

Chess Life

JUNE 2014

Dr. Ken Regan

CATCHING CHESS CHEATERS



A USCF Publication \$5.95



7 25274 64631 9

THE WORLD'S MOST WIDELY READ CHESS MAGAZINE

www.uschess.org

HOW TO CATCH A CHESS CHEATER

Ken Regan Finds Moves Out of Mind

By **HOWARD GOLDOWSKY**

Photography **LUKE COPPING**

“WHAT’S GOD’S RATING?” ASKS KEN REGAN, as he leads me down the stairs to the finished basement of his house in Buffalo, New York. Outside, the cold intrudes on an overcast morning in late May 2013; but in here sunlight pierces through two windows near the ceiling, as if this point on earth enjoys a direct link to heaven. On a nearby shelf, old board game boxes of Monopoly, Parcheesi, and Life pile up, with other nostalgia from the childhoods of Regan’s two teenage children. Next to the shelf sits a table that supports a lone laptop logged into the Department of Computer Science and Engineering’s Unix system at the University at Buffalo, where Regan works as a tenured associate professor. The laptop controls four invocations of his anti-chess-cheating software, which at this moment monitor games from the World Rapid Championships, using an open-source chess engine called Stockfish, one of the strongest chess-playing entities on the planet. Around the clock, in real-time, this laptop helps compile essential reference data for Regan’s algorithms. Regan and I are on our way to his office, where he plans to explain the details of his work. But the laptop has been acting up. First he must check its progress, and Regan taps a few keys. What he’s staring at on the screen reminds him to rephrase his question, but this time he doesn’t wait for my answer. “What’s the rating of perfect play?” he asks. “My model says it’s 3600. These engines at 3200, 3300, they’re knocking

at that door.” In Regan’s code, the chess engine needs to play the role of an omniscient artificial intelligence that objectively evaluates and ranks, better than any human, every legal move in a given chess position. In other words, the engine needs to play chess just about as well as God.

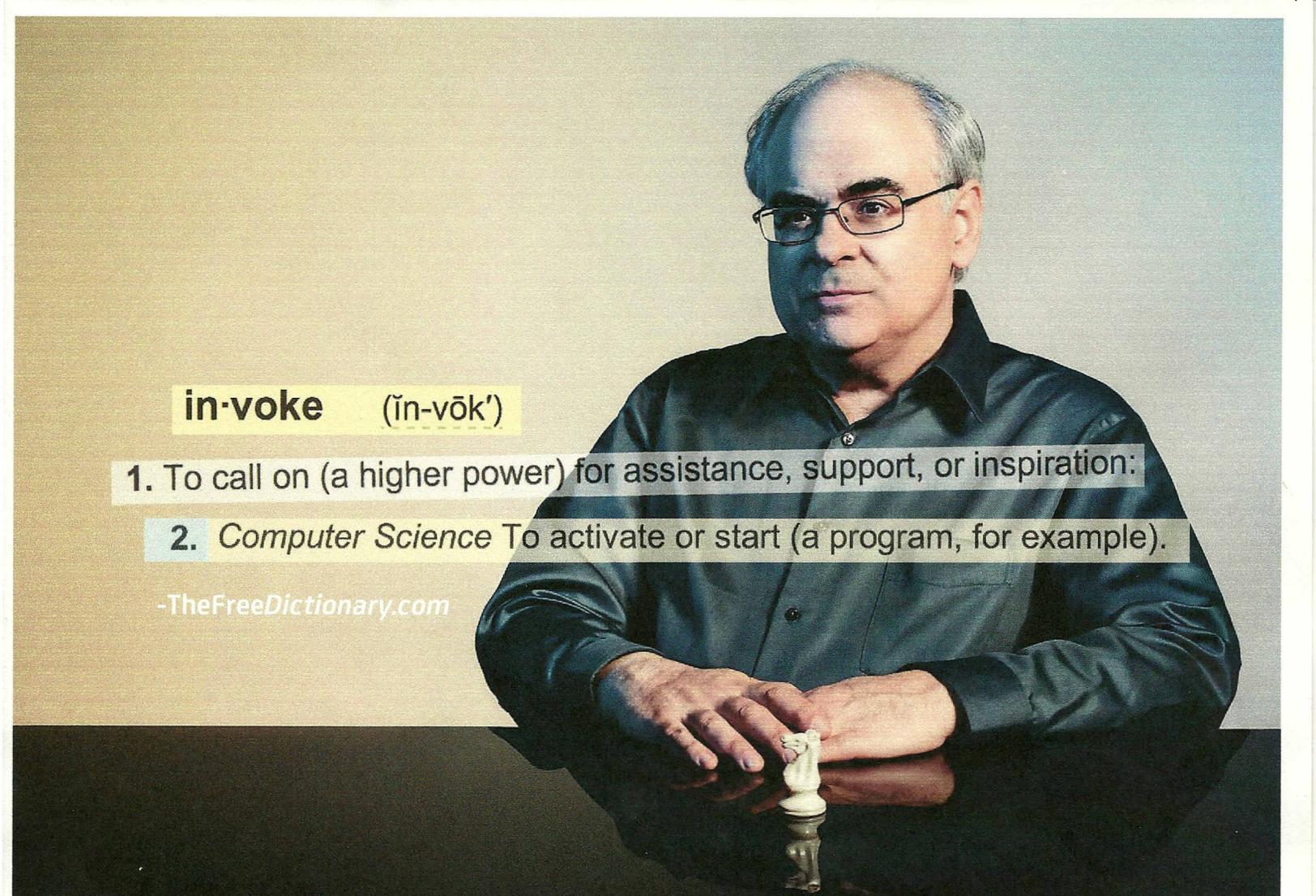
A ubiquitous Internet combined with button-sized wireless communications devices and chess programs that can easily wipe out the world champion make the temptation today to use hi-tech assistance in rated chess greater than ever (see sidebar). According to Regan, since 2006 there has been a dramatic increase in the number of worldwide cheating cases. Today the incident rate approaches roughly one case per month, in which usually half involve teenagers. The current anti-cheating regulations of the world chess federation (FIDE) are too outdated to include guidance about disciplining illegal computer assistance, so Regan himself monitors most major events in real-time, including open events, and when a tournament director becomes suspicious for one reason or another and wants to take action, Regan is the first man to get a call.

