Harvard Data Science Review • Issue 7.2, Spring 2025

An Investigation Into Probabilities of Streaks in Online Chess

Jeffrey S. Rosenthal[1]

[1]University of Toronto, Toronto, Ontario, Canada

The MIT Press

Published on: Apr 30, 2025

DOI: https://doi.org/10.1162/99608f92.6939122d

ABSTRACT

We investigate the probabilities of long chess winning streaks by the player Hikaru on the online site Chess.com, in response to recent high-level allegations of cheating. We find that, under a reasonable model of chess win and draw probabilities, the observed streaks are comfortably within the range of statistical expectation. We conclude that these streaks do not provide evidence of cheating or suspicious behavior.

Keywords: chess, rating, winning streak, hot hand, cheating

Media Summary

Former world chess champion Vladimir Kramnik has raised concerns about certain long winning streaks in online chess by the top-level player Hikaru Nakamura, as reported in the *New York Times* and elsewhere. In this article, we investigate the probabilities of such long winning streaks using statistical modeling. We conclude that these streaks are comfortably within the range of statistical expectation, and do not provide evidence of cheating or suspicious behavior.

---

1. Introduction

Chess is a strategy board game going back many hundreds of years (Cazaux & Knowlton, 2017). In recent years, online chess has become extremely popular, with millions of active players (Keener, 2022). Meanwhile, chess-playing computer programs or 'engines,' with names like Chessmaster, Fritz, Komodo, Houdini, Stockfish, and Chessbase, are now much better than humans and could easily be consulted (either manually or automatically) during online matches to achieve superior play, which is strictly prohibited (Chess.com, 2024a). The most popular online chess website, Chess.com, actively monitors and attempts to catch cheaters through various methods, including comparing player performance online versus in-person (Chess.com, 2024b) and sometimes even requiring that cameras be set up to monitor players in their homes (Chess.com, 2025). Nevertheless, concerns about cheating continue, including at the highest levels (McClain, 2023).

One recent issue involves long streaks of games that were all (or almost all) won by a specific player. In particular, former world champion V. Kramnik has raised concerns about winning streaks of top-level player Hikaru Nakamura (player name: Hikaru), including one recent streak of 46 games in which he won 45 and tied one (McClain, 2023; see also GothamChess, 2024).

I was contacted by Chess.com CEO Erik Allebest, who had seen my *Wired* interview (Wired, 2022), and asked to perform an independent statistical analysis of such winning streaks. To facilitate this, I was supplied (E. Allebest, personal communication, July 15, 2024) with data[1] showing results of all games on Chess.com of seven different top-level players, including Hikaru. I then conducted an independent statistical examination of evidence of unusual or surprising streaks in Hikaru's Chess.com game record. I first examined the nature of chess ratings, expected scores, win and draw probabilities, and game correlations, to establish a model for the probabilities of online chess outcomes. I then used this model to examine the probabilities corresponding to Hikaru's winning streaks.

When my original preliminary report (Rosenthal, 2024a) was publicized on Chess.com (Svensen, 2024) and elsewhere (Hikaru, 2024a), Kramnik posted a response video and comments (Kramnik, 2024a) with numerous criticisms. I responded to that in an addendum (Rosenthal, 2024b), and some of those issues are incorporated herein. Months later, while extensively revising this article, I discovered that Kramnik had posted a second video (Kramnik, 2024b) in response to my addendum, and some of the issues raised there are also incorporated herein. I note that these same win streaks have also been examined in other contexts, including a blog post (Bobyrev, 2023), a Chess.com response (Chess.com, 2023), and a Bayesian perspective (Maharaj et al., 2023), which each reached conclusions similar to mine through different approaches.

2. Informal Discussion

Before getting into the statistical analysis, we begin with some informal discussion. Over more than 10 years, Hikaru has played 57,421 games on Chess.com—a tremendous number. Of these, he has won 45,409, drawn 4,943, and lost 7,069, for a total score of 47,880.5 out of a maximum of 57,421 (83.4%). So, Hikaru has certainly had a very successful record. And, this record does indeed include many long winning streaks. For instance, on December 22, 2018, Hikaru played a total of 139 games, and won his last 116 in a row. So, if all long winning streaks are suspicious, then Hikaru's record would be very suspicious indeed.

However, this raises the question of whether long winning streaks are necessarily always suspicious. This might not be so, for two reasons. First, if Hikaru often plays against much weaker opponents, then his chance of winning each game is very high, so long winning streaks might be less surprising than they first appear. Second, Hikaru has played so many games total, that over such a long period, even surprising events might occur by chance alone.

The first reason—the effect of the relative strength of the two players—involves carefully considering chess players' relative abilities and probabilities of winning, as we shall do herein. For example, during the above 116-game winning streak, Hikaru's chess rating (as discussed herein) averaged 3,017, indicating top-level performance. Meanwhile, his opponents' ratings averaged just 1,526, indicating middle amateur level. This is a difference of +1,491, which indicates that Hikaru's opponents were often *much* weaker chess players than he was, which suggests that he had a very high probability of winning individual games.

The second reason—that even surprising events might occur by chance over a long period—can be illustrated using a simple analogy. Suppose your friend is repeatedly flipping a coin. If they get 12 heads in a row, that seems suspicious. Indeed, the 'raw probability' of getting 12 heads in a row, if you flip a fair coin 12 times, is just 1/4,096, extremely low. However, suppose your friend has flipped the coin a total of 10,000 times. Then they have many more opportunities to obtain 12 heads in a row. Although each specific opportunity has raw probability 1/4,096, the chance of getting *some* sequence of 12 heads at *some* point has much higher probability. This is an example of *multiple testing* (e.g., Ranganathan et al., 2016): in all those 10,000 flips, some sequence of 12 heads, somewhere along the way, is much more likely.

How likely? Well, as a theoretical calculation, this is not so simple, because potential heads streaks are overlapping. For example, if flips number 21 through 32 were all heads, then it is more likely that flips 22 through 33 will all be heads, too. Still, since 10,000/12 = 833.3, there are 833 nonoverlapping sequences of 12 flips contained within 10,000 flips. And these nonoverlapping sequences are independent. Each of them has probability 1/4,096 of being all heads, and hence probability $1-(1/4{,}096)$ of *not* being all heads. So, the probability that *none* of them are all heads is equal to $1-[1-(1/4{,}096)]^{833} \doteq 0.184$. This gives us a theoretical lower bound of 18.4% on the probability of obtaining a sequence of 12 consecutive heads somewhere over the course of 10,000 fair coin flips.

However, the true probability of obtaining such a sequence must be significantly higher than 18.4%, since the overlapping sequences give additional possibilities for success. Computing the true probability exactly is challenging, though it is possible for short sequences using recurrence relations (see, e.g., Szczepanek, 2024). Alternatively, we can run a Monte Carlo simulation (see, e.g., Robert & Casella, 2009), whereby we get our computer to *simulate* flipping 10,000 coins, over and over, and count the percentage of those simulations that have 12 heads in a row at some point. Indeed, I just ran a Monte Carlo simulation of 10,000 fair coin flips, repeated 1,000 times. And, in 69.8% of those repetitions, the longest streak of heads was at least 12. This tells us that the true probability of obtaining a sequence of 12 consecutive heads over the course of 10,000 fair coin flips is approximately 69.8%—quite large, and much more than the theoretical 18.4% lower bound.

In this article, we will apply similar reasoning, including both probability calculations and Monte Carlo simulations, to the more complicated case of winning streaks in online chess.

3. Chess Ratings and Expected Scores

To study the statistics of chess outcomes, we need to assess the probabilities of winning or drawing or losing each game. One way to do this is through chess ratings. Chess.com assigns every player on their site a chess *rating* for each game, based upon their past performance. We wish to use these ratings to compute an expected score (i.e., average outcome) in each game, where the score is 1 for a win, 1/2 for a draw, or 0 for a loss. (We note that, in addition to the Chess.com ratings, many players also have various chess ratings from the international chess federation FIDE. Those FIDE ratings are not public, and are not available for all players. They are usually similar to the Chess.com ratings, but with some differences, which could cause some changes in our results, but we do not consider that here.)

Suppose White has rating A, and Black has rating B. Then one possible formula for White's expected score is the well-known Elo (1978) logistic formula:

$$\frac{1}{1+10^{-(A-B)/400}}\,.$$

It was shown in Rosenthal (2024a) that this Elo formula does fit our Chess.com data reasonably well, but we can do better. The Elo formula is known to have various limitations (e.g., Glickman & Jones, 1999) and is not always used. In fact, Chess.com actually generates their ratings (Chess.com, 2024c) using the more sophisticated Glicko method (Glickman, n.d., 2022), based on the theoretical analysis in Glickman (1999). This method, in its original version, instead gives White's expected score as:

$$\frac{1}{1+10^{-\left[1+\frac{3}{\pi^2}\left(\frac{\ln 10}{400}\right)^2\left(RD_A^2+RD_B^2\right)\right]^{-1/2}(A-B)/400}}\,.$$

Here the factor $[\ldots]^{-1/2}$ in the exponent is a complicated multiplier, which in turn depends on each player's 'ratings deviation' (RD), a measure of the uncertainty in their rating.

Chess.com was unable to provide the individual player RD values, so we cannot use Glicko estimates directly. Furthermore, neither the above Elo nor Glicko formulae take into account the (small) advantage of playing White, that is, moving first (e.g., Glickman & Jones, 1999), even though the average score for White in the data is about 0.52, which is slightly higher than 0.50 indicating an advantage that should be included in the probability model.

Thus, inspired by the above Elo and Glicko formulae, but also allowing for White's small advantage, we model White's expected score as

$$S \;=\; \frac{1}{1+10^{-(A-B+c_1)c_2}}\,.$$

Here c_1 and c_2 are unknown constants, to be estimated from the data as accurately as possible. The value of c_1 represents White's (small) advantage from going first (taken as $c_1 = 0$ in the Elo and Glicko formulas), while c_2 measures the scale of influence of the ratings (taken as c_2 = 1/400 in the original Elo formula, and as the above complicated function of RD values in the Glicko formula). We wish to find the best values of c_1 and c_2 to fit the data, in terms of the rating differences $A - B$ and the average game outcomes.

4. Fitting Expected Scores to the Data

Our data consists of 293,047 chess game results played by the seven top-level players. In finding the best values of c_1 and c_2, the raw game data is difficult to work with, since each rating difference $A - B$ is an integer that could take on several thousand different possible values. To make it more manageable and collect enough games to give meaningful data points, we 'binned' the games together according to their rating difference $A - B$. Specifically, we defined the bin ranges as $\ldots, (-14, -5), (-4, 5), (6, 15), (16, 25), \ldots$. Then, for each bin, we computed the average score for White among all games whose rating difference falls within that bin. This provided us with a reasonable target expected score for White for that bin difference. In this section, we make use of all games in the data, though in later sections we will restrict to specific time controls and players.

Given these target expected scores, we wish to compute the best values of the constants c_1 and c_2. We did this using the *principle of least squares* (e.g., Dekking et al., 2005, Chapter 22), to find the values of c_1 and c_2, which minimize the sum of squares of residuals, that is, make the expected score curve fit the data points as accurately as possible with as small errors as possible. (It might also be possible to use a *weighted* least squares fit, in which each bin is adjusted according to its individual bin count, but we do not consider that here.)

