Thinking Time Versus Quality of Play: A Review of Results

Chess as a Research Laboratory Panel, STLCC 2026

Kenneth W. Regan[1]
University at Buffalo (SUNY)

23 October, 2026

[1]With grateful acknowledgment to co-authors Guy L. Haworth and Tamal Biswas, students in my graduate seminars, and UB's Center for Computational Research (CCR)

Chess as Testbed for Wider Social Behavior

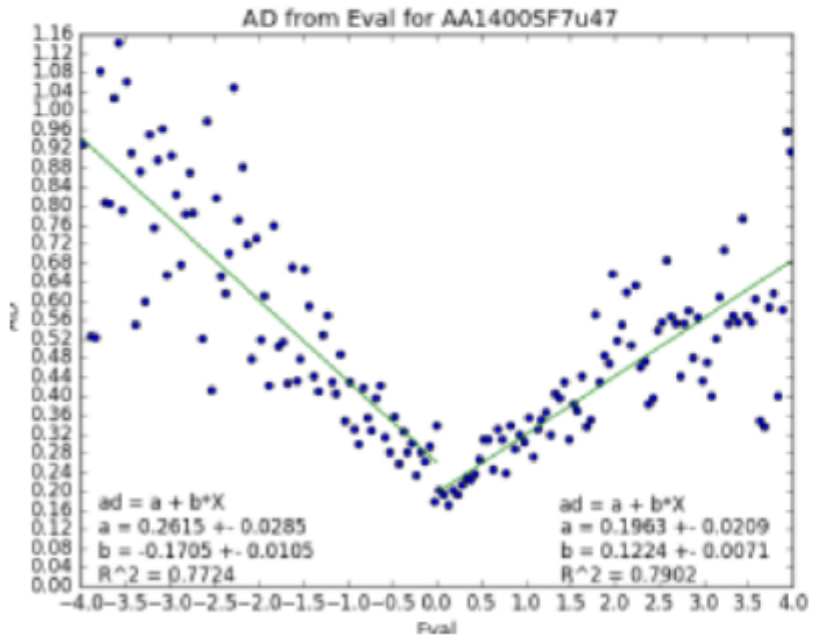
Premise of my 4/24 obit of Daniel Kahneman:

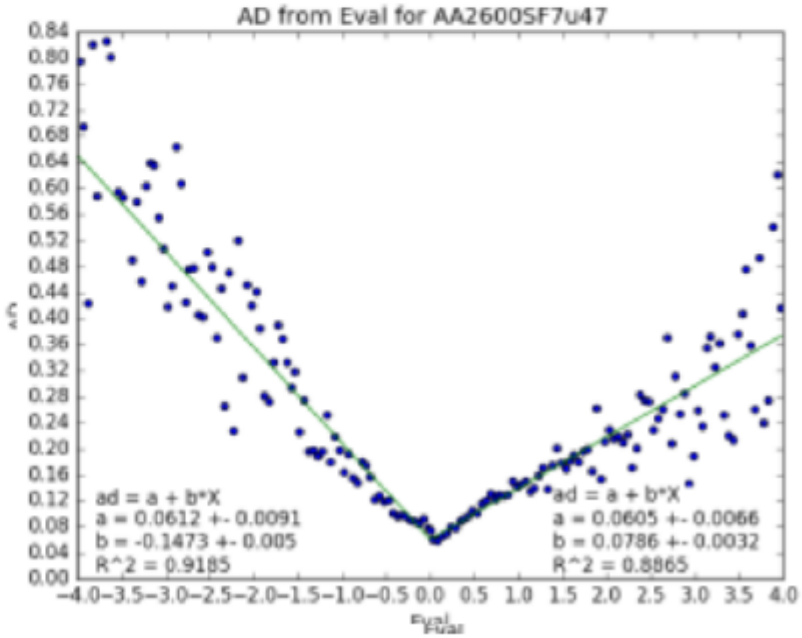
> In place of **well-targeted** but **small** direct studies of **low replication assurance**, transpose into **chess**, where *competition is real*, *data is plentiful*, and *metrics are clear*.

- Instance there: Penny-pinching disappears on bigger purchases $\longleftrightarrow$ **Average centipawn loss scales linearly with position value.**
- Instance here: **Longer deliberations yield lower-quality decisions...** *by which we mean chess moves.*
- Latter challenges Kahneman's "System 2" hypotheses.
- **Further theme:** clearinghouse for chess research.

Quoting Kahneman...

> "It is common . . . for someone to be willing to drive across town to save $10 on a $20 clock radio, but unwilling to do so to save $10 on a $1,000 television set."





Is this chess ACPL behavior commensurate? OK, on to time usage...

Some groups that could know more of each other

1. Ashton Anderson, Jon Kleinberg, and Sendhil Mullainathan, "Assessing Human Error Against a Benchmark of Perfection," in *Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), pp 705–714. Journal version *ACM Trans. Knowl. Discov. Data* **11** (4) (2017). Also ArXiv 1606.04956, June 2016.
   - 24M FICS online fast games + 1M OTB GM-level games, only FICS blitz used for time analysis.
2. Uwe Sunde, Dainis Zegners, and Anthony Strittmatter, "Speed, Quality, and the Optimal Timing of Complex Decisions: Field Evidence," ArXiv 2201.10808, January 2022. Publ. as "Speed and quality of complex strategic decisions," *PNAS* **123** (20), May 2026.
   - All on decision speed vs. quality.
   - Cites earlier 2020 paper by Kleinberg, Anderson et al.
   - IMHO the clear antecedent.

Some groups that could know each other, cont'd

3. Johannes Carow and Niklas M. Witzig, "Time pressure and strategic risk-taking in professional chess," Johannes Gutenberg-Universität Working Paper 2404, 2024. Published in J. Econ. Behav. Organ. 238, 107218 (2025).
   - FIDE World Cups. Includes tiebreaks and rating difference as isk-taking factors.
   - Remarks on Kahneman-Tversky work, but really addresses only the time remaining.
   - **Fulcrum**: Cites papers 1 and 2, cited by 2026 version of 2.
4. Michael Vidulich and Pamela Tsang, "Intuition and deliberation in elite expertise," *Cognitive Research: Principles and Implications* **11** (art. 35), April 2026.
5. Albert Hamood, "Long think, wrong think: Olympiad edition," *ChessDS Substack*, Oct. 4, 2026.

Also mention: papers in 2020–2024 by Kleinberg and Anderson with Reid McIlroy-Young, Siddhartha Sen, and.or Karim Hamade.

As we fly over these papers, ask how easily one can ascertain the magnitude and scope of the effect.

