Chess Research and Public Interests

New Faculty Academy 2026

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

8 April, 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)

First, some chess and education background on me

- Grew up in Paramus, NJ.
- Youngest USCF Master since Bobby Fischer, 1973.
- US Junior Champion (first equal), 1977.
- FIDE International Master title, 1981. (I am not a Grandmaster),
- BA (not BS) in Mathematics, Princeton University, 1981.
- Marshall Scholar to Oxford (Merton College), 1981–1984.
- Junior Fellow of Merton. 1984–1986 and 1988, interleaved with—
- Cornell MSI postdoc (under Anil Nerode and Juris Hartmanis), 1986–1987 and 1988–1989.
- Assistant Professor at UB, Department of Computer Science (now CSE) from August 1989; tenure in 1995.
- Tenure and great colleagues conferred blessings of Freedom...

Research Background

- Combinatorial Mathematics, then Computational Complexity.
- Central problem: P Versus NP.
- Expert on logical status: is it **undecidable** a-la Kurt Gödel? How to surmount barriers?
- In 1997, I thought I had an inside track on solving (the "algebraic version" of) it.
- Quantum Computing neighbors complexity. **Shor's Theorem**.
- Co-wrote textbook with Richard Lipton.
- Also co-wrote his weblog Gödel's Lost Letter and P=NP, 2009–2024.
- All the while, I resisted entreaties to do computer chess.
- Then came the Topalov-Kramnik 2006 World Championship Scandal.

My Reactions

1. (Try to) Reproduce the Accusation.
2. Observe Facts About Chess Programs:
   - As the search progresses, they "change their minds" about which move is best.
   - List top move at each depth of search and whether it "matches" the move that was played (MM).
   - (Un)fair to count a "coincidence" if the played move matches at any depth?
   - Can set chess program to give values of multiple possible moves.
3. Simple Principle: The wider the value gap between the best move and any other move, more likely a strong human player will find it.
4. Judge a move "forced" if the value loss of the next-best move ("delta") is catastrophic.
5. Also cases where multiple moves have Equal-optimal Value (EV).
6. Tally MM and EV cases and deltas over the games.
7. Find people to talk to...and compare results.

The Fortune and Blessings of Simplicity

- How to build a *quantitative* model out of the simple principle?
- Simple elements **Strategy** and **Tactics** take us far.
- **Depth of Thinking** should be next.
- Do weaker players **prefer** weaker moves?
- Or are they more easily **distracked?**
- How shall we handle the element of **Difficulty**?
- **Recognition** "Versus" **Thinking**.
  - See the 2007 National Geographic documentary "My Brilliant Brain" with Susan Polgar (crux here).
  - We will try to glean comparable insight from numerical analytics.

A 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.

Chess and Tests—With Partial Credits (Or LLMs?)

The ____ of drug-resistant strains of bacteria and viruses has ____ researchers' hopes that permanent victories against many diseases have been achieved.