Using this principle of least squares, we determined that the sum of squares is minimized numerically when c_1 = 18 and c_2 = 1/381. The resulting expected score curve is thus:

$$S_1 \;=\; \frac{1}{1 + 10^{-(A-B+18)/381}}\,. \tag{4.1}$$

Here 18 represents the small advantage to playing White (i.e., going first), and the denominator 381 is slightly smaller than the usual 400. This expected score curve fits the binned average scores quite well, as shown in Figure 1, so we will use it in our analysis below.

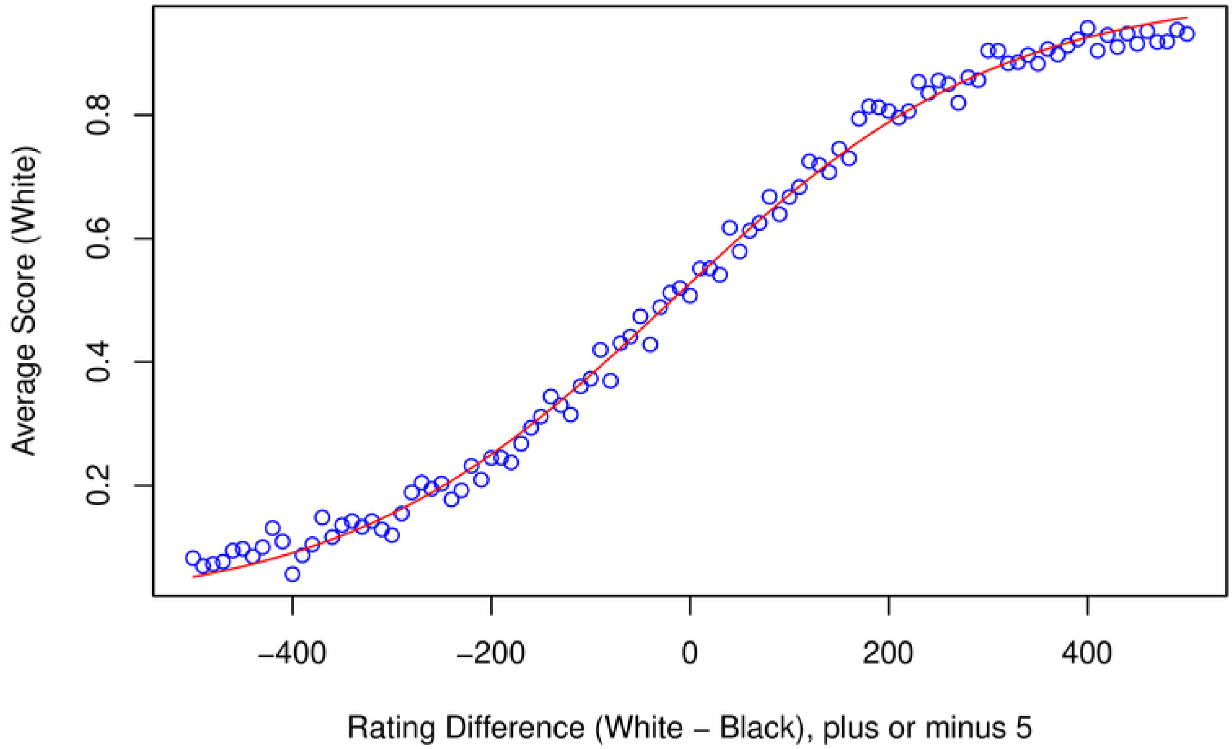

Figure 1. Average scores versus expected score, all games. The average score for White in all games as a function of the binned rating difference (blue), together with the score fit function S_1 from Equation 4.1 (red).

We note that our fit curves are all presented from White's perspective. However, in every game, Black's score equals one minus White's score, with rating difference the negative. So, presenting these graphs from Black's perspective would simply rotate them by 180 degrees, leading to exactly the same fit with no additional information.

5. Draw (Tie) Probabilities

The above expected scores S do not specify what fraction of the score should arise from wins versus from draws. To evaluate the likelihood of long streaks of wins and draws, it is necessary to consider not just the expected score S, but more specifically the probability W of a win and probability D of a draw (tie). Since wins give a score of 1 while ties give a score of $\frac{1}{2}$, we must have

$$S = W + \frac{1}{2}D .$$

In traditional chess tournaments with over-the-board games lasting many hours, draws are quite common. However, in online blitz chess they are less so: just 9.1% of the games in the data set resulted in draws. Binning the data again as above, we observe that the draw probabilities on a log scale are approximately a downward-quadratic function of the rating difference $A - B$, so we model the draw probability as:

$$c_3\, e^{-[(A-B+c_4)c_5]^2}$$

where c_3 and c_4 and c_5 are constants to be estimated. We then again use a least-squares analysis to find the values of these constants, which minimize the sum of squares of the residuals and thus best fit the data. We compute that this best fit occurs when c_3 = 0.120, c_4 = 19, and c_5 = 1/417. This leads to the exponential downward-quadratic function

$$D_1 \;=\; (0.120)\, e^{-[(A-B+19)/417]^2} \qquad (5.1)$$

for the probability of a draw (tie) game. This draw probability curve fits the binned average scores reasonably well, as shown in Figure 2, though with more uncertainty than Figure 1 since there are a smaller number of draws in the data.

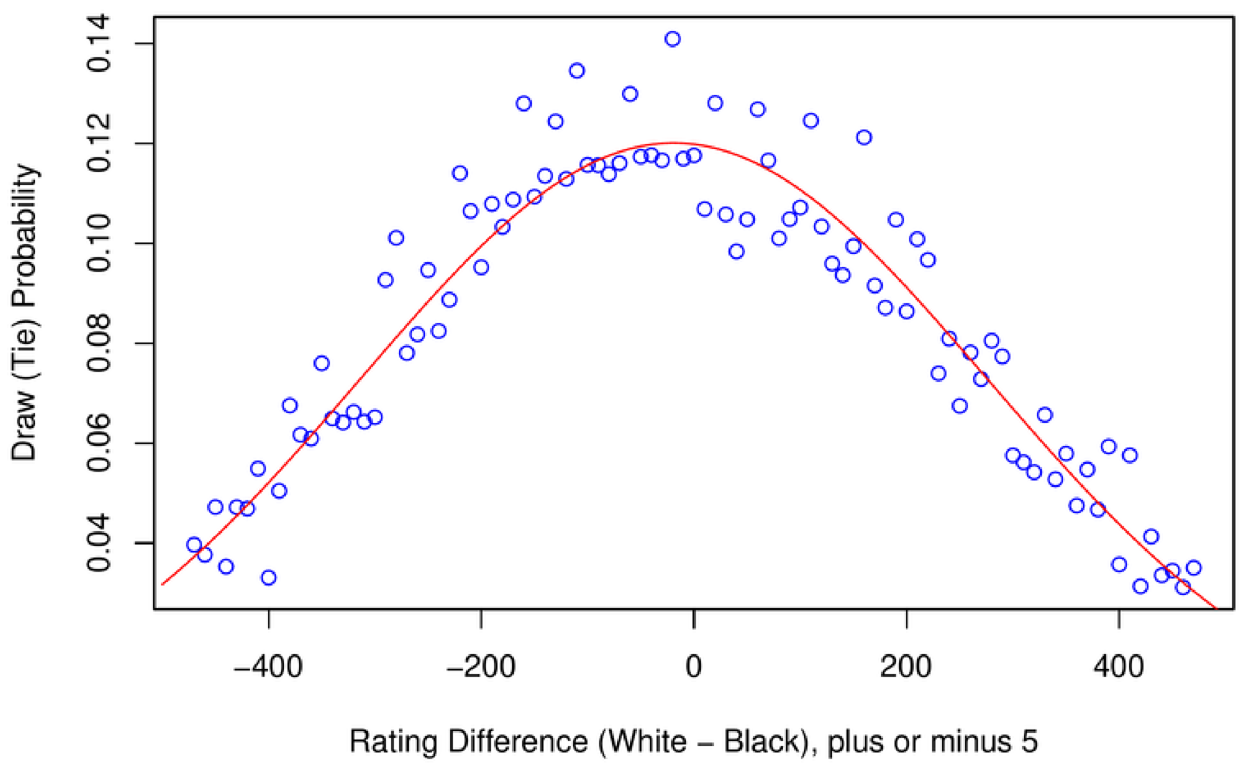

Figure 2. Probability of a draw (tie), all games. The draw fraction in all games as a function of the binned rating difference (blue), together with the draw probability function D_1 from Equation 5.1 (red).

Thus, in our analysis below, in addition to the formula S_1 from Equation 4.1 for expected score, we use the formula D_1 from Equation 5.1 for the probability of a draw. And, since we always have $S = W + \frac{1}{2}D$, it then follows that the probability of a win is given by

$$W_1 \;=\; S_1 - \frac{1}{2}\, D_1 \,.$$

As an aside, we note that the draw probability formula D_1 in Equation 5.1 is largest when $A - B + 19 = 0$, corresponding to the situation where Black's rating is 19 points higher than White's, just enough to overcome White's advantage from going first.

6. Autocorrelations of Excess Scores

To model probabilities of streaks, another issue is the extent to which different games are independent. There is a long history of statistical debate about 'hot hands' in basketball and other sports (e.g., Gilovich et al., 1985; Miller & Sanjurjo, 2018), whereby players are more likely to succeed the next time if they succeeded the previous time. So, it is quite plausible that there would be some 'hot hand' persistence of performance in chess games as well, especially for games played on the same day in rapid succession, perhaps even against the same opponent.

To investigate this, we examined the 57,421 games played by Hikaru on Chess.com. For each game, we computed Hikaru's 'excess score,' defined as his actual score (i.e., 1 or 0 or $\frac{1}{2}$) minus his expected score S_1 from Equation 4.1 for that particular game. This gives a time series list of excess score for all of his 57,421 games, in chronological order.

For such a time series, we can consider its 'autocorrelation function' (ACF), common in Monte Carlo and time series research (see e.g., Ross, 2022, Chapter 18). An autocorrelation is a measure, for each time lag, of the correlation between the excess score on games played at that spacing. For example, at lag = 1, this measures the correlation of excess score between successive games. The autocorrelation at lag = 0 is always equal to one, since games have perfect correlation with themselves. But the autocorrelations at positive lags show the extent to which Hikaru's excess score in one game is correlated with his excess score in subsequent games. The autocorrelation function for Hikaru's excess scores are presented in Figure 3.

