To understand strength and significance of frequency-dependent effects we plot the growth rates f_i' of strategies (Figure 4) and compute expected deviation of move counts from random choice (Figure 7). Other analyses could potentially be used, for example evaluating whether the function f_i is significantly different from a constant function.

Another statistical issue that could affect our inferences is the possibility of correlated input features, so that the $\boldsymbol{\beta}$ coefficients might not be easily identifiable. However, features for the games played by the top-50 players show behaviors that differ from those of the total population of around 15,000 players (Figure S2). Thus, for the factors we consider, it appears that distinguishing the influence of the top-50 players from a general influence of master-level players is indeed possible.

Our model incorporates only a subset of possible features that could be relevant to move choice, such as highly developed theory or objective strength as determined by computer evaluation (Section 3). However, the presence of significant success bias and prestige bias could correspond to mechanisms of social learning *about* these other features. For example, suppose a player observes several successful games in top tournaments with **h3** in the Najdorf Sicilian, and then studies the move. The player could learn about the enthusiasm of modern computer engines for this move and could incorporate it into future play. For our model, this mechanism is indistinguishable from the player simply copying a successful move. This reasoning about a player's mechanistic evaluation process suggests a potential direction for further modeling that would incorporate varying individual behaviors and knowledge about position evaluation.

7.3 Conclusion

Data from the last five decades of high-level chess games can be evaluated in the context of cultural transmission and evolution. We have shown that the cultural "features" of transmission can be measured from move choice decisions in various positions by professional players. In particular, we have inferred influences of frequency-dependent bias, success bias (win rate), and prestige bias (the use of the move by the very top players). The prevalence of anti-conformity and the lack of strong success bias for many strategies reflects the nature of opening play in chess, which involves extensive preparation and assessment of opponents' likely preparation. We have also connected the presence or absence of transmission biases with chess theory. The fact that many of our quantitative results correspond to ideas well-known to professional chess players suggests that our modeling could be useful to chess analysts and historians. In particular, many qualitative explanations are available for the popularity of certain strategies, and a quantitative evaluation of move frequency dynamics could help test the narratives familiar to chess players with statistical evidence. More broadly, our statistical approach could potentially be used to complement the historical study of cultural trends in other games that contain discrete choices, or even in other cultural domains in which circumscribed discrete data are recorded.

Data and code. The code to generate figures in this paper and links to access the dataset are available at github.com/EgorLappo/cultural_transmission_in_chess.

Funding. We acknowledge NSF grant BCS-2116322 and grant 61809 from the J. T. Templeton Foundation for support.

Acknowledgments. We thank Sharon Du for suggesting the problem and Kaleda Denton, Bret Beheim, and an anonymous reviewer for helpful comments on the manuscript.

References

Aplin, Lucy M, Ben C Sheldon, and Richard McElreath (2017). "Conformity does not perpetuate suboptimal traditions in a wild population of songbirds". In: *Proceedings of the National Academy of Sciences* 114.30, pp. 7830–7837.

Barrett, Brendan J, Richard L McElreath, and Susan E Perry (2017). "Pay-off-biased social learning underlies the diffusion of novel extractive foraging traditions in a wild primate". In: *Proceedings of the Royal Society B* 284.1856, p. 20170358.

Beheim, Bret Alexander, Calvin Thigpen, and Richard McElreath (2014). "Strategic social learning and the population dynamics of human behavior: The game of Go". In: *Evolution and Human Behavior* 35.5, pp. 351–357.

Bentley, R Alexander et al. (2007). "Regular rates of popular culture change reflect random copying". In: *Evolution and Human Behavior* 28.3, pp. 151–158.

Blasius, Bernd and Ralf Tönjes (2009). "Zipf's law in the popularity distribution of chess openings". In: *Physical Review Letters* 103.21, p. 218701.

Boyd, Robert and Peter J Richerson (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago press.

Caballero, Armando (2020). *Quantitative Genetics*. Cambridge: Cambridge University Press.

Canteloup, Charlotte et al. (2021). "Processing of novel food reveals payoff and rank-biased social learning in a wild primate". In: *Scientific Reports* 11.1, p. 9550.

Capablanca, Jose R (1935). *A primer of chess*. New York and London: Houghton Brace Jovanovich.

Cavalli-Sforza, Luigi Luca and Marcus W Feldman (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton: Princeton University Press.