(a) vigor . . corroborated

(b) feebleness . . dashed

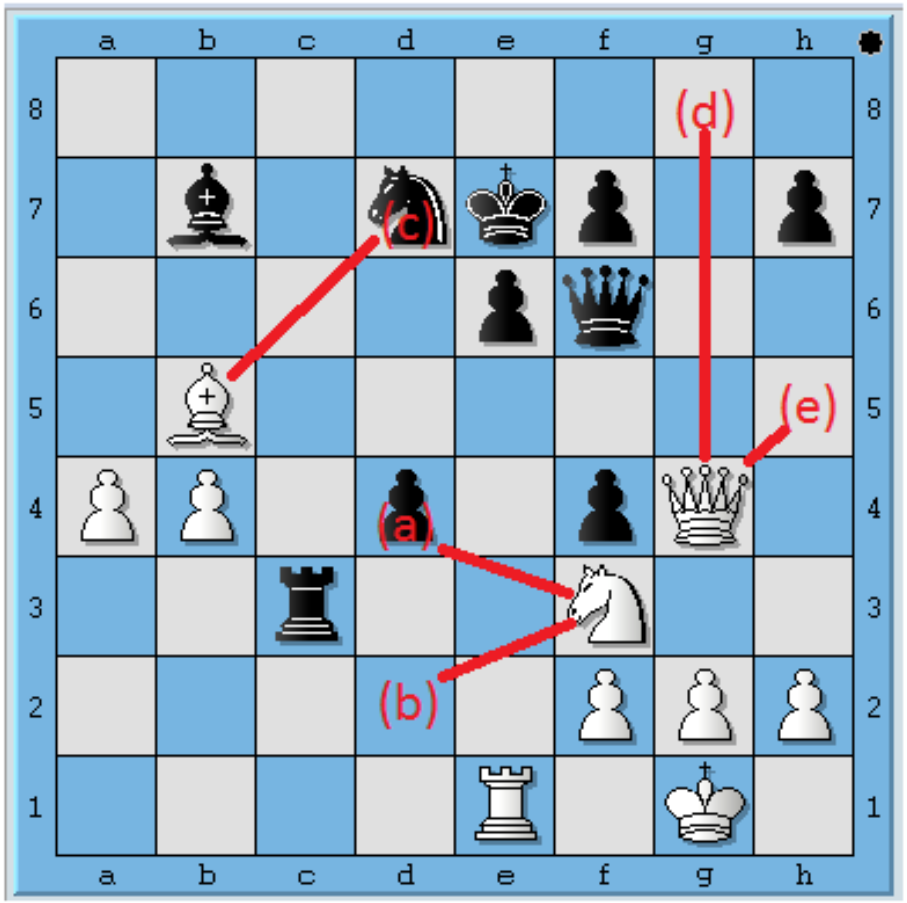
(c) proliferation . . blighted

(d) destruction . . disputed

(e) disappearance . . frustrated

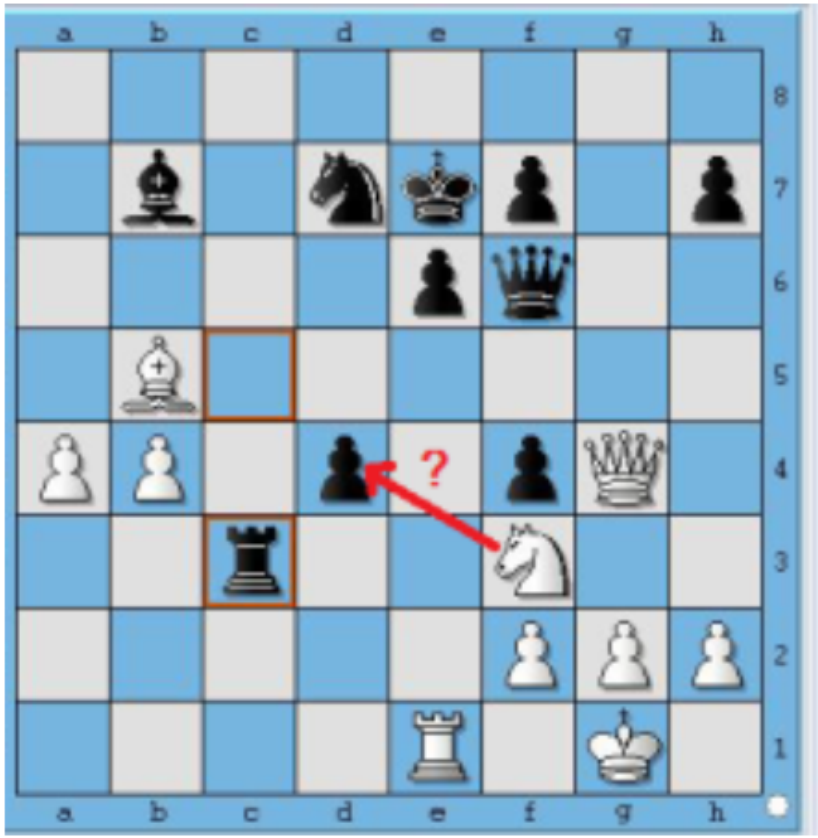
(source: itunes.apple.com)