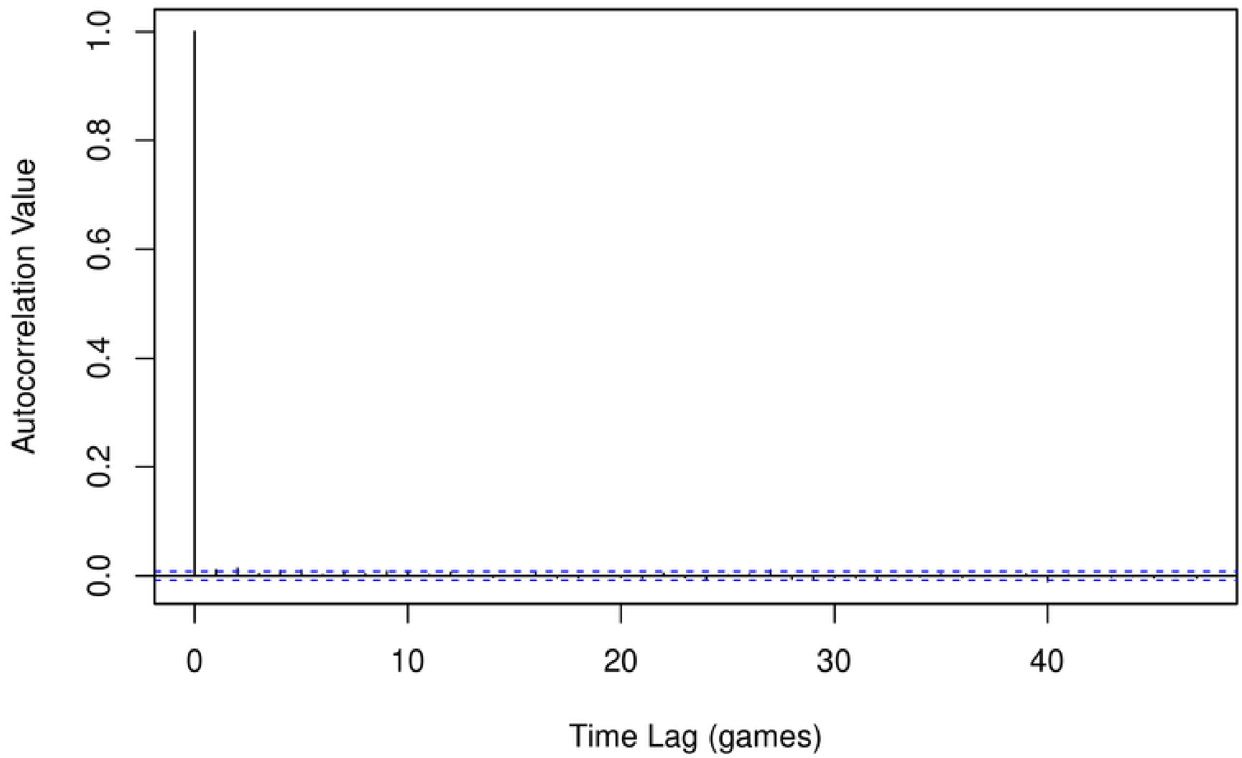

Figure 3. Autocorrelations of Hikaru excess scores. The autocorrelation function (ACF) between the excess scores (actual minus expected) in Hikaru's games, showing virtually no autocorrelation.

Figure 3 indicates that, to our surprise, the autocorrelations at all positive lags are all extremely close to zero. This indicates that there is virtually no correlation between Hikaru's excess scores on successive games. That is, for the excess scores in these games at least, there is no overall evidence of a hot hand effect. As a result, conditional on the player ratings and colors, the game outcomes can be reasonably treated as being conditionally independent.

We checked that this lack of autocorrelation of successive excess scores also holds for other players in the data. And Kramnik (2024a) and others agree with this conclusion. Nevertheless, it may seem surprising. Chess players are often thought to experience 'tilt,' whereby they get frustrated after one loss, which makes them more likely to lose again. Plus, players sometimes repeatedly play the same opponent, with various psychological and practical implications.

One possible explanation for the observed lack of autocorrelation could be that these psychology-related factors are generally not as large as previously thought. And, small correlations in certain specific games only might not significantly affect the overall value.

In addition, these excess scores are calculated by subtracting off the expected scores, which are based on the players' current ratings and hence updated after each game. They thus already take into account the choice of opponent plus some effects of previous games, and hence might counteract the correlations. That is, zero correlations do not mean that the game outcomes are independent, just that they are *conditionally* independent given the updated player ratings, which might themselves already account for various dependencies. Indeed, Hikaru's actual game scores (as opposed to excess scores) have autocorrelations that are more substantial, equal to nearly 0.1 for lags 1 through 5, indicating that the updated player ratings are a significant reason for uncorrelated excess scores.

Whatever the reason, the observed autocorrelations of excess scores are indeed extremely small. This indicates that the excess scores are approximately conditionally independent, conditional on the player ratings. So, we will use this property in our analysis below. As an aside, we note that if Hikaru's excess scores did actually have a positive correlation, then this would make his long winning streaks *more* likely, not less.

7. Identifying Winning Streaks

Next, we investigate winning streaks in the Hikaru game data.

Hikaru is recorded as playing a total of 57,421 games on Chess.com over the date range January 6, 2014, to to July 14, 2024. In this section, we combine all of these games together, in time order, to determine streaks. In later sections, we will also consider separating out the games played at specific time controls or dates.

To define a 'streak,' we have to decide how to handle draws. At the 'pure' extreme, we could define streaks to consist solely of wins, so that any draw or loss ends it. This is a reasonable definition, but it excludes such cases as the recent streak of 46 games in which he won 45 and tied one (GothamChess, 2024; McClain, 2023). At the other extreme, we could say that wins or draws both continue a streak, while only a loss ends it. This is a very loose definition, allowing many draws in a row to constitute a major 'streak,' which seems inappropriate.

So, as a compromise, since the most controversial of Hikaru's streaks involved just one draw, we shall use the 'in-between' definition that a streak consists of a maximal sequence of games with no losses and at most one draw. That is, a single draw continues a streak, but a second draw (or any loss) ends it.

Compared to pure winning streaks, this compromise streak definition is somewhat harder to work with. It also allows for *overlapping* streaks, which are not independent. For example, suppose a player had a first block of wins, then one draw, then a second block of wins, then another draw, then a third block of wins. In this case, the first two blocks of wins plus the one draw between them would count as a streak. And, the second and third blocks of wins plus the one draw between them would also count as a streak. There would thus be two overlapping streaks. Nevertheless, despite these minor challenges, we decided that this compromise definition is best under the circumstances, so we use it in our analysis below. We note that we did also separately look at Hikaru's 'pure' winning streaks, and confirmed that the overall conclusions are similar in that case too.

With this compromise streak definition, Hikaru has a total of 8,069 streaks (including some overlapping ones). Now, most of these are very short ‘streaks’; indeed 1,302 of them consist of just a single game. However, quite a few of them are reasonably large. Indeed, 226 of them are at least 30 games, and the largest are of lengths 121, 114, 107, 103, and 101. Next, we have to decide how to interpret and analyze such streaks.

8. Raw Probabilities of Winning Streaks

Even if a streak is very long, this does not necessarily mean that it is unexpected. We also need to consider the *probability* of each streak. We define the ‘raw probability’ of a streak as follows. If the streak consisted of winning all of the games, then its raw probability is the probability of a player winning all of those games, given the observed rating differences and color assignments. If the streak consisted of winning all but one game and tying the other, then its raw probability is the probability of a player either winning all of those games, winning all but (any) one of those games and tying the other, again given the observed rating differences and color assignments. This raw probability thus depends on the individual game win and draw probabilities, which are computed from the rating differences in that game using the fit formulae Equation 4.1 and Equation 5.1 (or their modifications in later sections), together with the conditional independence property described in Section 6.

With this definition of raw probability, it turns out that even some very long streaks have raw probabilities that are not particularly low. For example, Hikaru’s streak of length 121 began with his game number 20,940, and took place on December 22, 2018 (except for the final game). Over these 121 games, Hikaru had an average rating of 3016. But his *opponents* had an average rating of just 1,579. This means that Hikaru had an average rating advantage of +1,437. This is a *tremendous* advantage, corresponding to a win probability of 99.985%, that is, nearly certain. When computed over the course of all 121 games, his raw probability of scoring at least 120.5 on those 121 games then works out to 12.8%, or about one chance in 7.8, that is, not very low. We will discuss Hikaru’s opponents’ low ratings further below.

Since Hikaru has so many streaks that are short or have fairly high raw probability, we need to narrow down to those streaks that are striking in some sense. Here we focus on streaks that have a minimum length of 25 games—a reduction from the minimum 30 games used in Rosenthal (2024a), as requested in Kramnik (2024b). Among those streaks, we focus on those with small raw probabilities. In particular, there are 21 such streaks with raw probability less than 1/200 (including a few overlapping ones), which are presented in Table 1.

Table 1. Hikaru Streaks (all games, length ≥25, raw prob <1/200. A list of all 21 of Hikaru’s streaks (winning all, or all but one and drawing one) of at least 25 games with raw probability <1/200, in chronological order. The columns show the streak number, the date on which the streak ended, the streak’s starting game and length, the score Hikaru achieved in that streak, his expected score according to Equation 4.1, and the raw probability of each streak as defined herein.

line	streak	end date	start	length	score	expect	raw prob
1	589	March 6, 2016	7,027	54	54	48.9	1/1093.4
2	1,384	May 28, 2017	13,170	32	32	27.9	1/289.2
3	1,717	October 20, 2017	15,503	91	90.5	85.9	1/251.8
4	2,154	July 1, 2018	18,182	37	36.5	32.1	1/258.1
5	2,155	July 1, 2018	18,184	44	43.5	38.3	1/554.3
6	2,414	November 12, 2018	19,665	40	39.5	34.2	1/705.6
7	2,415	November 12, 2018	19,666	57	56.5	49.6	1/5247.9
8	2,527	November 28, 2018	20,436	25	25	20.9	1/319.0
9	3,569	November 15, 2019	27,244	32	31.5	27.3	1/201.7
10	3,734	January 25, 2020	28,227	30	29.5	24.1	1/907.4
11	3,805	February 16, 2020	28,644	25	25	21.0	1/242.5
12	3,917	March 12, 2020	29,340	41	40.5	33.6	1/6942.8
13	3,918	March 12, 2020	29,350	32	31.5	27.2	1/210.9
14	4,465	May 31, 2020	32,790	61	61	55.3	1/3058.4
15	4,551	June 30, 2020	33,483	53	52.5	47.4	1/524.5
16	5,029	December 20, 2020	36,631	26	25.5	20.8	1/505.6

17	5,030	December 20, 2020	36,633	26	25.5	20.6	1/614.2
18	6,519	January 9, 2023	45,881	26	25.5	19.9	1/2437.1
19	7,388	November 17, 2023	51,857	46	45.5	40.4	1/520.9
20	7,471	December 15, 2023	52,582	34	33.5	29.2	1/202.2
21	7,770	March 12, 2024	55,162	35	34.5	29.7	1/396.0

Table 1, line 19, corresponds to the most controversial streak, mentioned earlier, of length 46, ending on date November 17, 2023. As can be seen from the last column, this streak has raw probability nearly one chance in 500, which is small but not extremely small.