Chessgames.com (2023a). *Capablanca vs Alekhine WCM 1927*. URL: https://www.chessgames.com/perl/chesscollection?cid=1033458 (visited on 07/21/2023).

— (2023b). *Carlsen - Nepomniachtchi World Championship Match*. URL: https://www.chessgames.com/perl/chess.pl?tid=105291 (visited on 07/21/2023).

Corsini, Noemi and Cinzia Viroli (2022). "Dealing with overdispersion in multivariate count data". In: *Computational Statistics & Data Analysis* 170, p. 107447.

De Firmian, Nick (2008). *Modern Chess Openings*. 15th ed. New York: Random House Puzzles & Games.

Deffner, Dominik, Vivien Kleinow, and Richard McElreath (2020). "Dynamic social learning in temporally and spatially variable environments". In: *Royal Society Open Science* 7.12, p. 200734.
Denton, Kaleda Krebs et al. (2020). "Cultural evolution of conformity and anticonformity". In: *Proceedings of the National Academy of Sciences* 117.24, pp. 13603–13614.
Desjarlais, Robert R (2011). *Counterplay: An anthropologist at the chessboard*. Berkeley and Los Angeles, California: University of California Press.
Di Fatta, Giuseppe, Guy McC Haworth, and Kenneth W Regan (2009). "Skill rating by Bayesian inference". In: *2009 IEEEs Symposium on Computational Intelligence and Data Mining*. IEEE, pp. 89–94.
Elo, Arpad E (1978). *The Rating of Chessplayers, Past and Present*. London: Batsford.
Euwe, Max and John Nunn (1997). *The Development of Chess Style*. New York: David McKay Company Inc.
Ewens, Warren John (2004). *Mathematical Population Genetics: Theoretical Introduction*. Vol. 1. New York: Springer.
Gelman, Andrew, John B Carlin, et al. (2020). *Bayesian Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.
Gelman, Andrew, Daniel Lee, and Jiqiang Guo (2015). "Stan: A probabilistic programming language for Bayesian inference and optimization". In: *Journal of Educational and Behavioral Statistics* 40.5, pp. 530–543.
Gobet, Fernand (2018). *The Psychology of Chess*. New York: Routledge.
Gobet, Fernand, Jean Retschitzki, and Alex de Voogt (2004). *Moves in Mind: The Psychology of Board Games*. New York: Psychology Press.
Harrison, Joshua G et al. (2020). "Dirichlet-multinomial modelling outperforms alternatives for analysis of microbiome and other ecological count data". In: *Molecular Ecology Resources* 20.2, pp. 481–497.
Henrich, Joseph and Richard McElreath (2003). "The evolution of cultural evolution". In: *Evolutionary Anthropology* 12.3, pp. 123–135.
Hooper, David and Ken Whyld (1992). *The Oxford Companion to Chess*. Oxford: Oxford University Press.
Johnson, Norman Lloyd, Samuel Kotz, and Narayanaswamy Balakrishnan (1997). *Discrete multivariate distributions*. Vol. 165. Wiley Series in Probability and Statistics. New York: Wiley.
Kendal, Rachel L et al. (2018). "Social learning strategies: Bridge-building between fields". In: *Trends in Cognitive Sciences* 22.7, pp. 651–665.
Kotov, Aleksandr and Mikhail Mikhailovich Yudovich (1961). *The Soviet school of chess*. New York: Dover.
Laland, Kevin N (2004). "Social learning strategies". In: *Animal Learning & Behavior* 32.1, pp. 4–14.
Lappo, Egor, Kaleda K Denton, and Marcus W Feldman (2023). "Conformity and anti-conformity in a finite population". In: *Journal of Theoretical Biology* 562, p. 111429.
Madsen, Rasmus E, David Kauchak, and Charles Elkan (2005). "Modeling word burstiness using the Dirichlet distribution". In: *Proceedings of the 22nd international conference on Machine learning*, pp. 545–552.
Mesoudi, Alex (2020). "Cultural evolution of football tactics: strategic social learning in managers' choice of formation". In: *Evolutionary Human Sciences* 2, e25.
Miu, Elena et al. (2018). "Innovation and cumulative culture through tweaks and leaps in online programming contests". In: *Nature Communications* 9.1, p. 2321.
Newberry, Mitchell G and Joshua B Plotkin (2022). "Measuring frequency-dependent selection in culture". In: *Nature Human Behaviour*, pp. 1–8.
Osborne, Nathan, Christine B Peterson, and Marina Vannucci (2022). "Latent network estimation and variable selection for compositional data via variational EM". In: *Journal of Computational and Graphical Statistics* 31.1, pp. 163–175.
Paetkau, D et al. (1995). "Microsatellite analysis of population structure in Canadian polar bears". In: *Molecular Ecology* 4.3, pp. 347–354.
Perotti, Juan I et al. (2013). "Innovation and nested preferential growth in chess playing behavior". In: *Europhysics Letters* 104.4, p. 48005.
Regan, Kenneth W, Bartlomiej Macieja, and Guy McC Haworth (2011). "Understanding distributions of chess performances". In: *Advances in Computer Games*. Springer, pp. 230–243.
Regan, Kenneth Wingate, Tamal Biswas, and Jason Zhou (2014). "Human and computer preferences at chess". In: *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Ribeiro, Haroldo V et al. (2013). “Move-by-move dynamics of the advantage in chess matches reveals population-level learning of the game”. In: *PLoS One* 8.1, e54165.