Pages From Research Papers on Thinking Time at Chess

#1: Andersen-Kleinberg-Mullainathan

Assessing Human Error Against a Benchmark of Perfection

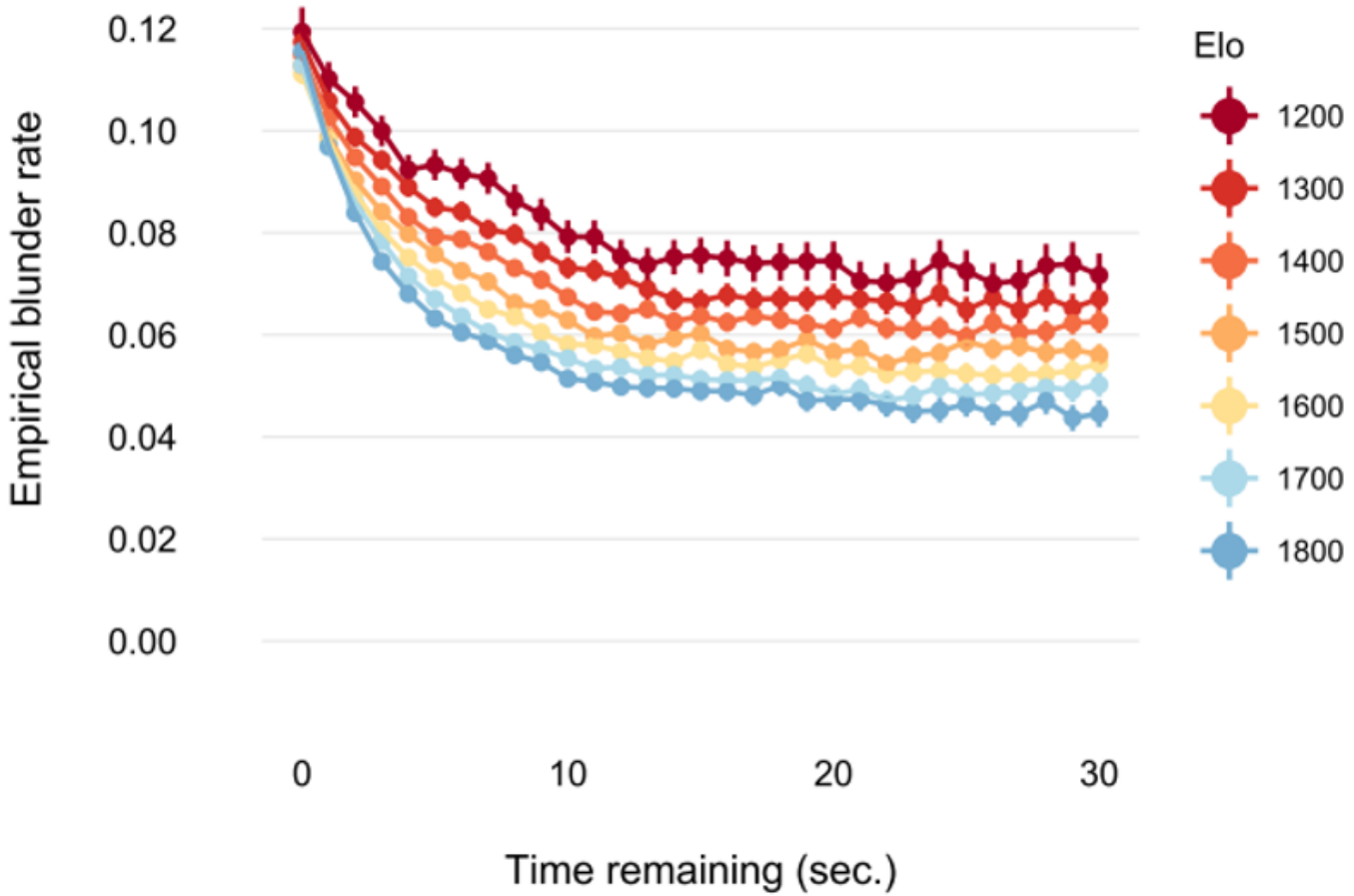


Fig. 10. The empirical blunder rate as a function of time remaining for several skill levels.