Table 1 also includes many other streaks, some of which have considerably smaller raw probabilities. For example, the streak on line 12, consisting of 41 games ending on March 12, 2020, has the smallest raw probability, about one chance in 6,943. Over those 41 games, Hikaru's opponents had a quite high average rating of 3,008. But Hikaru had an average rating of 3,261, and hence still an average rating advantage of +253. Now, a raw probability of one chance in 6,943 is fairly small. So, is it surprising to see such a streak? We consider this question theoretically in the next section, and then later through Monte Carlo simulations.

9. Theoretical Bound on Smallest Streak Probability

We have seen from Table 1 line 12 that Hikaru's streak of 41 games ending on March 12, 2020, had the smallest raw probability, of about one chance in 6,943.

Now, out of 57,421 games total, since 57,421 / 41 $\doteq$ 1,400.514, there are still 1,400 different independent opportunities to establish a streak of length 41, even without considering overlapping possibilities. So, as a first approximation lower bound, suppose there are 1,400 independent opportunities to establish a streak, each of independent probability 1/6,943. Then the probability of achieving such a streak would be given by

$$1 - \left[1 - \frac{1}{6,943}\right]^{1,400} \doteq 0.1826\,.$$

That is, under this approximation, the probability of achieving such a streak over the course of 57,421 games is roughly 18%, even without considering overlapping opportunities. That is not particularly surprising, and well above the usual 5% level for statistical significance.

So, even with a simple approximate theoretical lower bound, we can already see that Hikaru's individual streaks with small raw probabilities are not particularly surprising. However, much of the concern about Hikaru's streaks (e.g., Kramnik, 2024b) involves looking at *multiple* streaks together. This is too difficult to analyze theoretically, but can be done by Monte Carlo simulations, as we do next.

10. Monte Carlo Simulations of Streaks

The above calculation indicates that Hikaru's individual least-likely streak is not particularly surprising. However, the approximate 18% probability computed above is just a lower bound, which does not take into account the additional possibilities of long streaks in game sequences that are overlapping and hence not independent. Furthermore, they do not look at multiple win streaks beyond the single least-likely one, for which theoretical probabilities are too difficult to compute.

To analyze this further, we instead conduct a Monte Carlo (random) simulation (e.g., Robert & Casella, 2009). Specifically, using the actual player ratings and colors for each of Hikaru's 57,421 games, we simulated fresh game results using the probabilities of wins and ties implied by Equation 4.1 and Equation 5.1. We repeated this simulation 10,000 different times. Each time, we recorded the 21 streaks with smallest raw probability, in order, and checked whether their raw probability was smaller than the corresponding ordered raw probability from the actual data. The results are shown in Table 2.

Table 2. Hikaru Monte Carlo Percentages, all games. **A list of the streaks from Table 1, now ordered by raw probability, plus a final column showing what percentage of the 10,000 Monte Carlo runs using S_1 and D_1 had a smaller raw probability for that corresponding streak with that ordering.**

line	streak	end date	startgame	length	score	expect	raw prob	% less
1	3,917	March 12, 2020	29,340	41	40.5	33.6	1/6942.8	45.0%
2	2,415	November 12, 2018	19,666	57	56.5	49.6	1/5247.9	20.1%
3	4,465	May 31, 2020	32,790	61	61	55.3	1/3058.4	18.3%
4	6,519	January 9, 2023	45,881	26	25.5	19.9	1/2437.1	12.3%

5	589	April 6, 2016	7,027	54	54	48.9	1/1093.4	39.8%
6	3,734	January 25, 2020	28,227	30	29.5	24.1	1/907.4	38.2%
7	2,414	November 12, 2018	19,665	40	39.5	34.2	1/705.6	45.4%
8	5,030	December 20, 2020	36,633	26	25.5	20.6	1/614.2	44.9%
9	2,155	July 1, 2018	18,184	44	43.5	38.3	1/554.3	42.3%
10	4,551	June 30, 2020	33,483	53	52.5	47.4	1/524.5	35.6%
11	7,388	November 17, 2023	51,857	46	45.5	40.4	1/520.9	25.9%
12	5,029	December 20, 2020	36,631	36	25.5	20.8	1/505.6	19.7%
13	7,770	March 12, 2024	55,162	35	34.5	29.7	1/396.0	35.2%
14	2,527	November 28, 2018	20,436	25	25	20.9	1/319.0	52.4%
15	1,384	May 28, 2017	13,170	32	32	27.9	1/289.2	56.0%
16	2,154	July 1, 2018	18,182	37	36.5	32.1	1/258.1	62.6%
17	1,717	October 20, 2017	15,503	91	90.5	85.9	1/251.8	57.3%
18	3,805	February 16, 2020	28,644	25	25	21.0	1/242.5	54.0%
19	3,918	March 12, 2020	29,350	32	31.5	27.2	1/210.9	66.3%
20	7,471	December 16, 2023	52,582	34	33.5	29.2	1/202.2	64.4%

21	3,569	November 15, 2019	27,244	32	31.5	27.3	1/201.7	57.5%

The top line of Table 2 shows the streak with smallest raw probability 1/6,943 discussed above. Its last column indicates that, over the course of 10,000 separate Monte Carlo simulations, 45.0% of the simulations produced a streak with smaller (or equal) raw probability. As expected, this is considerably larger than the 18% theoretical lower bound computed earlier. It shows even more clearly that a streak of such small raw probability is not unexpected over the course of Hikaru's full record.

The second line of Table 2 shows the streak with second-smallest raw probability 1/5,248. In this case, the Monte Carlo simulation showed that 20.1% of the simulations had second-smallest streak raw probability less than (or equal to) 1/5,248. That is, about 20.1% of the simulations had two streaks whose raw probabilities were both less than 1/5,248. This is somewhat less than the 45.0% from line 1, indicating that Hikaru's two smallest raw probability streaks combined are somewhat less likely than just his smallest one. Nevertheless, they are not particularly surprising, and still well above the usual 5% level for statistical significance.

The third and fourth lines of Table 2 give the percentage of Monte Carlo simulations whose third or fourth smallest streak raw probability is less than (or equal to) that of Hikaru. These percentages get somewhat smaller, down to 18.3% and 12.3%. This indicates that, in Hikaru's record, the single most surprising streak fact is that he had four streaks that all had raw probability ≤1/2,437. This would only occur about 12.3% of the time. However, even this is well above the usual 5% threshold.

The rest of the last column of Table 2 shows that the remaining streaks each have a Monte Carlo percentage that is higher than 12.3%, usually considerably so. The only other percentage that is remotely small is for the 12th-smallest streak raw probability, which had a Monte Carlo percentage 19.7%, still not particularly low. Of course, even if one of the 21 streaks had a Monte Carlo percentage slightly lower than 5%, this would not necessarily indicate statistical significance due to issues of multiple testing (Ranganathan et al., 2016). But in fact, none of them even come close to the 5% threshold. So, this extensive Monte Carlo simulation indicates that none of Hikaru's streaks, even when taken as a group, are particularly surprising or unexpected given his and his opponents' chess ratings.

11. Restricting to Just a Single Time Control

One important aspect of any chess game is the *time control*, that is, how much time each player is allotted to play the game. While traditional chess tournaments allotted several hours to each player, the modern online trend is toward much faster games. Indeed, the large majority of Hikaru's recorded games used the time control 3m+0s, meaning that each player gets 3 minutes for the entire game, with no bonus increment for completed moves. Of his 57,421 total games, there were 35,449 games (61.7%) at this time control. Second most common was the time control 1m + 0s, where each player gets just one minute for the entire game, again with no bonus increment, corresponding to 15,569 of his games (27.1%). Third most common was the time control 3m+1s, in which each player gets 3 minutes plus a one-second bonus for each completed move, at 3,310 games (5.8%).

Now, our above analysis combined all of Hikaru's different games together, in chronological order, regardless of time control. This may be reasonable, since any cheating and so on might be expected to continue during different types of games. However, in the response Kramnik (2024a) vigorously objected to this, since different time controls substantially change the nature of the game. Indeed, Chess.com even uses different chess ratings for 'blitz' games (total time between 3 and 14 minutes each) and 'bullet' games (total time less than 3 minutes each). So, in this section, we redo the entire curve fit and Monte Carlo analysis again, restricted solely to games with most common 3m+0s time control.

Of the 293,047 total chess games in our data, 131,445 of them (44.9%) are at the time control 3m+0s. For White's expected score, redoing the least squares analysis on just this 3m+0s data, we find that the sum of squares is minimized numerically when c_1 = 21 and c_2 = 1/356. The resulting expected score curve is thus:

$$S_2 \;=\; \frac{1}{1+10^{-(A-B+21)/356}}\,, \tag{11.1}$$

quite similar to the fit S_1 in Equation 4.1 but with a somewhat smaller exponent divisor, which fits the 3m+0s data well (see Figure 4).

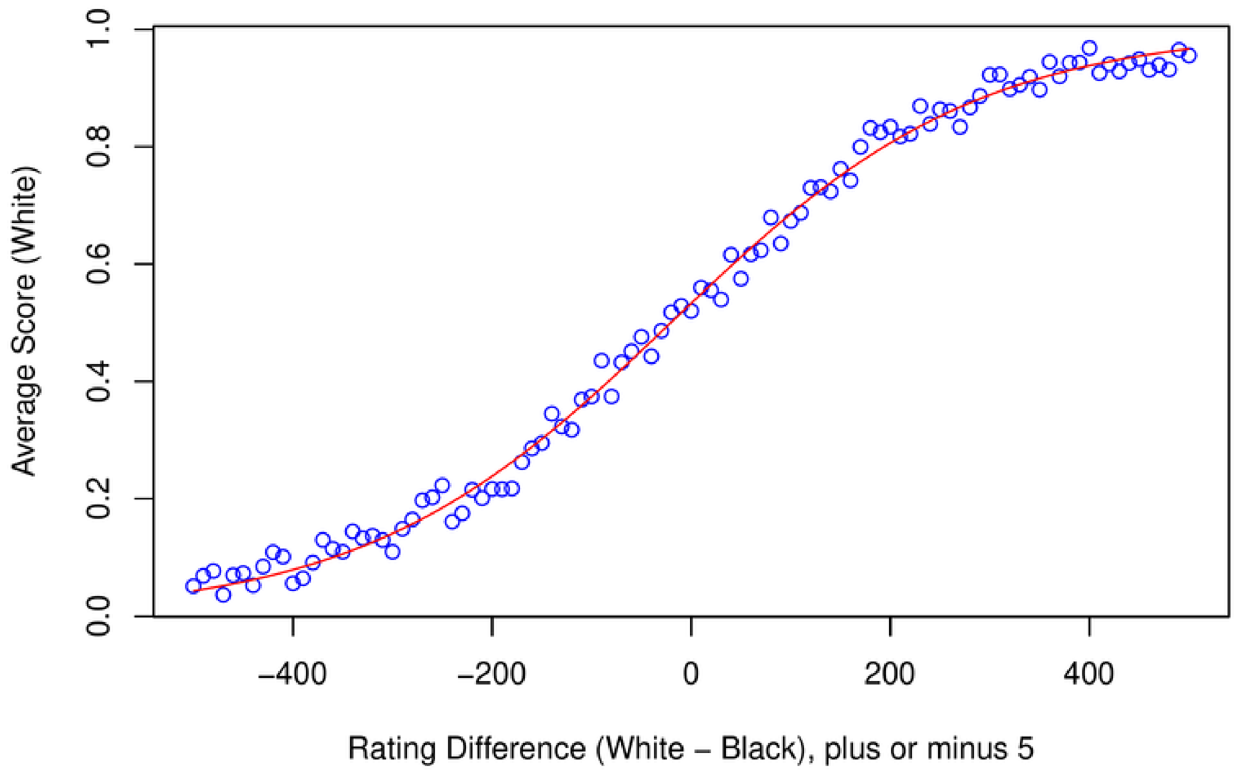

Figure 4. Average scores versus expected score, 3m+0s only. The average score for White in games at time control 3m+0s only, as a function of the binned rating difference (blue), together with the score fit function S_2 from Equation 11.1 (red).