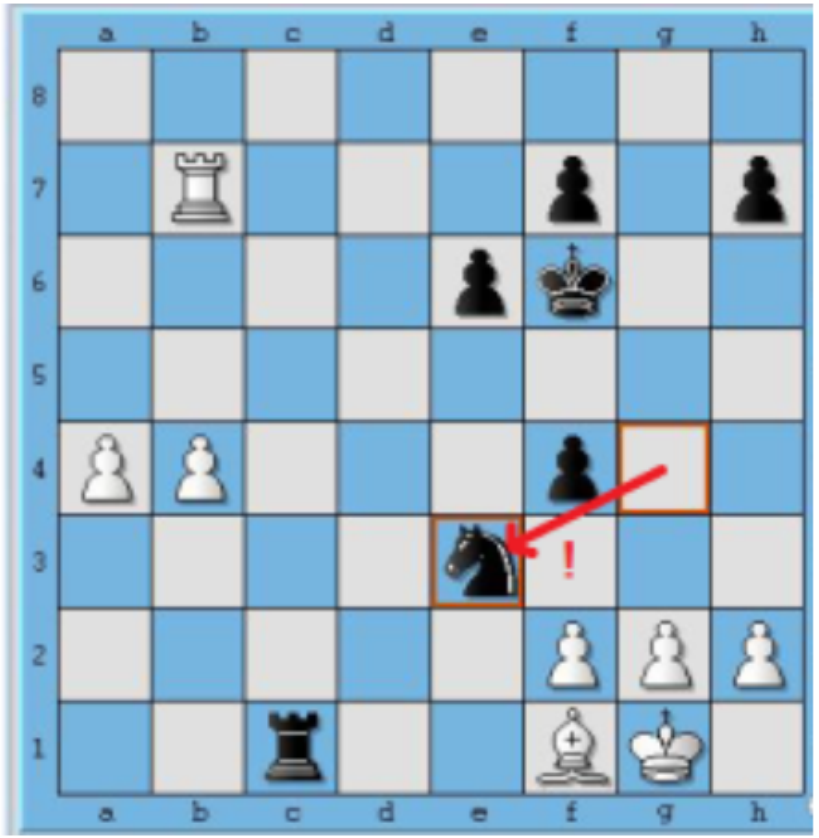
=



Here (b,c) are **equal-optimal** choices, (a) is bad, but (d) and (e) are reasonable—worth part credit.

Move Utilities Example (Kramnik-Anand, 2008)





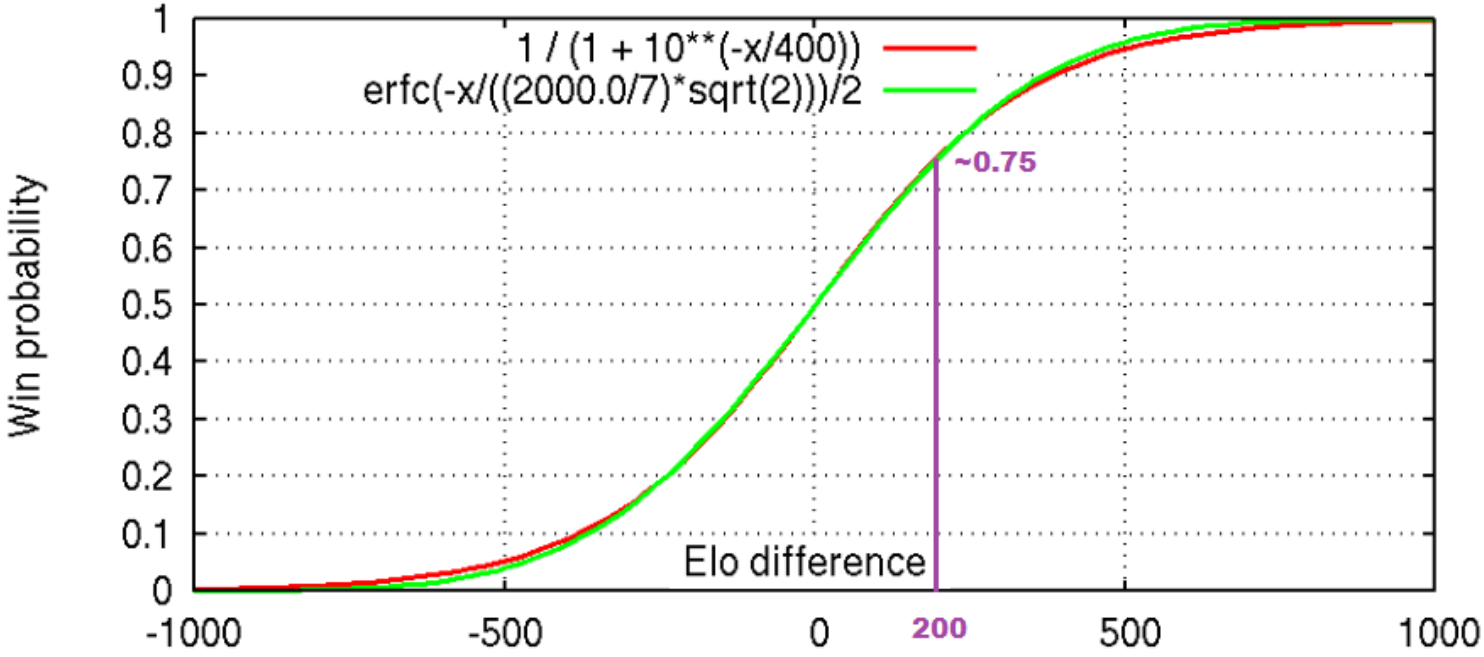
Depths...