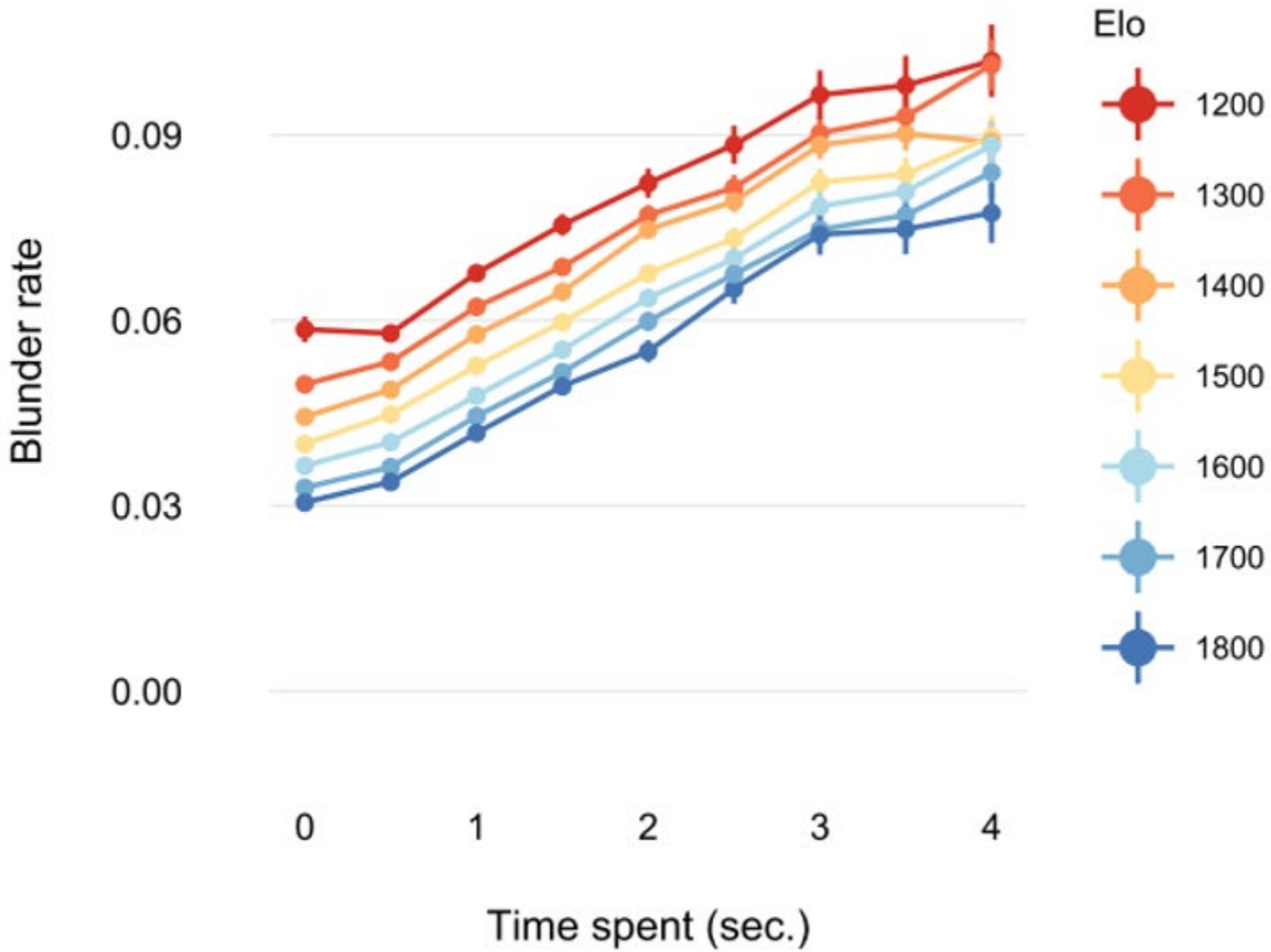


Fig. 13. The empirical blunder rate as a function of time spent on the move separated by rating bucket. The data is restricted to positions where the player to move have at least 20 seconds remaining.

When a player spends more time on a move, should we predict they are less likely to blunder (because they gave the move more consideration) or more likely to blunder (because the extra time suggests they did not know what to do)? The data turns out to be strongly consistent with the latter view

Time Spent on a Move. Thus far, we have looked at how the empirical blunder rate depends on the amount of time remaining in the game. But we can also ask how the probability of a blunder varies with the amount of time the player actually spends considering their move before playing it. When a player spends more time on a move, should we predict they are less likely to blunder (because they gave the move more consideration) or more likely to blunder (because the extra time suggests they did not know what to do)?

The data turns out to be strongly consistent with the latter view. As shown in Figure 13 (where we restrict to instances in which players have at least 20 seconds remaining), the empirical blunder rate is higher for players who spend more time playing a

[Did not do on their GM dataset. See ICLR 2024 paper including Andersen and Kleinberg for more.]

#2 Sunde-Zegners-Strittmatter

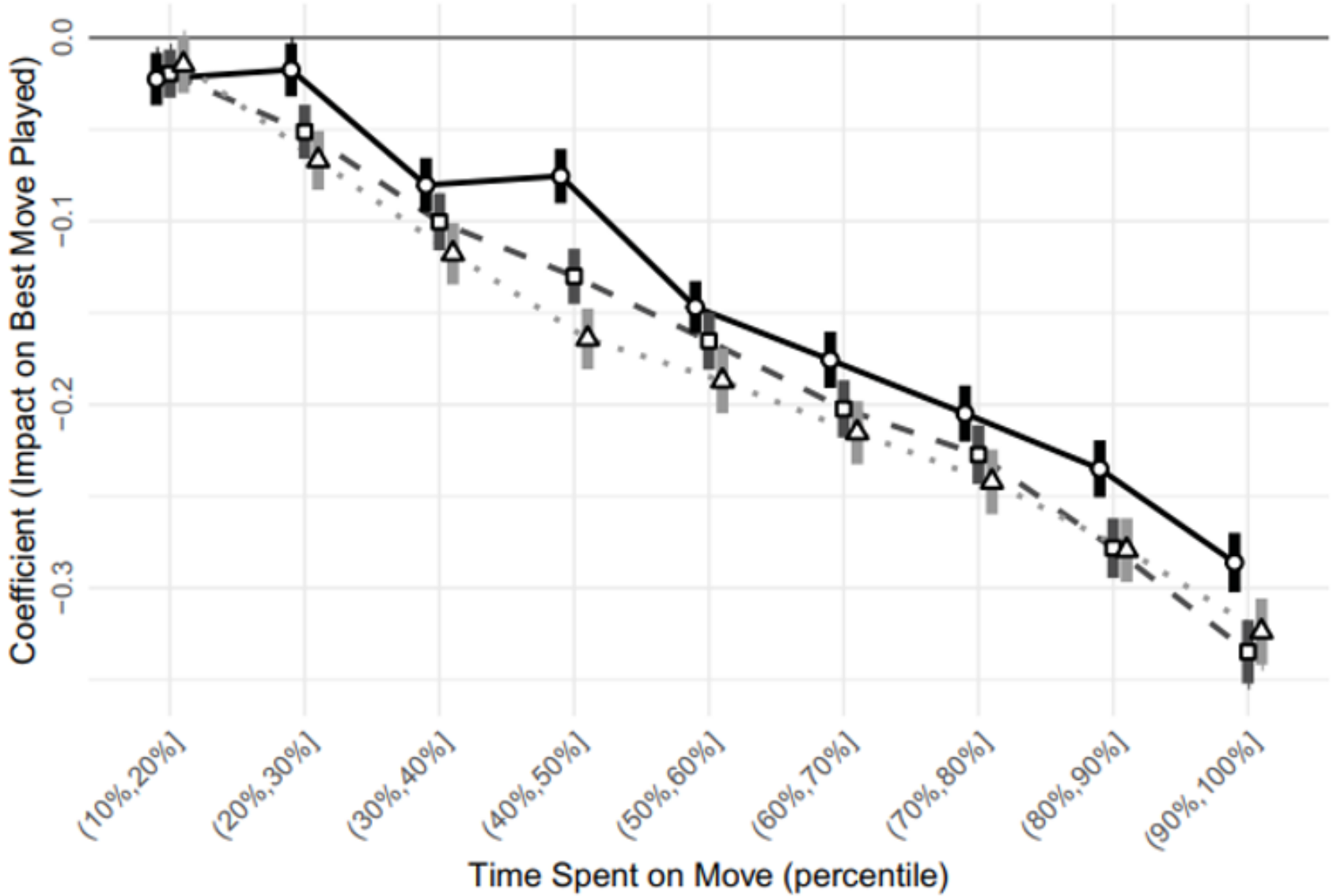


Fig. 1. Decision time and move quality across the time distribution. *Note:* The figure presents estimates from a flexible model. Decision times are divided into ten equally sized bins (by number of observations), defined separately for Classical, Rapid, and Blitz games. All coefficients are expressed relative to the 10% quickest moves (the [0, 10%] bin). For example, the coefficient for the bin (70%, 80%] shows the probability of playing the best move for that bin relative to the fastest 10% of moves, within each time control format. The regression specification is otherwise the same as in Table 1, including controls for available time budget, computational complexity, and distinctiveness, as well as player–game fixed effects, move-number fixed effects, and indicators for favorable or unfavorable positions. Error bars show 95% CIs with SEs clustered at the game level.