And, for the probability of a draw, the sum of squares is minimized when $c_3 = 0.138$, $c_4 = 13$, and $c_5 = 1/373$, leading to the draw probability curve

$$D_2 \;=\; (0.138)\, e^{-[(A-B+13)/373]^2} \qquad (11.2)$$

for the probability of a draw game, a small adjustment of D_1 in Equation 5.1, which fits the 3m+0s data reasonably well (see Figure 5).

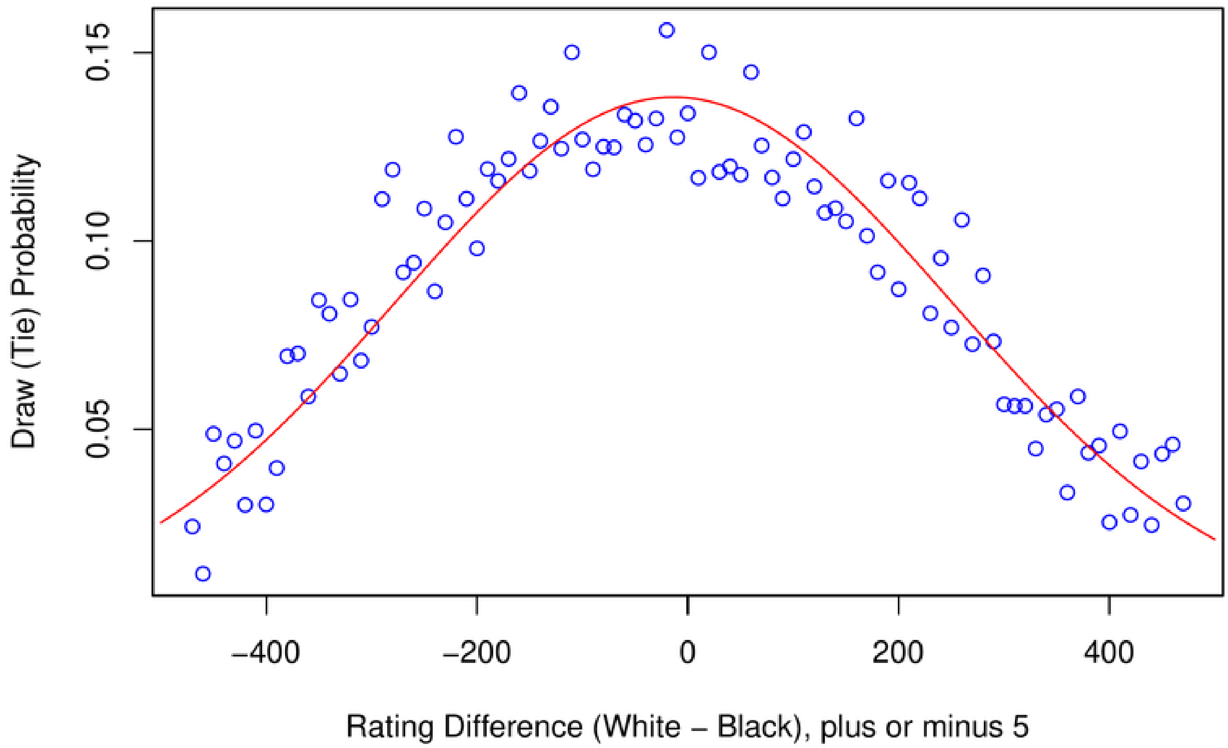

Figure 5. The draw fraction in all games at time control 3m+0s as a function of binned rating difference (blue), with the draw fit function D_2 from Equation 11.2 (red).

When restricting to just Hikaru's 35,449 games at time control 3m+0s, and using the new fits S_2 and D_2, we find that he has just seven streaks with raw probability less than 1/200, and a total of 16 streaks with raw probability less than 1/100, shown in Table 3.

Table 3. Hikaru Streaks (3m+0s only, length ≥25, raw prob <1/200). A table similar to Table 1, but now based on just Hikaru's games at the specific time control 3m+0s, and with raw probability $< 1/100$.

line	streak	end date	start	length	score	expect	raw prob
1	8	May 14, 2014	45	46	46	41.5	1/590.6
2	539	May 28, 2017	4,514	32	32	28.4	1/167.3
3	577	June 26, 2017	4,747	25	25	21.7	1/107.6
4	1,188	July 1, 2018	8,755	37	36.5	32.6	1/138.0
5	1,189	July 1, 2018	8,757	44	43.5	39.0	1/253.0

6	2,274	November 15, 2019	15,733	32	31.5	27.8	1/119.7
7	2,400	January 25, 2020	16,508	52	51.5	47.1	1/306.4
8	2,462	February 16, 2020	16,906	25	25	21.5	1/152.3
9	2,557	March 12, 2020	17,551	41	40.5	34.2	1/3380.1
10	2,558	March 12, 2020	17,561	33	32.5	28.6	1/127.8
11	3,063	May 31, 2020	20,723	61	61	55.9	1/1697.6
12	4,275	January 10, 2023	29,512	29	28.5	24.4	1/183.3
13	4,553	July 3, 2023	31,454	33	32.5	28.6	1/129.3
14	4,575	July 10, 2023	31,607	47	46.5	42.2	1/207.8
15	4,675	November 17, 2023	32,499	46	45.5	41.1	1/231.0
16	4,721	December 17, 2023	32,904	36	35.5	31.6	1/125.0

Comparing Table 3 with Table 1, we see that several of the previous streaks remain, including the least likely one (now line 9), and the most controversial one (now line 15), though some other ones disappear or are modified, and the raw probabilities are all changed due to the new fits.

To examine the significance of these streaks at just time control 3m+0s, we conducted a fresh Monte Carlo simulation of 10,000 random simulations of just these 35,449 games, using the new fits S_2 and D_2 from Equation 11.1 and Equation 11.2. Specifically, using the actual player ratings and colors for each of Hikaru's 35,449 games at 3m+0s time control, we simulated fresh game results using the probabilities of wins and draws from Equation 11.1 and Equation 11.2. We repeated this simulation 10,000 different times. Each time, we recorded the 16 streaks with smallest raw probability, in order, and checked whether their raw probability was smaller than (or equal to) the corresponding raw probability from the actual data. The results are shown in Table 4.

Table 4. Hikaru Monte Carlo Percentages, 3m+0s only. A list of the streaks from Table 3, now ordered by raw probability, plus a final column showing what percentage of the 10,000 Monte Carlo runs using S_2 and D_2 had a smaller raw probability for that corresponding streak with that ordering.

line	streak	end date	startgame	length	score	expect	raw prob	% less
1	2,557	March 12, 2020	17,551	41	40.5	34.2	1/3380.1	46.2%
2	3,063	May 31, 2020	20,723	61	61	55.9	1/1697.6	39.9%
3	8	May 14, 2014	45	46	46	41.5	1/590.6	77.2%
4	2,400	January 25, 2020	16,508	52	51.5	47.1	1/306.4	95.4%
5	1,189	July 1, 2018	8,757	44	43.5	39.0	1/253.0	96.0%
6	4,675	November 17, 2023	32,499	46	45.5	41.1	1/231.0	94.5%
7	4,575	July 10, 2023	31,607	47	46.5	42.2	1/207.8	94.0%
8	4,275	January 10, 2023	29,512	29	28.5	24.4	1/183.3	94.5%
9	539	May 28, 2017	4,514	32	32	28.4	1/167.3	94.2%
10	2,462	February 16, 2020	16,906	25	25	21.5	1/152.3	94.4%
11	1,188	July 1, 2018	8,755	37	36.5	32.6	1/138.0	95.1%
12	4,553	July 3, 2023	31,454	33	32.5	28.6	1/129.3	94.9%
13	2,558	March 12, 2020	1,7561	33	32.5	28.6	1/127.8	92.4%
14	4,721	December 17, 2023	32,904	36	35.5	31.6	1/125.0	89.2%

15	2,274	November 15, 2019	15,733	32	31.5	27.8	1/119.7	87.5%
16	577	June 26, 2017	4,747	25	25	21.7	1/107.6	90.8%

Examining the first line of Table 4, we see that 46.2% of the new Monte Carlo simulations had a streak with raw probability less than (or equal to) that of Hikaru's smallest raw probability. This is very similar to (and slightly more than) the 45.0% Monte Carlo probability in the first line of Table 2.

Then, examining the subsequent lines of Table 4, we see that their Monte Carlo probabilities are actually *larger* than those in Table 2. Indeed, the smallest Monte Carlo percentage, in line 2 of Table 4, is still nearly 40%. The others are all larger, mostly considerably so, with many of them above 90%.

The Monte Carlo simulation of Table 4 thus indicates that, when restricted to just 3m+0s games with appropriately adjusted curve fits, Hikaru's streaks are still not at all unexpected over his full history. Restricting to just 3m+0s time control games makes small changes to the expected score and draw probability formulae, but does not show any additional evidence of surprising win streaks.

12. Fitting the Curves While Omitting One Player

The response Kramnik (2024b) argued that, in addition to restricting to just the single time control 3m+0s, the curve fits should also be done *excluding* games played by Hikaru, so that his potentially irregular play does not affect the assessment of probabilities. So, we now consider what would change if we restrict our fits solely to games with 3m+0s time control that do *not* involve Hikaru.

The data includes 92,872 games at 3m+0s time control not involving Hikaru. For this data, again using the principle of least squares, we determined that the expected score sum of squares is minimized when $c_1 = 21$ and $c_2 = 1/348$, giving the expected score function

$$S_3 \;=\; \frac{1}{1+10^{-(A-B+21)/348}}\,. \qquad (12.1)$$

This is very similar to the fit S_2 in Equation 11.1, and fits this new data well (see Figure 6).