Rosenberg, Noah A (2005). “Algorithms for selecting informative marker panels for population assignment”. In: *Journal of Computational Biology* 12.9, pp. 1183–1201.

Sadler, Matthew and Natasha Regan (2019). *Game Changer. AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*. Amsterdam: New In Chess.

Schaigorodsky, Ana L, Juan I Perotti, and Orlando V Billoni (2016). “A study of memory effects in a chess database”. In: *PLoS One* 11.12, e0168213.

Schandorff, Lars (2021). *Playing the Caro-Kann*. Glasgow: Quaity Chess UK.

Shannon, Claude E (1950). “A chess-playing machine”. In: *Scientific American* 182.2, pp. 48–51.

Shenk, David (2011). *The Immortal Game: A History of Chess*. Canada: Anchor Canada.

Sousa, João Dinis de (2002). “Chess moves and their memomics: a framework for the evolutionary processes of chess openings.” In: *Journal of Memetics* 6.2, pp. 18–43.

Stan Development Team (2023). “Stan Modeling Language Users Guide and Reference Manual, version 2.31”. https://mc-stan.org.

Vale, Gillian L et al. (2017). “Acquisition of a socially learned tool use sequence in chimpanzees: Implications for cumulative culture”. In: *Evolution and Human Behavior* 38.5, pp. 635–644.

Wadsworth, W Duncan et al. (2017). “An integrative Bayesian Dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data”. In: *BMC Bioinformatics* 18.1, pp. 1–12.

Wang, Richard J et al. (2023). “Human generation times across the past 250,000 years”. In: *Science Advances* 9.1, eabm7047.