	Dependent variable: Best move (Dummy)								
	Classical			Subset:Rapid			Blitz		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Decision time									
Time spent on move (min.)	−0.023***	−0.018***	−0.021***	−0.134***	−0.108***	−0.131***	−0.666***	−0.550***	−0.744***
	(0.0005)	(0.0005)	(0.0007)	(0.003)	(0.003)	(0.005)	(0.019)	(0.018)	(0.025)
Time budget									
Remaining time (min.)		0.0008***	0.0006***		0.006***	0.005***		0.022***	0.010
		(0.0002)	(0.0002)		(0.002)	(0.002)		(0.008)	(0.008)
Complexity									
N Mega-nodes computed (log)		−0.029***	−0.034***		−0.052***	−0.054***		−0.067***	−0.069***
		(0.004)	(0.004)		(0.005)	(0.005)		(0.004)	(0.004)
Evaluation gap									
Distance second best move (log)		0.197***	0.239***		0.174***	0.194***		0.152***	0.159***
		(0.004)	(0.007)		(0.005)	(0.008)		(0.005)	(0.006)
Interactions									
Time spent × Remaining time			0.0003***			0.010***			0.312***
			(0.00002)			(0.0008)			(0.022)
Time spent × N Mega-nodes			−0.004***			−0.010			−0.069*
			(0.0008)			(0.008)			(0.036)
Time spent × Dist. second best			0.021***			0.072***			0.190***
			(0.002)			(0.017)			(0.056)
Move observations	80,633	80,633	80,633	62,825	62,825	62,825	73,945	73,945	73,945
Game observations	1,497	1,497	1,497	1,007	1,007	1,007	1,181	1,181	1,181
Player-game fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control move number	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control evaluation position	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents OLS estimates. Columns (1)–(3) only include games played with a classical time control; columns (4)–(6) only games with a rapid time control; and columns (7)–(9) games with a blitz time control. The evaluation of the current position is controlled for using two dummy variables indicating whether the player to move is in a favorable (>0.5 pawn units) or unfavorable (< −0.5 pawn units) position. All independent variables have been demeaned, allowing main effects to be interpreted as marginal effects at the mean when interaction terms are present. SEs are clustered at the game level, and significance levels are indicated as follows: *: $P < 0.1$, **: $P < 0.05$, and ***: $P < 0.01$.

#3 Carow-Witzig

6. Discussion and interpretation of main result

The main result of our paper is a positive effect of available thinking time on revealed strategic risk-taking behavior of chess players. While the effect sizes are small-to-medium, the effect is robust across various subsamples and for different weighting schemes of our risk measure and replicates in analyses that use a different outcome for constructing the risk measure. In addition, the effect materializes comparably in a dataset where concerns about endogenous selection of players into playing modes are less likely to apply. Overall, we read these findings as evidence in favor of a genuine effect.

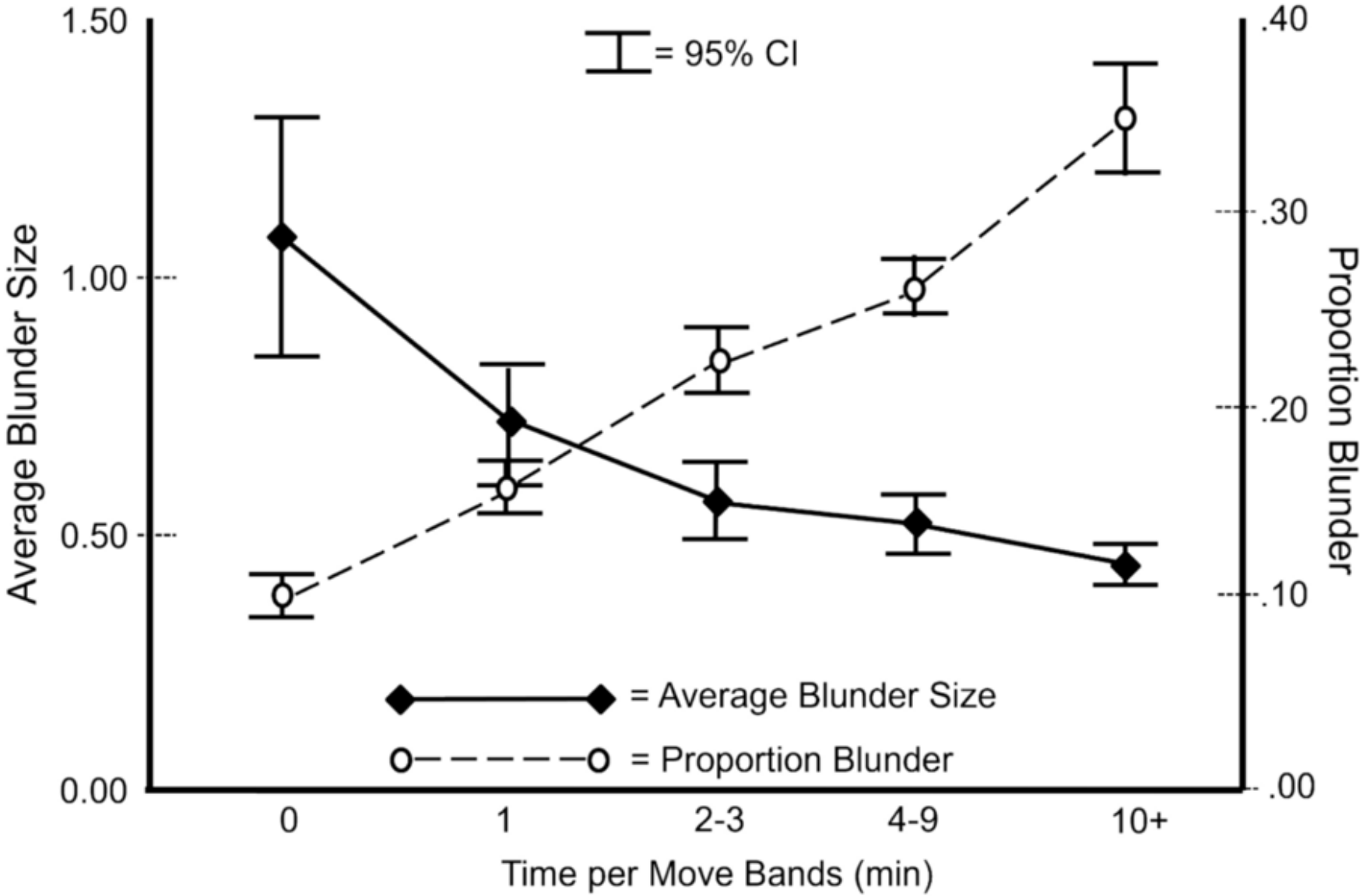
...

Nonetheless, our paper innovates on multiple fronts: We develop a novel measure for the degree of strategic risk entailed in single chess moves. We thus advance existing approaches that focus entirely on chess openings. In addition, we introduce a way to measure the difficulty of a chess position, an interesting and transportable quantity by itself. Both our measure of strategic risk on single moves and the difficulty of given chess positions should be insightful for a plethora of possible research questions, putting the dynamic development of strategic risk and difficulty during play into focus.