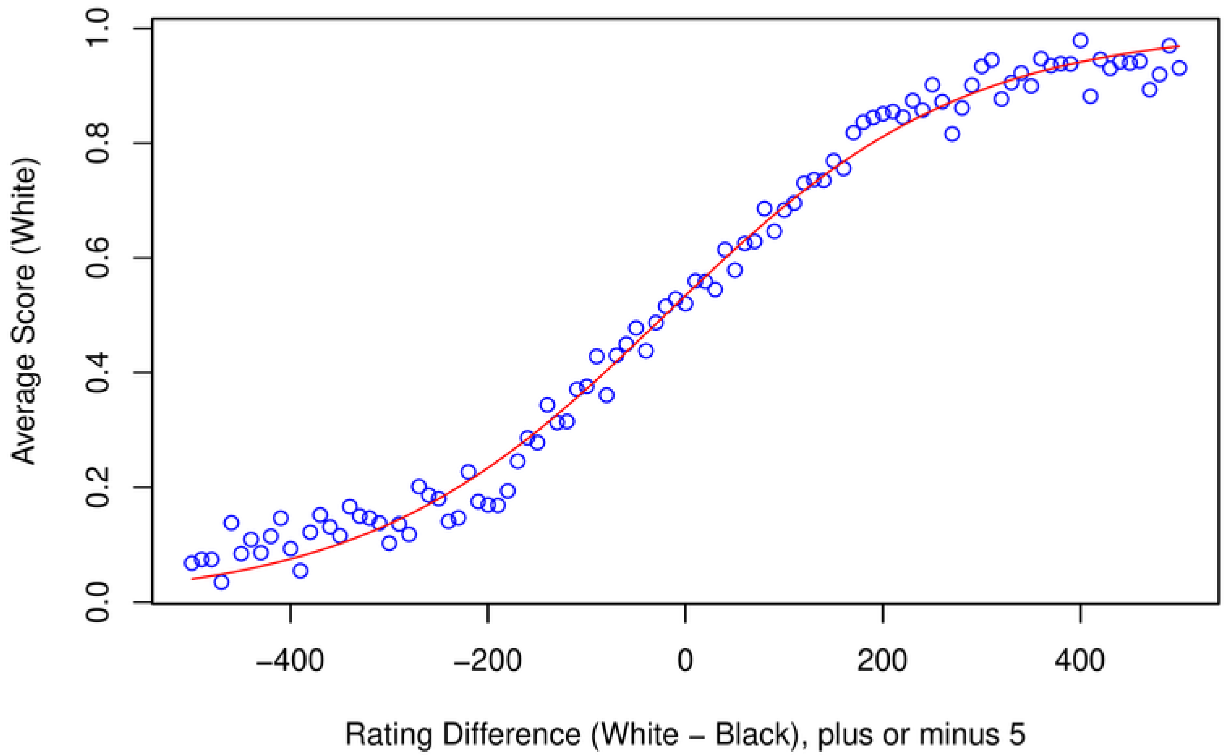

Figure 6. Average scores versus expected score, 3m+0s excl. Hikaru. The average score for White in games at time control 3m+0s only, excluding games involving Hikaru, as a function of the binned rating difference (blue), together with the score fit function S_3 from Equation 12.1 (red).

And, for the probability of a draw, the sum of squares is minimized when c_3 = 0.126, c_4 = 15, and c_5 = 1/418, leading to the draw probability curve

$$D_3 \;=\; (0.126)\, e^{-[(A-B+15)/418]^2}\,, \qquad (12.2)$$

an adjustment to Equation 11.2, which fits the new data reasonably well (see Figure 7).

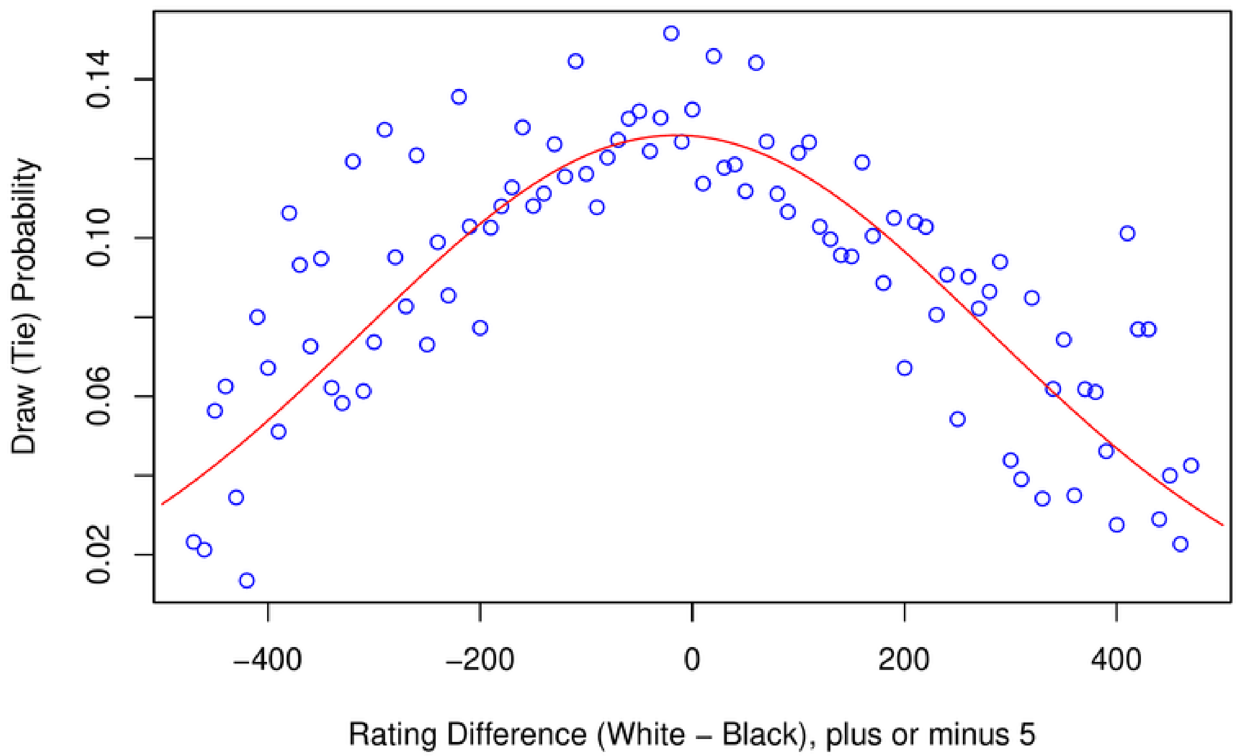

Figure 7. Probability of a draw (tie), 3m+0s excl. Hikaru. The draw fraction in all games at time control 3m+0s only, excluding games involving Hikaru, as a function of the binned rating difference (blue), together with the draw fit function D_3 from Equation 12.2 (red).

To proceed, we again used the actual player ratings and colors for each of Hikaru's 35,449 games at 3m+0s time control. We simulated fresh game results, now using the formulae S_3 and D_3 from Equation 12.1 and Equation 12.2. We repeated this simulation 10,000 different times. Each time, we again recorded the 15 streaks with smallest raw probability, in order, and checked whether their raw probability was smaller than (or equal to) the corresponding raw probability from the actual data. The results are shown in Table 5.

Table 5. Hikaru Monte Carlo Percentages, 3m+0s only, using S_3 and D_3. A list of the streaks from Table 3 plus Monte Carlo percentages, similar to Table 4, but now instead using the fits S_3 and D_3.

line	streak	end date	startgame	length	score	expect	raw prob	%less
1	2,557	March 12, 2020	17,551	41	40.5	34.4	1/2662.3	54.8%
2	3,063	May 31, 2020	20,723	61	61	56	1/1410.0	48.2%

3	8	May 14, 2014	45	46	46	41.7	1/519.9	82.5%
4	2,400	January 25, 2020	16,508	52	51.5	47.2	1/258.2	97.9%
5	1,189	July 1, 2018	8,757	44	43.5	39.2	1/209.8	98.6%
6	4,675	November 17, 2023	32,499	46	45.5	41.3	1/192.1	98.2%
7	4,575	July 10, 2023	31,607	47	46.5	42.4	1/173.6	98.0%
8	4,275	January 10, 2023	29,512	29	28.5	24.5	1/153.5	98.2%
9	539	May 28, 2017	4,514	32	32	28.5	1/149.1	97.0%
10	2,462	February 16, 2020	16,906	23	25	21.6	1/132.9	97.9%
11	1,188	July 1, 2018	8,755	37	36.5	32.8	1/118.4	98.5%
12	4,553	July 3, 2023	31,454	33	32.5	28.8	1/109.3	98.6%
13	2,558	March 12, 2020	17,561	33	32.5	28.8	1/108.6	97.6%
14	4,721	December 17, 2023	32,904	36	35.5	31.8	1/106.5	96.5%
15	2,274	November 15, 2019	15,733	32	31.5	27.9	1/104.3	95.2%

Comparing Table 5 with Table 4, we see that they are very similar, and in fact the Monte Carlo percentages are slightly *higher*. This means that Hikaru's streaks at time control 3m+0s are actually slightly *less* surprising when using the fits S_3 and D_3 than when using the fits S_2 and D_2. In any case, the conclusion is the same: once again, by this new measure, Hikaru's game record does not show evidence of surprising winning streaks.

13. Examination of Two Specific Days

In the critique, Kramnik (2024b) mentions a number of specific recent Hikaru 'pure' winning streaks, that is, streaks consisting of wins only with no draws, and wonders why they were not considered particularly surprising in Rosenthal (2024a, 2024b). Several of these involve Hikaru's games on February 16–17, 2024, so we take a closer look at those two specific dates now.

Hikaru played a total of 230 games on those two days, a huge amount, in his games numbered 54665 through 54894. Of these, 129 were at time control 3m+0s, and 101 were at 1m + 0s. We consider separately the cases of grouping all 230 games together, and looking at just the 3m+0s time control games separately as requested in Kramnik (2024a, 2024b).

For the 230 games all grouped together, the longest pure winning streak (by far) consisted of 59 games, numbered 54762 through 54820, spanning both days. For these 59 games, Hikaru's mean rating was 3,317, while his opponents' mean rating was 2,793, giving the very large average rating advantage of +524. Using the fit S_1 and D_1 from Equation 4.1 and Equation 5.1 to combine all time controls, his raw probability of winning all 59 games works out to 1/158.

For the 129 of these games that are at time control 3m+0s, there are two long pure winning streaks, each involving 'skipping over' some 1m + 0s games with losses or draws. The first is the 34 games from 54685 through 54698 and from 54719 through 54738 on February 16, skipping over games 54699 through 54718 at 1m + 0s. The second is the 36 games from 54762 through 54787 and from 54869 through 54878 spanning both days, skipping over games 54788 through 54868 at 1m + 0s.

For the 34-game streak, Hikaru's mean rating was 3,276, while his opponents' mean rating was 2,882, an average rating advantage of +394. And, using the fit S_3 and D_3 from Equation 12.1 and Equation 12.2 based on just time control 3m+0s excluding Hikaru games, his raw probability of winning all 34 of these games works out to 1/71.8.

For the 36-game streak, Hikaru's mean rating was 3,265, while his opponents' mean rating was 2,831, an average rating advantage of +434. And, again using the fit S_3 and D_3 from Equation 12.1 and Equation 12.2, his raw probability of winning all 36 games works out to 1/19.7.

In summary, although Hikaru's long streaks on February 16–17, 2024, do seem striking at first glance, in fact they were against far weaker opponents and hence have raw probabilities that are not particularly small or surprising.