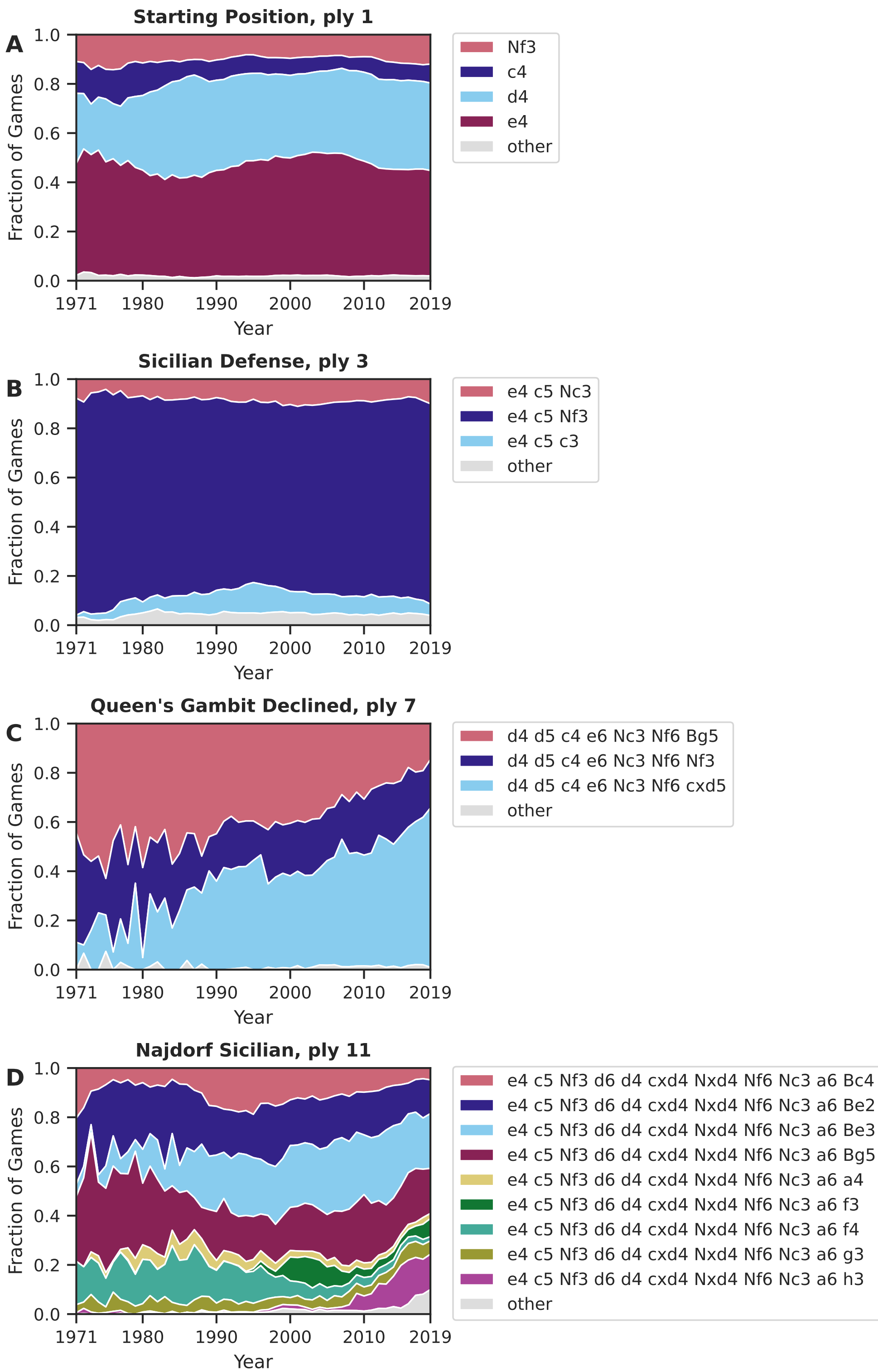

Figure 3: Move frequencies over time. For each panel, the legend presents the whole sequence of moves from the start of the game, with odd moves played by white, and even moves by black. The "other" category contains all rare moves that individually have average frequency less than 2%, with the average taken over all years. (A) Starting Position. (B) Sicilian Defense. (C) Queen's Gambit Declined. (C) Najdorf Sicilian. See the main text for a discussion of each position.