Regan is a devoted Christian. His faith has inspired in him a moral and social responsibility to fight cheating in the chess world, a responsibility that has become his calling. As an interna-

“Religion is responsibility or it is nothing.”
—Jacques Derrida

tional master and self-described 2600-level computer science professor with a background in complexity theory—he holds two degrees in mathematics, a bachelor’s from Princeton and a doctorate from Oxford—he also happens to be one of only a few people in the world with an ability to commit to such a calling. “Ken Regan is one of two or three people in the world who have the quantitative background, chess expertise, and computer skills necessary to develop anti-cheating algorithms likely to work,” says Mark Glickman, a statistics professor at Boston University and chairman of the USCF ratings committee. Every time Regan starts an instance of his anti-cheating code he does not merely run a piece of software—he invokes it. The dual meaning of “invoke” conveys Regan’s inspired relationship to the anti-cheating work that he does.

His work began on September 29, 2006, during the Topalov-Kramnik World Championship match. Vladimir Kramnik had just forfeited game five in protest to the Topalov team’s accusation that Kramnik was consulting a chess engine during trips to his private bathroom. This was the reunification match to unite the then-separate world champions, a situation created when Garry Kasparov and Nigel Short broke from FIDE in 1993. Topalov qualified for the 2006 match because he held the FIDE title. Kramnik qualified because he had defeated Garry Kasparov in 2000 to claim a spot through historical lineage. Due to the schism, chess had suffered 13 years of heavy declines in sponsorship, stability, and respect. Kramnik’s forfeiture of game five not only threatened



in-voke (ĩn-vōk')

1. To call on (a higher power) for assistance, support, or inspiration:

2. *Computer Science* To activate or start (a program, for example).

-TheFreeDictionary.com

the reunification but also the future of the sport.

Kramnik agreed to play game six, which ended in a draw. After game six, on October 4, Topalov's team published a controversial press release trying to prove their previous allegations. Topalov's manager, Silvio Danailov, wrote in the release, "... we would like to present to your attention coincidence statistics of the moves of GM Kramnik with recommendations of chess program Fritz 9." The release went on to report at what frequency Kramnik's moves for games one, two, three, four, and six matched the "first line" (Danailov's words) of Fritz's output.