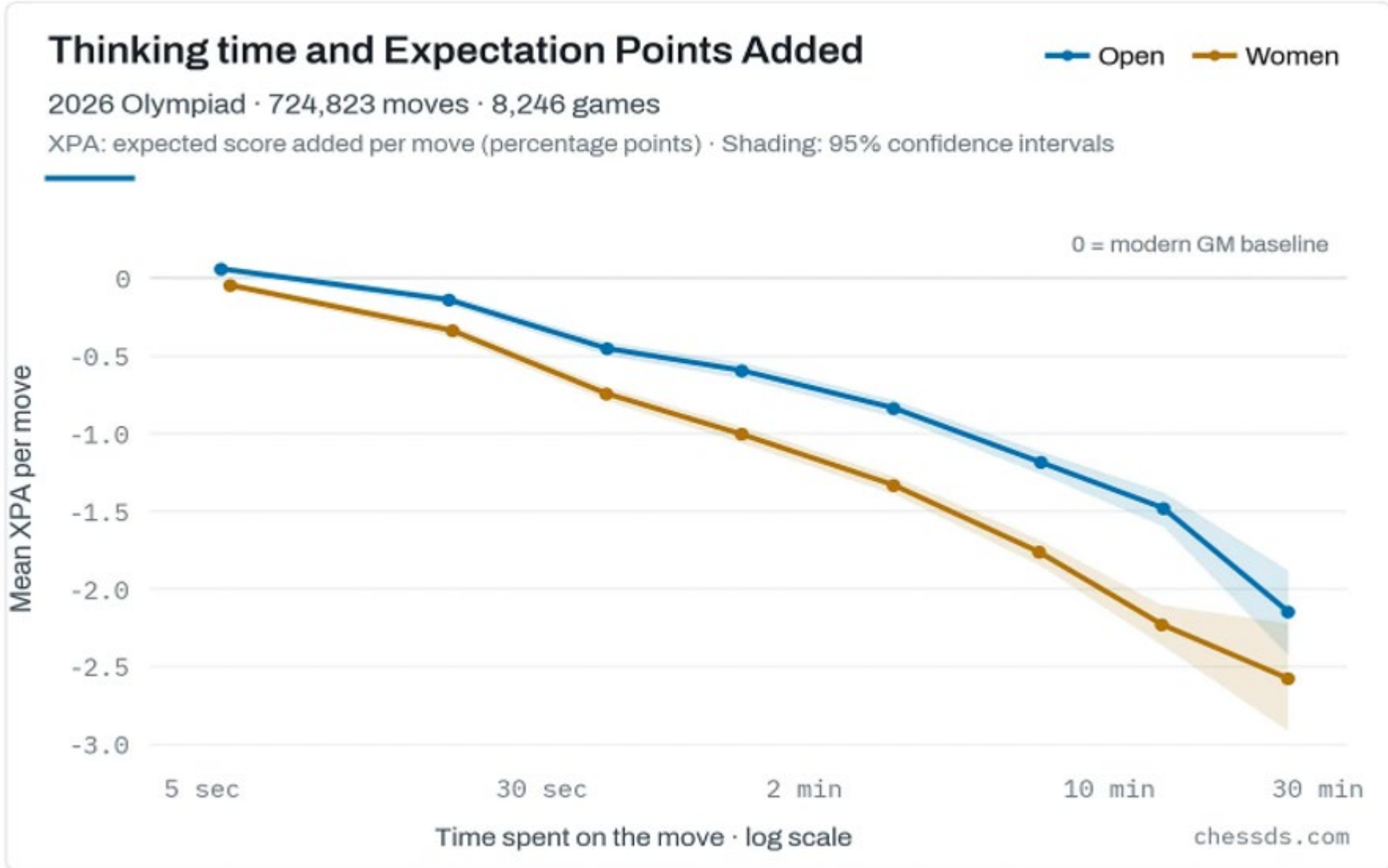
#4 Vidulich-Tsang

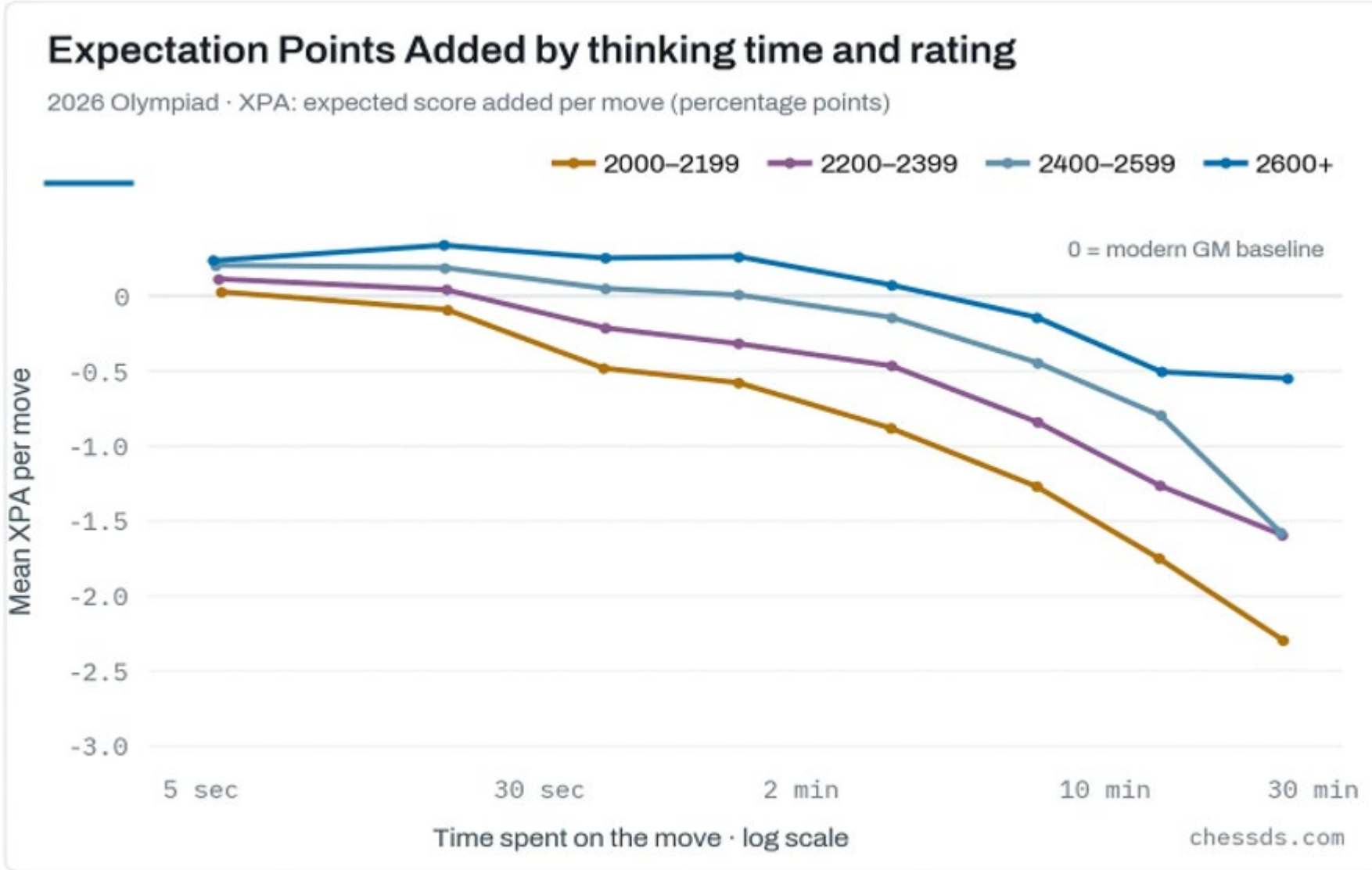
Fig. 4

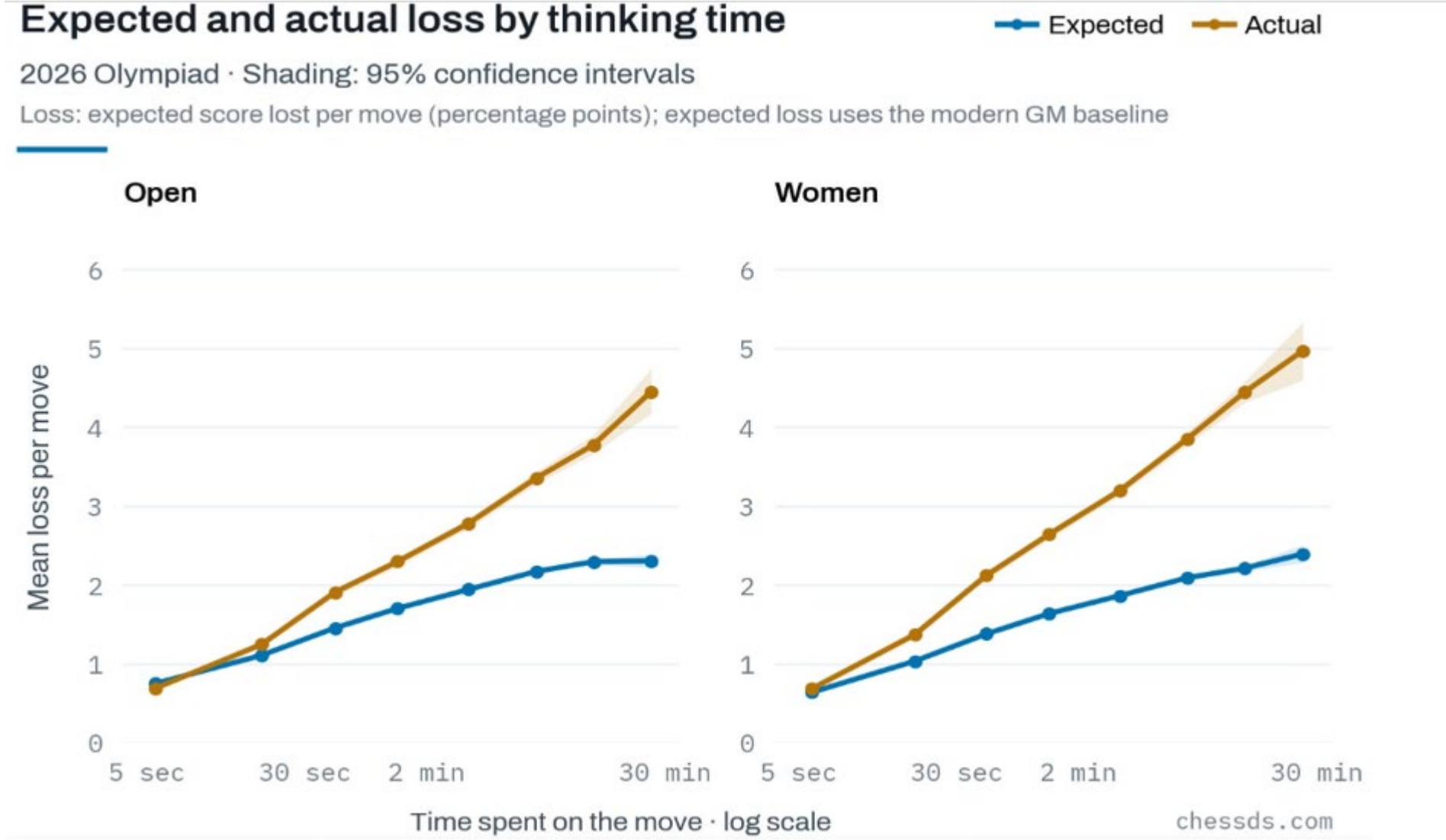