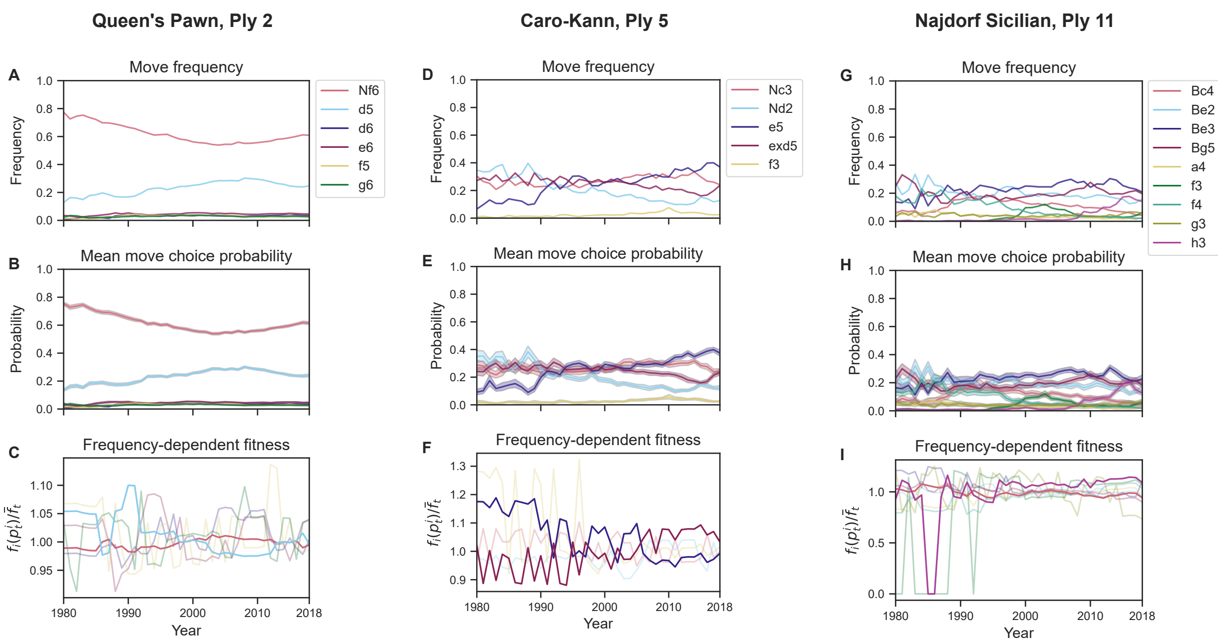

Figure 4: Dirichlet-multinomial model fits for move choice in three different positions: the Queen's Pawn opening at ply 2 (**1. d4**), the Caro-Kann opening at ply 5 (**1. e4 c6 2. d4 d5**), and the Najdorf Sicilian at ply 11 (**1. e4 c5 2. Nf3 d6 3. d4 cxd4 4. Nxd4 Nf6 5. Nc3 a6**). Panels A, D, and G show move frequencies x_t^i/N_t. Panels B, E, and H show posterior means of probabilities of move choice in the year t, with grey lines marking the range containing the middle 98% of the posterior density. Panels C, F, and I show frequency-dependent fitness $f_i(x_t^i/N_t)/\bar{f}_t$ of moves over time, with the values computed using posterior medians of the f_i. (A) Move frequencies, Queen's Pawn, ply 2. (B) Mean move choice probability, Queen's Pawn, ply 2. (C) Frequency-dependent fitness, Queen's Pawn, ply 2. (D) Move frequencies, Caro-Kann, ply 5. (E) Mean move choice probability, Caro-Kann, ply 5. (F) Frequency-dependent fitness, Caro-Kann, ply 5. (G) Move frequencies, Najdorf Sicilian, ply 11. (H) Mean move choice probability, Najdorf Sicilian, ply 11. (I) Frequency-dependent fitness, Najdorf Sicilian, ply 11. The curves for the "other" category are omitted in all plots as the category is too rare to give meaningful results. The model was fitted for years 1980-2019, and the move fitnesses are estimated for all years except 2019.