An online battle commenced between pundits who took Danailov's "proof" seriously versus others, like Regan, who insisted that valid statistical methods to detect computer assistance did not yet exist. For the first time, a cheating scandal was playing a role in top-level chess. There remained all kinds of uncertainties, including how much time Fritz used to process each move, how many forced moves were played, whether the engine was in single-line or multi-line mode (in multi-line mode machines play slower but stronger, because they enable extra heuristics and do less pruning of unpromising moves), what constituted a typical matching percentage for super-grandmaster play, all kinds of questions that prohibited scientific reproduction of Danailov's accusation. In just a few weeks, the greatest existential threat to chess had gone from a combination of bad politics and a lack of financial support to something potentially more sinister: scientific ignorance. In Regan's mind, this threat seemed too imminent to ignore. "I care

about chess," he says. "I felt called to do the work at a time when it really did seem like the chess world was going to break apart."

When Regan satisfies himself with the laptop's data collection, he walks me out of his basement to the end of his driveway, where he points to a neighbor's house down the block. Regan's neighbor's brother happens to be a college friend

with whom Regan toured England before studying at Oxford, and with whom Regan spent a lot of time while on sabbatical at the University of Montreal. (The friend is a professor at McGill.) Regan loves to call attention to the connections and coincidences that surround his life, and as much as his

faith drives a moral influence

in his anti-cheating work and his interests in chess and mathematics drive a technical influence, his fascination with coincidence drives its own quirky influence. "Social networking theory is interesting," he says. "Cheating is about how often coincidence arises in the chess world."

"I felt called to do the work at a time when it really did seem like the chess world was going to break apart."

~Ken Regan

In Regan's Honda Accord, we talk about how his chess work has spawned non-chess-related ideas, from how to use computers to grade massive open online courses, to how to think about the future economy. Tyler Cowen, Regan's childhood friend and an economics professor at George Mason University, is the author of *Average is Over*, which came out in 2013, and Cowen fills a chapter with predictions extrapolated from Regan's research. Cowen reports how freestyle (human-computer) chess teams play stronger than computers do on their own and argues that the future economy will consist of high-performing human-computer teams in all aspects of society. Regan takes pride in playing a prominent part in his friend's book.

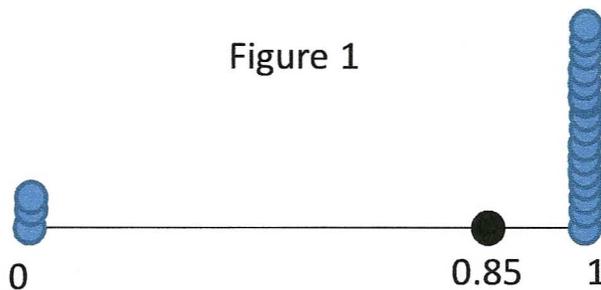
Randomness affects all aspects of Regan's life. His wallet oozes scraps of paper that contain names, numbers, and reminders. He doesn't own a smartphone. When we enter his office, unopened boxes crowd the floor, and spewed across every shelf and workspace lie papers, stacks of books, piles of notebooks, an ancient monitor, a 90's-era radio, and milk crates full of ephemera. A few months earlier, Regan moved to a new building constructed by the university and he claims he hasn't had time to unpack. A clean spot the size of two cafeteria trays makes room for a monitor and keyboard. On another small clearing, conspicuously placed across from us, sits the only item in the room besides the computer equipment to have received Regan's apparent care: a framed portrait of his wife.

A tab on Regan's browser is open to a fantasy baseball site. He loves baseball, and he was watching the 2006 baseball playoffs and logged into *PlayChess.com*, an online chess server, when he first heard about the Kramnik forfeit.

Regan feels a responsibility to do for professional chess what steroid testing has done for professional baseball. The Mitchell Report was commissioned in 2006 to investigate performance enhancing drug (PED) abuse in the major leagues, around the time Regan began his anti-cheating work. While baseball enters its post-PED era, FIDE has yet to put a single performance enhancing device—the chess world's PED—regulation into place. It wasn't until mid-2013 that the Association of Chess Professionals (ACP) and FIDE organized a joint anti-cheating committee, of which Regan is a prominent member. In mid-2014, the committee plans to ratify a protocol about how to evaluate evidence and execute punishment.