From: Intuition and deliberation in elite expertise



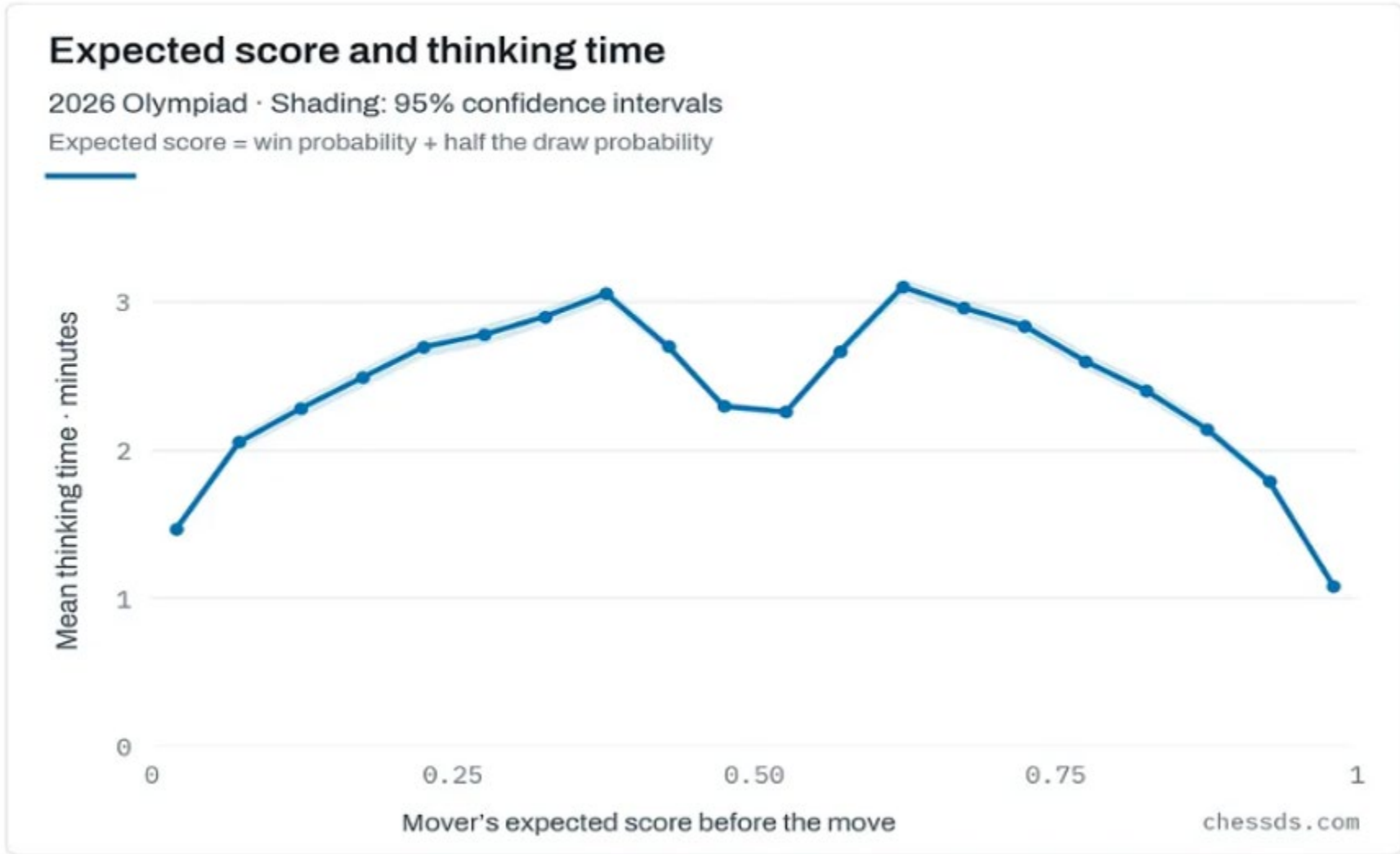
#5 Albert Hamood:




Expectation Points Added by thinking time and rating
2026 Olympiad · XPA: expected score added per move (percentage points)
2000–2199
2200–2399
2400–2599
2600+
0 = modern GM baseline
Mean XPA per move
0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
5 sec
30 sec
2 min
10 min
30 min
Time spent on the move · log scale
chessds.com


Expected and actual loss by thinking time
Expected
Actual
2026 Olympiad · Shading: 95% confidence intervals
Loss: expected score lost per move (percentage points); expected loss uses the modern GM baseline
Open
Women
Mean loss per move
0
1
2
3
4
5
6
5 sec
30 sec
2 min
30 min
Time spent on the move · log scale
chessds.com

move. Note the dip around 0.5 – these buckets contain many clearly drawn positions. The peaks correspond to positions where the player is at an advantage or disadvantage that is not yet decisive:



Enter a Fully Predictive Analytic Model

Means that the model:

- Addresses a series of events or decisions, each with possible outcomes $m_1, m_2, \ldots, m_j, \ldots$
- Assigns to each m_j a probability p_j.
- Projects risk/reward quantities associated to the outcomes.
- Also assigns *confidence intervals* for p_j and those quantities.

In a **utility-based** model, each m_i has a utility or cost u_i.
Main risk/reward quantity then becomes $E = \sum_i p_i u_i$.

- **Insurance**: m_i are risk factors; costs u_i need not influence p_i.
- **Chess**: m_i are legal moves; u_i are engine values and influence p_i.
- **Multiple-choice tests**: m_i are possible answers to a test question, u_i = gain/loss for right/wrong answer.

How The Model Operates

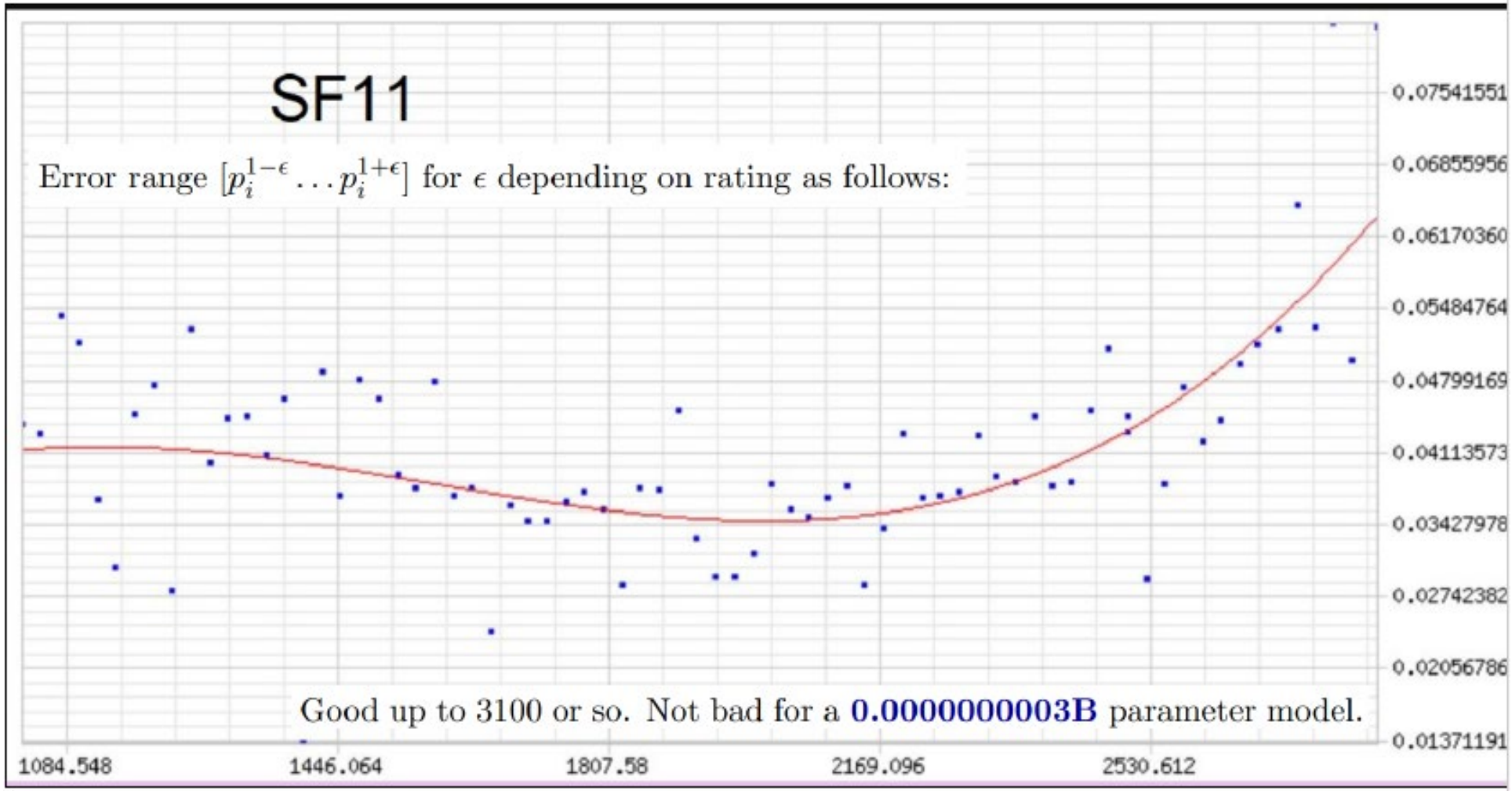
- Three main parameters s, c, h are trained to Elo ratings R.
- $s_R, c_R, h_R + \{u_i\} \mapsto \{p_i\}$ for each legal move m_i.
- (**No other chess knowledge** than R and the move values.)
- Paint m_i on a 1,000-sided die **1,000**p_i times **and roll it**. Bernoulli!
- (Correct after-the-fact for chess decisions not being independent.)