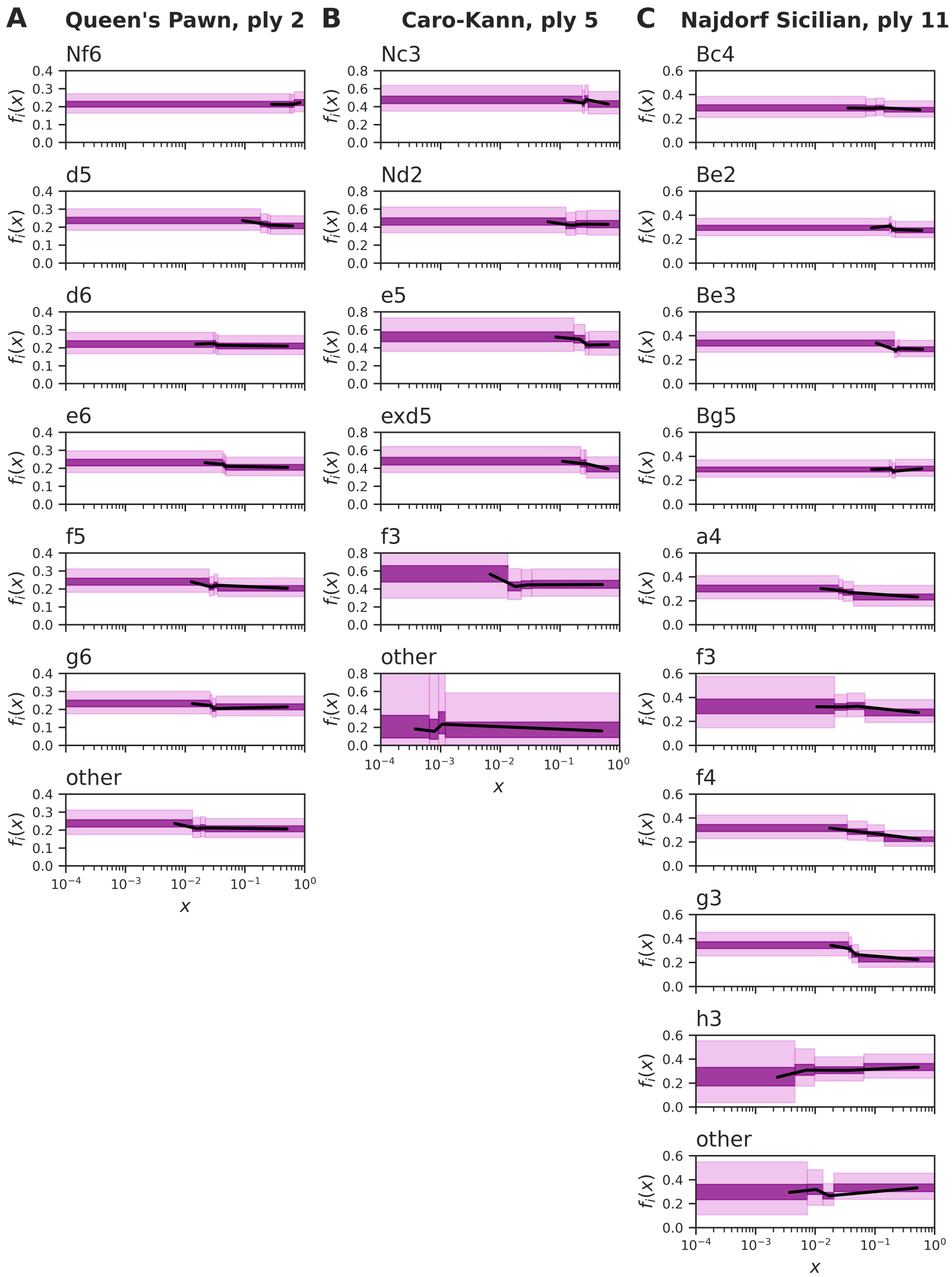

Figure 5: Estimated frequency-dependent fitness functions f_i. The black line connects the posterior medians for the four constant segments, bright purple shows regions containing 60% of the posterior density, and light purple shows regions containing 98% of the posterior density. (A) Queen's Pawn, ply 2. (B) Caro-Kann, ply 5. (C) Najdorf Sicilian, ply 11.