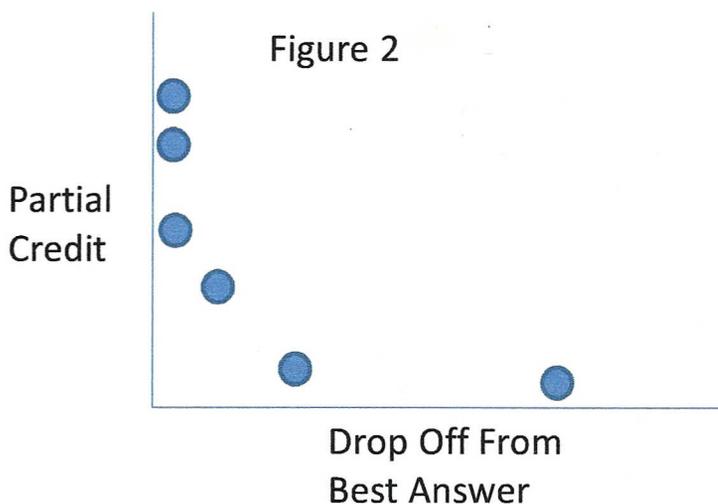
Regan clicks a few times on his mouse and then turns his monitor so I can view his test results from the German Bundesliga. His face turns to disgust. "Again, there's no physical evidence, no behavioral evidence," he says. "I'm just seeing the

Figure 1



On a standard 25-question multiple-choice exam, if a student answered 22 questions correctly and three questions incorrectly (shown here), then such a result would be equivalent to stacking 22 points at location 1 and three points at location 0. The number 0.85 represents the average or "best fit" location that summarizes the student's overall score.

Figure 2



Instead of stacking points at location 0 and location 1 on a number line, Regan distributes partial credit over a two-dimensional plot.

numbers. I'll tell you, people are doing it." Regan is 53. His hair has turned white. What remains of it, billows up in wild tufts that make him look the professor. When Regan acts surprised his thick, jet-black eyebrows rise like little boomerangs that return a hint of his youth. His enthusiasm for work never wanes; his voice merely shifts modes of erudition that make him *sound* the professor.

TO CATCH AN ALLEGED CHEATER, Regan takes a set of chess positions played by a single player—ideally 200 or more but his analysis can work with as few as 20—and treats each position like a question on a multiple-choice exam. The score on this exam translates to an Elo rating, a score Regan calls an Intrinsic Performance Rating (IPR). There are, however, three main differences between a standard multiple-choice exam and Regan's anti-cheating exam. First, on a standard exam each question has a fixed number of answers, usually four or five choices; on Regan's exam, the number of answers for each position equals the number of legal moves. Second, on a standard exam, one answer per question receives full credit, while the other answers receive zero credit; on Regan's exam, every legal move is given partial credit in proportion to how good it is relative to the engine's top choice. (Partial credit falls off as a complicated nonlinear relationship based on the engine's evaluations. Credit also abides by the constraint that all moves taken together for a position must sum to full credit.)

The third difference is the scoring method. (See Figure 1) A standard multiple-choice exam is scored by dividing the number of correct answers by the total number

of questions. This gives a percentage, which translates to an arbitrary grade like A, B-, C+, etc. What matters is not just the percentage but how one interprets the percentage. If a test is especially difficult and most students do poorly on it, then an 85 percent might translate to an 'A' rather than the more typical 'B'. This is called grading on a curve.

Figure 2 shows the conceptual relationship between a player's chosen moves for a set of positions and how an engine might distribute partial credit. Each point represents a move. Good moves fall into the top left corner of the plot, while poor moves fall into the bottom right. Since average players and grandmasters both make relatively poor moves compared to an engine, all human players' plots take on the same general L-shape. This method of converting engine evaluations into objective partial credit is the original aspect of Regan's work. He calls it "Converting Utilities into Probabilities." (Regan uses the technical term "probability" instead of "partial credit," because after the partial credits conform to the constraint that they must sum to full credit, they mathematically behave like probabilities.) "I made it up," he says. "I've been astounded, actually, that there doesn't seem to be precedent in the literature for it. I was dead sure people were doing this problem."