Main Outputs:

- **Statistical z-scores** for various (*actual*−*projected*) quantities:
  - **T1-match**: Agreement with the move listed first by the computer.
  - **EV-match**: Includes moves of equal-optimal value not listed first.
  - **ASD**: Average *scaled* difference in value from inferior moves.
- An **Intrinsic Performance Rating** (**IPR**) for the set of games.

Clock times are **not** used for training or prediction (in real cases they are often absent), so **IPR vv. Time** is fully downstream of the model.

Predictive Accuracy



Long Thinks in Samarkand, By Rating Cohort

Rtgs	IPR	15+m	10-15m	5-10m	3-5m	1-3m	15+ sz
1400s	1433	670	910	873	1085	1216	134
1500s	1586	941	1015	987	1221	1380	506
1600s	1715	1028	998	1153	1335	1536	867
1700s	1794	1046	1182	1287	1430	1623	1334
1800s	1922	1205	1287	1423	1549	1748	2129
1900s	2027	1370	1383	1559	1685	1846	2576
2000s	2141	1462	1489	1712	1806	1955	2808
2100s	2220	1560	1585	1774	1896	2042	3043
2200s	2292	1611	1657	1840	1979	2120	2870
2300s	2382	1731	1787	1943	2084	2221	2874
2400s	2465	1781	1890	2020	2161	2318	2338
2500s	2541	1779	2053	2126	2231	2384	1639
2600s	2609	1948	2059	2202	2227	2455	791
2700s	2664	2095	2186	2206	2298	2537	314

Short Thinks (with at least 1 minute left)

IPR	31-60s	11-30s	6-10s	4-5s	2-3s	$\leq$ 1s	2-3s sz	1s sz.
1433	1511	2088	2528	3080	2745	2084	120	41
1586	1696	2191	2790	2872	2836	2151	382	131
1715	1821	2310	2819	2901	2892	2426	660	199
1794	1853	2317	2869	2966	2901	2603	1207	354
1922	1984	2451	2977	3032	2996	2828	1842	555
2027	2116	2547	3056	3163	3101	2830	2234	646
2141	2218	2627	3096	3184	3217	2885	2634	775
2220	2286	2688	3138	3236	3252	3079	3045	901
2292	2352	2736	3146	3206	3245	3111	3092	887
2382	2412	2804	3183	3238	3253	3184	3326	947
2465	2512	2870	3204	3221	3256	3157	2909	834
2541	2615	2904	3221	3239	3260	3181	1995	540
2609	2689	2988	3253	3276	3292	3259	1088	262
2664	2730	3026	3208	3233	3260	3288	467	119

Losses From Long Thinks, By Rating Cohort

Rtgs	IPR	15+m	10-15m	5-10m	3-5m	1-3m	15+ sz
1400s	1433	-763	-523	-560	-348	-217	134
1500s	1586	-645	-571	-599	-365	-206	506
1600s	1715	-687	-717	-562	-380	-179	867
1700s	1794	-748	-612	-507	-364	-171	1334
1800s	1922	-717	-635	-499	-373	-174	2129
1900s	2027	-657	-644	-468	-342	-181	2576
2000s	2141	-679	-652	-429	-335	-186	2808
2100s	2220	-660	-635	-446	-324	-178	3043
2200s	2292	-681	-635	-452	-313	-172	2870
2300s	2382	-651	-595	-439	-298	-161	2874
2400s	2465	-684	-575	-445	-304	-147	2338
2500s	2541	-762	-488	-415	-310	-157	1639
2600s	2609	-661	-550	-407	-382	-154	791
2700s	2664	-569	-478	-458	-366	-127	314

Gains in Short Thinks (with at least 1 minute left)

IPR	31-60s	11-30s	6-10s	4-5s	2-3s	$\leq$ 1s	2-3s sz	1s sz.
1433	78	655	1095	1647	1312	651	120	41
1586	110	605	1204	1286	1250	565	382	131
1715	106	595	1104	1186	1177	711	660	199
1794	59	523	1075	1172	1107	809	1207	354
1922	62	529	1055	1110	1074	906	1842	555
2027	89	520	1029	1136	1074	803	2234	646
2141	77	486	955	1043	1076	744	2634	775v
2220	66	468	918	1016	1032	859	3045	901
2292	60	444	854	914	953	819	3092	887
2382	30	422	801	856	871	802	3326	947
2465	47	405	739	756	791	692	2909	834
2541	74	363	680	698	719	640	1995	540
2609	80	379	644	667	683	650	1088	262
2664	66	362	544	569	596	624	467	119

Long & Short Thinks, This Wk. By Champs & Tots

Event	IPR	15+m	10-15	5-10m	3-5m	1-3m	15+sz
USAchO 2687	2771	2095	2431	2257	2525	2629	113
USAchW 2531	2462	1418	1852	1986	2424	2321	89
UKCC U08,10	1515	n.a.	n.a.	452	1297	996	0

Event	31-60	11-30	6-10s	4-5s	2-3s	$\leq$ 1s	2-3s sz	1s sz.
USA O	2761	3203	3264	3355	3375	3212	174	18
USA W	2647	2912	3253	3169	3260	2687	168	24
UKCC	1268	1535	1866	2427	2532	2436	508	93

The effect is **huge** and its scope is everyone under 8 to over 80, all time controls.

A Hub Is Needed—Apart From Me

- Frank Niro tried with the *Journal of Chess Research* in late 2014.
- My 2015 update took until late 2019 to tame. Cubic regressions!
- Then "Pandemic Lag" + 2024 "Sonas Correction" roiled ratings.
- Back in 2013 I hoped for this division:
  - Screening data public (the "model" for it is just normalizing);
  - Screening data is not for-judgment, does not assess difficulty.
  - Full test data public, model **as-coded** sensitive and private.
- But FIDE believes my screening data is too sensitive as well.
- Witch-hunting, cognitive misconceptions too much in evidence.
- Mehmet Ismail ("Mars Seven") has "screening+" model; worked for NorwayChess and some other events.
- The Maia Project does full prediction but mostly for fast chess.
- What to Do? $\Longrightarrow$ Q & A.