Values by Stockfish 6

Move	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Nd2	103	093	087	093	027	028	000	000	056	-007	039	028	037	020	014	017	000	006	000
Bxd7	048	034	-033	-033	-013	-042	-039	-050	-025	-010	001	000	-009	-027	-018	000	000	000	000
Qg8	114	114	-037	-037	-014	-014	-022	-068	-008	-056	-042	-004	-032	000	-014	-025	-045	-045	-050
...			...			...			...			...			...			...	
Nxd4	-056	-056	-113	-071	-071	-145	-020	-006	077	052	066	040	050	051	-181	-181	-181	-213	-213

Aptitude—Via Elo Grades (calculator)

- Named for **Arpad Elo**, number R_P rates skill of player P.
- E.g. **1000** = bright beginner, **1600** = good club player, **2200** = master, **2800** = world championship caliber.
- Computer **engines** are far higher, e.g.: **Stockfish 16 = 3544, Torch 1.0 = 3531, Komodo Dragon 3.3 = 3529.**
- Expectation given by rating *difference* via this logistic curve:



Main Parameters and Inputs

The (only!) player parameters trained against chess Elo Ratings are:

- s for "sensitivity"—strategic judgment. *Like Anatoly Karpov.*
- c for "consistency" in tactical minefields. *Like Mikhail Tal.*
- h for "heave" or "Nudge"—obverse to depth of thinking.

Trained on all available in-person classical games in 2010–2019 with both players near the same Elo marker 1025, 1050, . . . , 2775, 2800, 2825.

Being retrained on new FIDE range **1400**. . . 2825, **from 1/1/25 on**.

- Given an Elo rating R, "central slice" gives corresponding s_R, c_R, h_R.
- Only other input is the grid of move utilities $u_{i,d}$ at various depths d of search, further **scaled** to make (perceived) values v_i (and ρ_i).
- Then $\delta_i = v_1 - v_i$ is difference to best move.
- Other than these, **my model knows nothing about chess.**

One Wonky Slide: Log-Linear Versus Loglog-Linear

The generic **log-linear** model puts

$$\log\left(\frac{1}{p_i}\right) = \alpha + \beta u_i, \quad \text{or equivalently,} \quad \log\left(\frac{1}{p_i}\right) - \log\left(\frac{1}{p_1}\right) = \beta\delta_i$$

- Solved by **softmax** giving $p_i = p_1 \cdot \exp(-\beta u_i)$.
- Each p_i is represented as a **multiple** of the top probability p_1.
- Ubiquitous in AI—but does not work for chess.

The **loglog-linear** model puts $\log\log(\frac{1}{p_i}) - \log\log(\frac{1}{p_1}) = \beta\delta_i$, i.e.:

$$\frac{\log(1/p_i)}{\log(1/p_1)} = \exp(\beta\delta_i).$$

- Gives $p_i = p_1^{\exp(\beta\delta_i)}$.
- So p_i are represented as **powers** of the best-move probability p_1.
- In place of $\beta\delta_i$, I really have $\left(\frac{\delta_i - h\rho_i}{s}\right)^c$, with h tightly clamped.

How The Model Operates

- Take s, c, h from a player's rating (or wider skill profile).
- Generate probability p_i for each legal move m_i.
- Paint m_i on a 1,000-sided die, **1,000**p_i times.
- **Roll the die** to give confidence intervals that go with the p_i.
- (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.

Fit s, c, h by making T1,EV,ASD be **unbiased estimators** on the training sets, which are stratified by Elo ratings.

Karpov & Tal at Montreal "Tourney of Stars" 1979

- Tied for first with 12/18 in star-studded double round-robin.
- Karpov was rated **2705**, Tal only **2615**.
- Karpov (per Stockfish 11): $s = \mathbf{0.016}$, $c = \mathbf{0.307}$.
- Tal (per Stockfish 11): $s = \mathbf{0.026}$, $c = \mathbf{0.365}$.
- Lower s is better—so Karpov was more "Karpovian."
- Higher c is better—so my model with Tal's parameters would make fewer large mistakes.

Are these grainy parameters enough to mimic human tendencies?

- IPRs: Karpov **2625 +- 155**, Tal **2730 +- 185**.
- Whole tourney IPR is (only!) **2575 +- 50** ($s = 0.041$, $c = 0.385$).
- Average Elo of players, **2621**, is within error bars. Surprise is that the IPR is not near 2700s range. Today's elite regularly hit 2800+.

Simplicity and Public Outreach

- Originally I intended to use *distributional distance measures*, of which **fidelity** is one.
- I realized that results would be difficult to **explain**.
- Hence I mapped everything to rolling dice, with math known since 1800.
- Other simple matters: Should Metrics Be Linear?
- "Pandemic Lag" in updating ratings.
- Are their players who are majorly better, relative to their peers at slow chess, in fast chess?
- [Segue to public-outreach cases and demos]
- Q & A — And Thanks.