(Regan's literature search nourished his penchant for coincidence as well. As a serious Christian he sometimes gets asked if he believes in the theory of evolution, which he does. But, he says, "Intelligent Design papers featured large in my initial literature search. There's no direct connection to my work, but some of the mathematical ingredients are the same." Intelligent Design's leading complexity theorist is William Dembski, and Regan noted that his wife's old roommate's husband is Robert Sloan, chair of the computer science department at the University of Illinois, Chicago, where Dembski earned his Ph.D.)

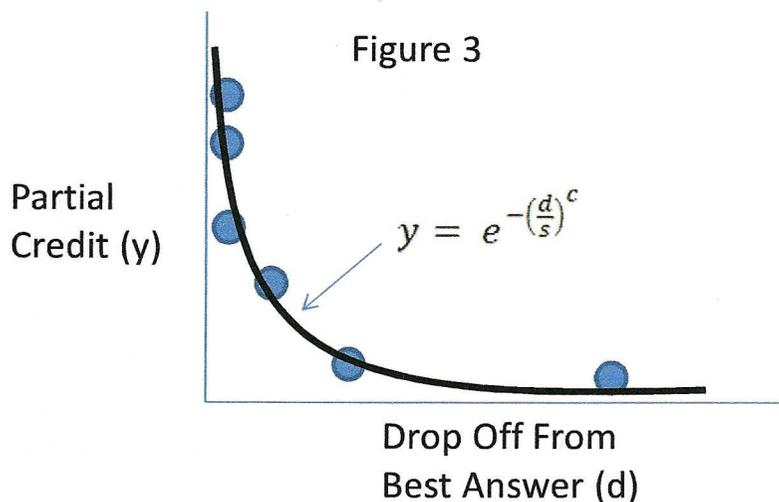
In Regan's algorithms it is the relative differences in move quality that matter, not the absolute differences. So if, for example, three top candidate moves are judged by the engine to be only slightly apart, then these top three moves will each earn approximately 30 percent credit (the remaining 10 percent left for the remaining candidate moves). This emphasis on relative differences rather than absolute value explains why cheaters who use moves that are not always the engine's first choice will still get caught. This also

explains why it's not possible for partial credit to be greater against weak opponents.

After a player's partial credit is plotted for a set of positions, Regan graphically scores his exam by drawing a curve averaged through the data (**See Figure 3**). (In statistical jargon, this process is called a "least squares best fit." The score on a standard multiple-choice exam can be thought of as a "best fit" too, but in this case its best fit is calculated between the points zero and one on a number line rather than between multiple points on a two-dimensional plot. **See Figure 1** again.) The best fit produces a curve (shown as 'y' in Figure 3) and two values, 's' and 'c,' which characterize the bend in the curve. Regan calls 's' the sensitivity. It shifts the curve left and right and correlates to a player's ability to sense small differences in move quality. Regan calls 'c' the consistency and it thins or thickens the tail of the curve. A larger 'c' represents a player's avoidance of gross blunders ("gross" being somewhat relative to the interpretation of the engine). Regan has found that different values of 's' and 'c' translate into well-defined categories that align with Elo ratings, similar to the way that a 95 percent and an 85 percent on an exam typically translate to an A and B, respectively. Back in the 1970s, when Arpad Elo designed the USCF and FIDE rating systems, he arbitrarily picked 2000 to mean expert, 2200 to mean master, etc. This arbitrary assignment means chess ratings are based on a curve, and specific values of 's' and 'c' can be mapped directly to specific Elo. The mapped rating is the Intrinsic Performance Rating.

It's more reliable to call someone, say, a B-player in chess than it is to call someone a B-student in school. A student can study for an individual test, but chess strength tends to change slowly. If Regan knows a player's Elo before subjecting the player's moves to an anti-cheating exam, he can compare how well each move's partial credit matches the typical partial credit earned by a player with that Elo. Regan represents this difference as a z-score, which is a fancy name for the ratio of how many standard deviations a player's test performance is from that player's typical Elo performance. The greater the z-score, the more likely a person has cheated. (**See Figure 4**)

The IPR and z-score are two separate results that emerge from the same test, but the z-score is much more reliable. If Regan were to compute an IPR with only a few moves, it would be like marking an exam with very few questions. This would translate to an unreliable letter grade. The z-score, however, is more accurate. "The IPR does not have forensic standing," says Regan. "But the cheating test [z-score] is based on settings that come from training 8,500 moves of world championship games." These moves act like questions



Instead of averaging between two locations on a number line and finding a "best fit" point like in Figure 1, Regan's scoring method finds a "best fit" curve between distributed point locations. Each player's moves result in a unique curve (shown by expression 'y') that can be characterized by the parameters 's' and 'c'.