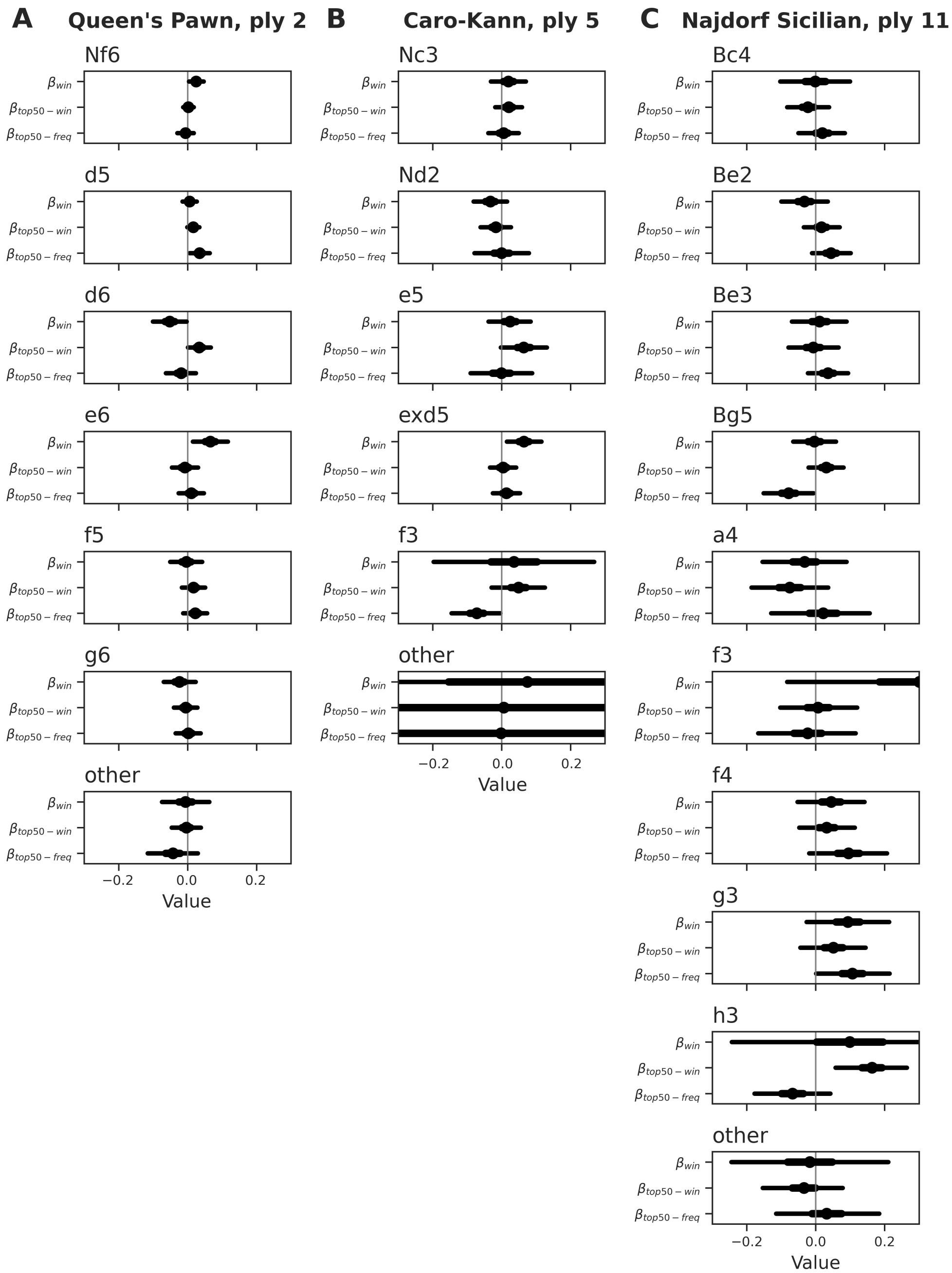

Figure 6: Estimated coefficients $\boldsymbol{\beta}_i$. A point marks the posterior median, the thick line marks the region containing 60% of the posterior density, and the thin line shows the region containing 98% of the posterior density. The coefficients presented are: β_{win}, the effect of the average outcome of games in the year previous to that in which a given move was played; $\beta_{\text{top50-win}}$, the effect of the average outcome of games involving players in the top 50 in the previous year; and $\beta_{\text{top50-freq}}$, the effect of the frequency of a given move in games involving players in the top 50 in the previous year (see Section 5.2.3). (A) Queen's Pawn, ply 2. (B) Caro-Kann, ply 5. (C) Najdorf Sicilian, ply 11.

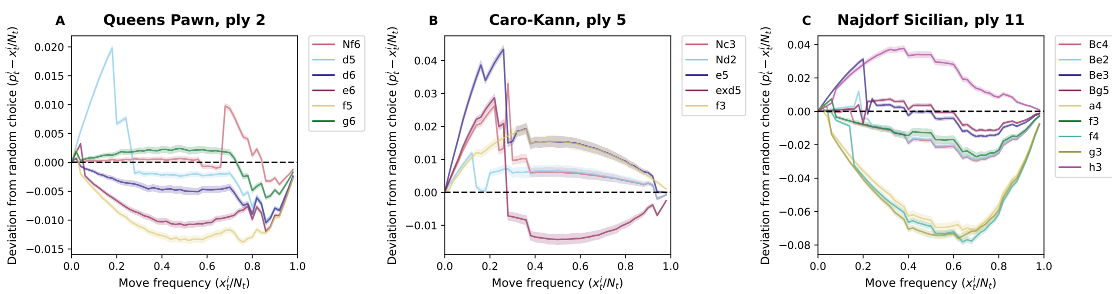

Figure 7: Dependence of move choice probability on strategy frequency. For each strategy, the corresponding curve shows the deviation from random choice of posterior move choice probability in year $t+1$ as the frequency of that strategy in year t varies from 0 to 1. The values were computed using 1000 samples from the posterior for each initial move frequency in year t, as described in Supplementary Information S3. Shading around the curves corresponds to 98% bootstrap confidence intervals for the mean. The curves for the “other” category are omitted in all plots, as these are too rare to give meaningful results. (A) Queen’s Pawn, ply 2. (B) Caro-Kann, ply 5. (C) Najdorf Sicilian, ply 11.