14. Streaks in the Year 2024 Only

The critique of Kramnik (2024b) expressed concern with various Hikaru streaks in the year 2024, not just the two discussed in the previous section. So, as a final test, we consider Hikaru's games at time control 3m+0s during just the year 2024, specifically from January 1, 2024 through July 14, 2024 (when our data ends), and look at all of his streaks within that time period.

For this data, again using the fits S_3 and D_3, Hikaru had 10 streaks with raw probability < 1/10. So, using the actual player ratings and colors for each of Hikaru's 2,312 games at 3m+0s in the year 2024, we simulated fresh game results, again using the fits S_3 and D_3 from Equation 12.1 and Equation 12.2. We repeated this simulation 10,000 different times. Each time, we recorded the 10 streaks with smallest raw probability, in order, and checked whether their raw probability was smaller than the corresponding raw probability from the actual data. The results are shown in Table 6.

Table 6. Hikaru Monte Carlo Percentages, 3m+0s, 2024 only. **Hikaru's streaks for 3m+0s games in the year 2024 only, ordered by raw probability, plus a final column showing what percentage of 10,000 Monte Carlo runs using S_3 and D_3 had a smaller raw probability for that corresponding streak.**

line	streak	end date	startgame	length	score	expect	raw prob	% less
1	112	February 16, 2024	1077	37	36.5	33.1	1/75.8	84.4%
2	139	March 19, 2024	1318	33	32.5	29.2	1/69.4	61.8%
3	57	January 20, 2024	425	49	49	46.3	1/53.3	51.9%
4	67	January 26, 2024	553	37	37	34.8	1/23.5	83.2%
5	117	February 17, 2024	1137	36	36	34	1/19.7	80.7%
6	188	May 14, 2024	1753	31	30.5	28.1	1/19.6	67.2%
7	130	March 3, 2024	1253	40	39.5	37.2	1/18.7	56.0%
8	187	May 11, 2024	1746	33	32.5	30.2	1/18.3	42.9%
9	142	March 27, 2024	1368	31	30.5	28.3	1/15.8	41.9%

10	49	January 15, 2024	339	26	25.5	23.5	1/13.2	44.3%

We see that Table 6, line 1, is a slight extension—due to allowing one draw instead of just pure streaks of wins only—of the streak of length 34 on February 16, 2024, discussed in the previous section. And Table 6, line 5, is the same pure streak of length 36 on February 16–17, 2024, discussed in the previous section.

In any case, the last column of Table 6 indicates that the Monte Carlo percentages of smaller corresponding raw probabilities are all above 40% and hence not particularly small. That is, although Hikaru did have a number of fairly long streaks in the year 2024, observing streaks with those raw probabilities was not particularly surprising. So, when considering just the year 2024, the conclusion is still the same: Hikaru's game record again does not show evidence of unexpected win streaks.

15. Comparing Hikaru and Carlsen Opponents

The critique of Kramnik (2024b) raised the issue of why Hikaru has so many more long winning streaks than top player Magnus Carlsen. Indeed, we have seen that Hikaru had 21 streaks (as defined herein) with raw probability <1/200, but meanwhile Carlsen had just one—a huge difference. Furthermore, Hikaru had several streaks of more than 100 games, while Carlsen's longest one was just 32 games—another huge difference.

So, by any measure, Hikaru had far more win streaks than Carlsen. That raises the question, does this provide evidence of irregularities? Or can it be explained in a way that is consistent with statistical expectations?

A first explanation is that Hikaru has played a total of 57,421 games on Chess.com, compared to just 5,104 for Carlsen. So, we would expect Hikaru to have more and longer streaks just by virtue of having more games to choose from. Okay, but what else?

Further insights can be gained by looking at their opponents' ratings. Hikaru's rating averaged 3,130 over his entire record, while Carlsen averaged 3,227, that is, even higher. However, Hikaru's *opponents'* mean rating was just 2,730, giving him a mean rating advantage of +400. By contrast, Carlsen's opponents' mean rating was 2,985, giving him a mean rating advantage of +242, much lower. Even more dramatically, as summarized in Table 7, Hikaru played a total of 699 games (1.2% of his total) against players with ratings ≤1,000, that is, extremely weak opponents, while Carlsen has played just eight (0.16%). And Hikaru has played 3,286 (5.7%) against players ≤2,000, compared to just 30 (0.59%) for Carlsen. And 8,048 (14.0%) against players ≤2,500, compared to just 156 (3.1%) for Carlsen. In short, Hikaru's opponents have generally had much lower ratings than Carlsen's.

Table 7. Hikaru versus Carlsen. A comparison of the records of Hikaru Nakamura and Magnus Carlsen, showing their total number of games played, average rating advantage over their opponent, and number of opponents with rating ≤1,000, and ≤2,000, and ≤2,500.

Player	NumGames	avr adv	# ≤ 1000	# ≤ 2000	# ≤ 2500
Hikaru Nakamura	57,421	+400	699 (1.2%)	3286 (5.7%)	8048 (14.0%)
Magnus Carlsen	5,104	+242	8 (0.16%)	30 (0.59%)	156 (3.1%)

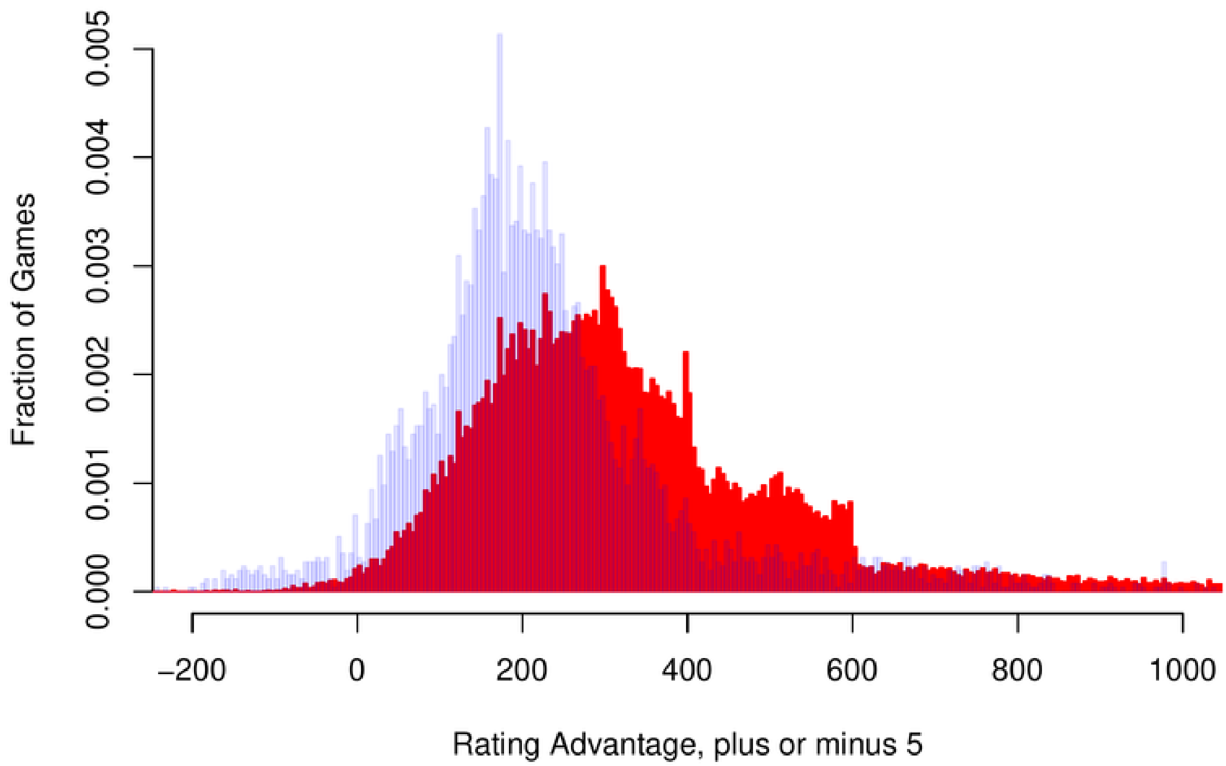

Figure 8. Histogram of rating advantages. Histograms of the rating advantages of Hikaru Nakamura (red) and Magnus Carlsen (blue), as fractions of all of their games played on Chess.com.

This can also be seen by looking at histograms of the two players' rating advantages over their opponents, in Figure 8. That graph shows that Hikaru's rating advantage (in red) spreads out from about 150 to 600, with tail extending beyond 1,000, while Carlsen's is concentrated between about 50 and 350. So, compared to Carlsen, Hikaru has played a lot more opponents with much lower ratings. This fact, in addition to simply playing many more games, provides a sound statistical explanation for why Hikaru has so many more winning streaks than Carlsen does.

16. Discussion

The statistical analysis presented here shows that Hikaru did indeed have many long winning streaks in his online chess play, including his recent controversial 46-game streak and many others. And some of them do have quite low raw probabilities. However, Monte Carlo simulations indicate that his streaks are well within the range of what would be expected statistically, given his rating advantages over his opponents. Hence, they are not particularly surprising, and do not provide any statistical evidence of irregularities.

Of course, every statistical analysis requires making some choices regarding definitions, scope, modeling, and so on, all of which can affect the results. However, contrary to the claims in Kramnik (2024a, 2024b), I believe that the choices in this report are all fair and defensible, consistent with the available data, and lead to accurate conclusions. Furthermore, various alternative choices, such as restricting to a single time control, excluding his own games when fitting the curves, or looking at just the most recent year, all lead to similar results. Hence, the overall conclusion, that the streaks observed in Hikaru's Chess.com record are fairly typical given the player ratings over Hikaru's long record of games, appears to be quite robust.

This conclusion is heavily influenced by Hikaru's rating advantages over his opponents. As discussed herein, Hikaru has played many games against opponents with far lower ratings, much more so than other top-level players such as Carlsen. This difference may be related to Hikaru's apparent practice of 'farming,' that is, intentionally seeking lower rated opponents (Nakamura, 2024a, 2024b). Now, it might be possible that Hikaru's chess rating is inflated for whatever reason, an issue that is not investigated here. However, conditional on Hikaru's ratings and his opponents' ratings and his extensive play record, the evidence indicates that his long winning streaks, though initially striking, are not unexpected from a statistical point of view.

It is important to note that the existence of unexpected streaks is distinct from the issue of cheating. For example, one player might improve their skill over a short period of time due to higher concentration and motivation and preparation, and thus perform much better than their rating would indicate. This might lead to many unexpected streaks, even without any cheating. Conversely, another player who cheats in all games, in a consistent and regular manner, might manage to obtain a very high chess rating. This would in turn make their winning streaks appear to be expected, so that a study of streaks would never arouse suspicion, despite their cheating ways. Or, even simpler, a player might decide to cheat only in every *second* game, leading to an unfairly successful record, but nevertheless creating no long win streaks since they would lose many of the intervening honest games. Hence, the examination presented herein should be viewed as merely investigating the presence of unexpected streaks, not the broader issue of cheating in online chess.