<u>IPR</u>	<u>s</u>	<u>c</u>
2700	.082	.515
2300	.118	.485
1900	.154	.455

Figure 4

the College Board uses to normalize its scoring on standardized tests. For example, if the College Board wanted to catch a cheater on the SAT, it could easily do so by analyzing a small sample of suspicious answers to questions it knows to be difficult. Cheaters would perform uncharacteristically well on these questions. The same red flags go up when a cheating chess player consistently receives more partial credit on each move than his Elo would predict he deserves.

Because the proper construction of statistical evidence against alleged cheaters requires such technical expertise, Regan believes that it's necessary to establish a centralized authority responsible for the administration of anti-cheating protocol. Eventually he would like to oversee the conversion of his 35,000 lines of C++ code into a Windows-driven program or portable app. "I see other people using my methods but not necessarily using my program," he says. Regan also believes that a centralized authority can best fix public confusion about what constitutes scientific versus unscientific procedure. It's too easy for people with a poor methodology to spread rumors online.

The most notorious public cheating case to date has been that of the then-26-year-old Bulgarian Borislav Ivanov. He was first accused of using computer assistance in December, 2012, at the Zadar Open in Croatia, where, barely a 2200-player, he scored six out of nine in the Open section, including wins over four grandmasters. Allegedly he had cheated in at least three open tournaments before that, too. Finally, Ivanov was disqualified from both the Bladovgrad Open, in October, 2013 and the Navalmoral de la Mata Open in December, 2013, after both times refusing the inspection of his shoes, where he had allegedly hid a wireless communications device.

The Ivanov case was widely publicized in the Bulgarian media and at the news site *ChessBase.com*, which prompted amateur bloggers and *You Tube* aficionados to post their own move "matching" analysis, but none of it was worthy enough or contained high enough confidence intervals to persuade the Bulgarian Chess Federation to take action. Regan's analysis, however, found that Ivanov's moves earned a z-score of 5.09, which translates to the odds of him independently making these moves to less than one chance in five million. Regan's statistical evidence, along with Ivanov's refusal to submit to a search, resulted in the Bulgarian Chess Federation suspending Ivanov for four months.

Statistical evidence is immune to concealment. No matter how clever a cheater is in communicating with collaborators, no matter how small the wire less communications device, the actual moves produced by a cheater cannot be hidden.

Nevertheless, non-cheating outliers happen from time to time, the inevitable false positives. In any large open tournament with at least a thousand non-cheating players, the chances are very high that at least one of those honest players will earn a z-score of 3.0 or more, an ostensibly suspicious value. Tamal Biswas, one of Regan's two graduate students and a class-A player, has used a database of previously played games to run simulations of large open tournaments and verify these numbers.

By the summer of 2014, the ACP-FIDE anti-cheating committee hopes to work out the logistical details about what amounts and combinations of statistical, physical, and behavioral evidence should be considered conclusive if an alleged cheater is not caught red-handed. Regan proposes that a single z-score above 5.0 (the threshold for scientific discovery) or multiple instances of slightly lower z-scores should be enough statistical evidence on their own. But in other cases, one would need supporting behavioral or physical evidence, such as suspicious behavior in the restroom or tournament hall.

... one may conclude that Hikaru Nakamura's peak FIDE rating of 2789 beats Bobby Fischer's peak of 2785 for best American chess player of all time ...

Regan grew up a chess prodigy during the 1960s and early 1970s, a few miles outside New York City, in Paramus, New Jersey.

This area swarmed with chess opportunities and, in 1973, at the age of 13, Regan earned the master title, the youngest American at the time to do so since Bobby Fischer. A photo of Regan from that time shows a boyish round face and the thick, black eyebrows he maintains today.

But before Regan finished high school,

mathematics proved too alluring, and he decided he didn't want to make chess a career. His final two competitive triumphs came in 1976, when he was the only non-Soviet to win a gold medal at the now-defunct Student Chess Olympics, and in 1977 when he co-won the U.S. Junior Championship. After graduating from Princeton and Oxford, and then serving a post-doc at Cornell, Regan was hired by the University at Buffalo in 1989, where he has worked ever since. During the 1990s to 2006 Regan didn't think much about chess. His kids were young, and he was busy immersing himself into the study of P versus NP, the holy grail of computer science problems. He now "leads three lives" as he likes to say: his main research and teaching duties, his anti-cheating work, and as co-author (with Richard J. Lipton) of the blog, "Godel's Lost Letter and P=NP." In December of 2013, Springer published a book Regan co-wrote with Lipton about the blog, titled *People, Problems, and Proofs*.

The blog publishes not only technical amusements but occasional fodder about coincidences. "[MIT Professor] Scott Aaronson bet \$100,000 that scalable quantum computing can be done," says Regan. "The media picked up on this. The impetus for this bet was my post entitled 'Perpetual Motion of the 21st Century.' But my post was edited by Lipton. Lipton, Lance Fortnow, and I co-wrote some papers in the early 1990s, and Fortnow co-writes his own blog with Bill Gasarch; and Bill Gasarch is a friend of mine and one of my confidants because he is also Christian." At times Regan goes on like this, and it can be argued that his advanced research requires less energy to follow than his personal connections.

P versus NP stands for Polynomial time versus Nondeterministic Polynomial time. A P-type problem requires relatively few computations, like the solution to tic-tac-toe or 8x8 checkers. Computations for an NP-type problem scale up extremely quickly, however—too quickly to find a solution based on the current theories of computer science. (An example would be the Travelling Salesman problem, where the goal is to find the shortest tour between a large number of cities.) Regan's research includes ways to reduce the number of computations in an NP-type problem, so it behaves more like a P-type.

If Regan manages to prove the theoretical equivalence of P- and NP-type problems—the meaning of "P=NP" in his blog title but an unlikely event, not because of a lack of technical proficiency on Regan's part but because of the general consensus in the field that the relationship is false—then the result would change the world: cryptographic techniques would become obsolete, perfect language translation and facial recognition algorithms would become possible, and there would be a tremendous leap in artificial intelligence. For good reason, such a discovery would earn him the \$1 million Millennium Prize.

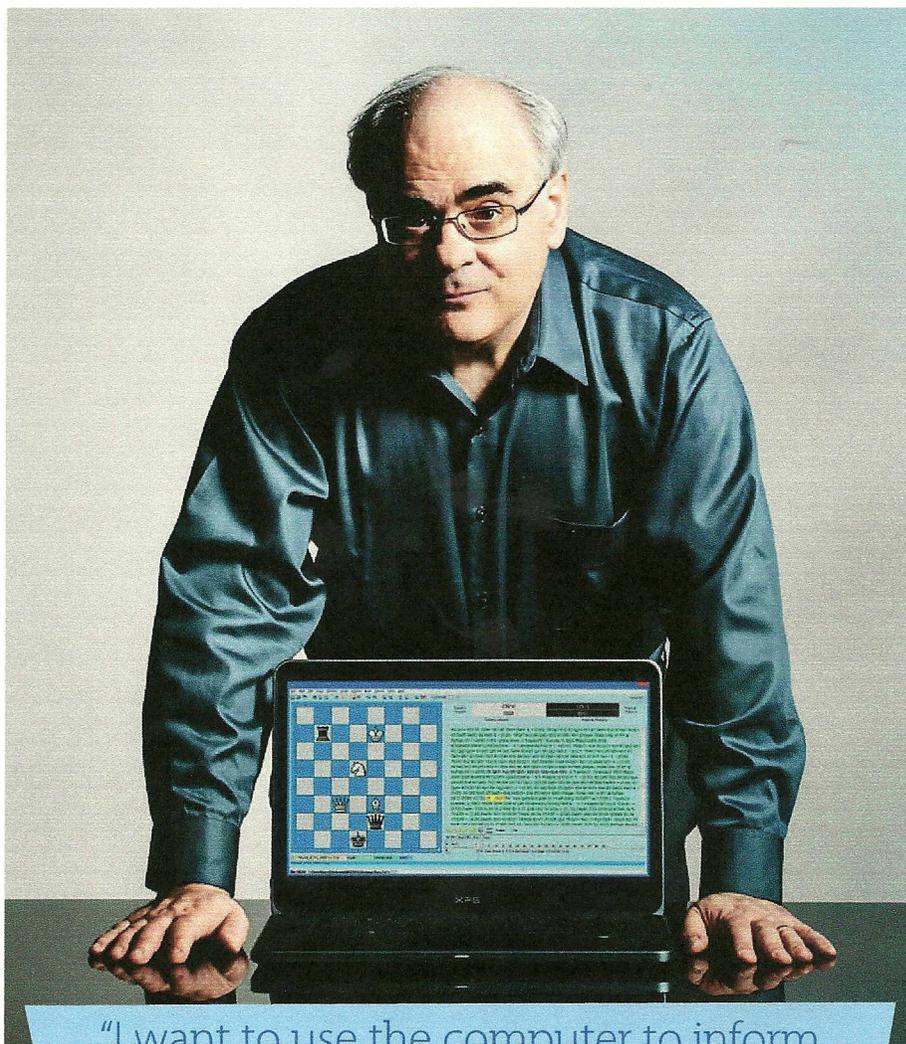
The solution to chess is not defined as an NP-type problem (although some vari-