Finally, it was insinuated in the comments of Kramnik (2024a, 2024b) that I might be biased, that is, manipulating the statistical analysis in a deliberate effort to protect Hikaru and deny the existence of unexpected streaks. I am confident that this is not the case. As a professional academic, I always attempt to seek the truth in an unbiased way. Furthermore, as I learned from my work on the high-profile lottery retailer scandal (Rosenthal, 2014), finding clear evidence of cheating or other irregularities would have led to *more* excitement and attention and glory for me, not less. So, to the extent that I had any personal motivation in this analysis, I actually hoped to find some such evidence. However, despite my best efforts, I did not find any evidence of suspicious behavior in Hikaru's online chess winning streaks.

Acknowledgments

I thank Erik Allebest and Dan Rozovsky of Chess.com for providing me with the chess results data and answering my questions, while allowing me the freedom to analyze the data as I wished. I thank Daniel Rosenthal for several helpful conversations about these issues. I thank V. Kramnik for his critiques (2024a, 2024b) which, though not always fair in my opinion, did raise some legitimate issues that spurred me to improve my analysis. I thank the referees and editors for detailed comments which helped me to improve the paper.

Disclosure Statement

The author has no conflicts of interest to declare, and received no compensation from Chess.com for this work.

References

Bobyrev, K. (2023, November 29). Analyzing long win streaks in online chess. *Kirill Bobyrev.* https://kirillbobyrev.com/blog/analyzing-long-win-streaks/

Cazaux, J.-L., & Knowlton, R. (2017). *A world of chess: Its development and variations through centuries and civilizations*. McFarland & Company.

Chess.com. (2023, November 29). Regarding recent accusations. https://www.chess.com/blog/CHESScom/regarding-recent-accusations

Chess.com. (2024a). What do I need to know about Fair Play on Chess.com? https://support.chess.com/en/articles/8568369-what-do-i-need-to-know-about-fair-play-on-chess-com

Chess.com. (2024b, July 18). Fair Play report: Estimating cheating rates in titled Tuesday. https://www.chess.com/blog/FairPlay/chess-com-fair-play-report-estimating-cheating-rates-in-titled-tuesday

Chess.com. (2024c). How do ratings work on Chess.com? https://support.chess.com/en/articles/8566476-how-do-ratings-work-on-chess-com

Chess.com. (2025, January 7). Fair Play on Chess.com. https://www.chess.com/fair-play

Dekking, F. M., Kraaikamp, C., Lopuha, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how*. Springer. https://doi.org/10.1007/1-84628-168-7

Elo, A. E. (1978). *The rating of chess players, past and present* (1st ed.). Arco.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

Glickman, M. E. (n.d.). *The Glicko rating systems.* Retrieved January 15, 2025, from http://www.glicko.net/glicko/glicko.pdf

Glickman, M. E. (2022, March 22). *Example of the Glicko-2 system.* http://www.glicko.net/glicko/glicko2.pdf

Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(3), 377–394. https://doi.org/10.1111/1467-9876.00159

Glickman, M. E., & Jones, A. C. (1999). Rating the chess rating system. *CHANCE, 12*(2), 21–28.

GothamChess. (2024, July 6). *I confronted Kramnik* [Video]. YouTube. https://www.youtube.com/watch?v=yDeQ-CmeZaU

Keener, G. (2022, September 27). Chess is booming. *The New York Times*. https://www.nytimes.com/2022/06/17/crosswords/chess/chess-is-booming.html

Kramnik, V. (2024a, November 30). *Another research full of mistakes* [Video]. YouTube. https://www.youtube.com/watch?v=qEsUXqnuDa8

Kramnik, V. (2024b, September 29). *Let us talk, Professor* [Video]. YouTube. https://www.youtube.com/watch?v=lz2I_WEG8lY

Maharaj, S., Polson, N., & Sokolov, V. (2023). *Kramnik vs Nakamura or Bayes vs p-value.* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4648621

McClain, D. L. (2023, December 25). Elite chess players keep accusing each other of cheating. *The New York Times*. https://www.nytimes.com/2023/12/25/crosswords/chess-hikaru-vladmir-kramnik-cheating.html

Miller, J. S., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica, 86*(6), 2019–2047. https://doi.org/10.3982/ECTA14943

Nakamura, H. (2024a, September 2). *Prof. of Stats Reports | Kramnik: "Lies!!"* [Video]. YouTube. https://www.youtube.com/watch?v=bfn9gAEY0Ck

Nakamura, H. (2024b, November 13). *Farming 3000s* [Video]. YouTube. https://www.youtube.com/watch?v=SvqPGAwS2Wg

Szczepanek, A. (2024, January 18). *Coin toss streak calculator.* Omni Calculator. https://www.omnicalculator.com/statistics/coin-flip-streak

Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research, 7*(2), 106–107. https://doi.org/10.4103/2229-3485.179436

Robert, C., & Casella, G. (2009). *Introducing Monte Carlo methods with R* (1st ed.). Springer. https://doi.org/10.1007/978-1-4419-1576-4

Rosenthal, J. S. (2014). Statistics and the Ontario lottery retailer scandal. *CHANCE, 27*(1), 4–9. https://doi.org/10.1080/09332480.2014.890864

Rosenthal, J. S. (2024a, August 9). *Probabilities of streaks in online chess* (original version). http://probability.ca/jeff/ftpdir/chessstreaksorig.pdf

Rosenthal, J. S. (2024b, September 5). *Probabilities of streaks in online chess* (addendum version). http://probability.ca/jeff/ftpdir/chessstreaksadd.pdf

Ross, K. (2022). *An Introduction to Bayesian reasoning and methods*. Bookdown. https://bookdown.org/kevin_davisross/bayesian-reasoning-and-methods/

Svensen, T. J. (2024, September 7). *Professor finds Nakamura's winning streaks statistically normal, Kramnik dismisses it.* Chess.com. https://www.chess.com/news/view/nakamura-winning-streaks-statistically-normal-professor-says

Wired. (2022, February 21). *Statistician answers stats questions from twitter | Tech Support | WIRED* [Video]. YouTube. https://www.youtube.com/watch?v=QW3KRaz4aI4

Footnotes

1. Data available at: https://probability.ca/chessstreakdata/ ↩

References

- Bobyrev, K. (2023, November 29). Analyzing long win streaks in online chess. *Kirill Bobyrev.* https://kirillbobyrev.com/blog/analyzing-long-win-streaks/

  ↩
- Cazaux, J.-L., & Knowlton, R. (2017). *A world of chess: Its development and variations through centuries and civilizations*. McFarland & Company.

  ↩
- Chess.com. (2023, November 29). Regarding recent accusations. https://www.chess.com/blog/CHESScom/regarding-recent-accusations

↩

- Chess.com. (2024a). What do I need to know about Fair Play on Chess.com? https://support.chess.com/en/articles/8568369-what-do-i-need-to-know-about-fair-play-on-chess-com

↩

- Chess.com. (2024b, July 18). Fair Play report: Estimating cheating rates in titled Tuesday. https://www.chess.com/blog/FairPlay/chess-com-fair-play-report-estimating-cheating-rates-in-titled-tuesday

↩

- Chess.com. (2024c). How do ratings work on Chess.com? https://support.chess.com/en/articles/8566476-how-do-ratings-work-on-chess-com

↩

- Chess.com. (2025, January 7). Fair Play on Chess.com. https://www.chess.com/fair-play

↩

- Dekking, F. M., Kraaikamp, C., Lopuha, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how*. Springer. https://doi.org/10.1007/1-84628-168-7

↩

- Elo, A. E. (1978). *The rating of chess players, past and present* (1st ed.). Arco.

↩

- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

↩

- Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(3), 377–394. https://doi.org/10.1111/1467-9876.00159

↩

- Glickman, M. E. (2022, March 22). *Example of the Glicko-2 system.* http://www.glicko.net/glicko/glicko2.pdf

↩

- Glickman, M. E. (n.d.). *The Glicko rating systems.* Retrieved January 15, 2025, from http://www.glicko.net/glicko/glicko.pdf

↩

- Glickman, M. E., & Jones, A. C. (1999). Rating the chess rating system. *CHANCE, 12*(2), 21–28.

↩

- GothamChess. (2024, July 6). *I confronted Kramnik* [Video]. YouTube. https://www.youtube.com/watch?v=yDeQ-CmeZaU

↩

- Keener, G. (2022, September 27). Chess is booming. *The New York Times*. https://www.nytimes.com/2022/06/17/crosswords/chess/chess-is-booming.html

↩

- Kramnik, V. (2024a, November 30). *Another research full of mistakes* [Video]. YouTube. https://www.youtube.com/watch?v=qEsUXqnuDa8

↩

- Kramnik, V. (2024b, September 29). *Let us talk, Professor* [Video]. YouTube. https://www.youtube.com/watch?v=lz2I_WEG8lY

↩

- Maharaj, S., Polson, N., & Sokolov, V. (2023). *Kramnik vs Nakamura or Bayes vs p-value*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4648621

↩

- McClain, D. L. (2023, December 25). Elite chess players keep accusing each other of cheating. *The New York Times*. https://www.nytimes.com/2023/12/25/crosswords/chess-hikaru-vladmir-kramnik-cheating.html

↩

- Miller, J. S., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica, 86*(6), 2019–2047. https://doi.org/10.3982/ECTA14943

↩

- Nakamura, H. (2024a, September 2). *Prof. of Stats Reports | Kramnik: "Lies!!"* [Video]. YouTube. https://www.youtube.com/watch?v=bfn9gAEY0Ck

↩

- Nakamura, H. (2024b, November 13). *Farming 3000s* [Video]. YouTube. https://www.youtube.com/watch?v=SvqPGAwS2Wg

↩

- Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research, 7*(2), 106–107. https://doi.org/10.4103/2229-3485.179436

↩

- Robert, C., & Casella, G. (2009). *Introducing Monte Carlo methods with R* (1st ed.). Springer. https://doi.org/10.1007/978-1-4419-1576-4

↩

- Rosenthal, J. S. (2014). Statistics and the Ontario lottery retailer scandal. *CHANCE, 27*(1), 4–9. https://doi.org/10.1080/09332480.2014.890864

↩

- Rosenthal, J. S. (2024a, August 9). *Probabilities of streaks in online chess* (original version). http://probability.ca/jeff/ftpdir/chessstreaksorig.pdf

↩

- Rosenthal, J. S. (2024b, September 5). *Probabilities of streaks in online chess* (addendum version). http://probability.ca/jeff/ftpdir/chessstreaksadd.pdf

↩

- Ross, K. (2022). *An Introduction to Bayesian reasoning and methods*. Bookdown. https://bookdown.org/kevin_davisross/bayesian-reasoning-and-methods/

↩

- Svensen, T. J. (2024, September 7). *Professor finds Nakamura's winning streaks statistically normal, Kramnik dismisses it.* Chess.com. https://www.chess.com/news/view/nakamura-winning-streaks-statistically-normal-professor-says

↩

- Szczepanek, A. (2024, January 18). *Coin toss streak calculator.* Omni Calculator. https://www.omnicalculator.com/statistics/coin-flip-streak

↩

- Wired. (2022, February 21). *Statistician answers stats questions from twitter | Tech Support | WIRED* [Video]. YouTube. https://www.youtube.com/watch?v=QW3KRaz4aI4

↩