speaker inside, similar to the way a chess engine appears to conceal a mini super-grandmaster. Searle argued that that when the English-speaking person (or a computer) follows a set of instructions to translate a language, no matter how well, they do not understand the language in the same way a native human speaker understands language. The same skepticism can be applied to whether or not computers understand chess.

Chess has often been described as a form of language, and when I propose to Regan that today's chess engines approach perfect play by following a set of rules embedded in source code, similar to the way the translator inside the Chinese Room follows a flowchart, he carefully considers his response. The implication that the best chess-playing entities on the planet follow rules revisits the ongoing debate in the chess community about whether or not human chess players also use rules to evaluate positions. Ironically, chess computers are commonly believed to play in a style that ignores rules.

Regan speaks English, Spanish, German, Italian, and French, and he approaches the debate by distinguishing rules written in human language from those written in computer code. "When we get to the tunable parameters in the program," says Regan, "all of the magic constants that define the value of the queen, the value of a rook, the value of a knight, the value of certain positional play, the values of squares, of attacks, these parameters are tuned by performance, linear regression. Programmers don't necessarily have a theory about what values or rules for those parameters work well. They have a general idea, but the final values are determined by [the engine] playing lots of very fast games against itself and seeing which values perform best." When I insist that the ones and zeroes of an engine's compiled code remain static, similar to rules written in a book, he leans back and restates his point. "Yes, that's true. But computers use regression."

What Regan means by regression is this: While some ones and zeroes remain static in the engine's initial program, other ones and zeroes essential for the engine's evaluation function—those essential for the way it "thinks"—rapidly update in short-term random access memory (RAM). This process mimics training and enhances a computer's ability to do more than just calculate. Regression creates real-time feedback that allows engines to "think" about each position unburdened by context, similar to the way a human weighs imbalances. But computers calculate much faster. Deep Blue, the first computer to defeat a world champion in a standard time control match, succeeded despite its relatively poor evaluation function and made up for this deficiency via fast



"I want to use the computer to inform us about the human mind."
~Ken Regan

calculation. Today's top engines would destroy Deep Blue, because they evaluate better—because, ironically, they "think" more like a human.

"[ALAN] TURING WANTED TO MODEL human cognition with a computer, but I'm going in the opposite direction," Regan says. "I want to use the computer to inform us about the human mind." Regan's data has reproduced a result in psychology first discovered by Nobel Prize-winning economist Daniel Kahneman and colleague Amos Tversky, which states that human perception of value is relative. "You'll drive across town to save \$4 on a \$20 purchase, but you wouldn't do it for a \$2,000 purchase," says Regan. His data shows that players make 60 percent to 90 percent more errors when half a pawn ahead or behind than when the game is even. Regan claims that this is an actual cognitive effect, not a result of high-risk-high-reward play, because it is observed with players who have both the advantage and disadvantage.

Chess has been called the *drosophila* (a small fruit fly, used extensively in genetic research because of its large chromosomes, numerous varieties, and rapid rate of reproduction) of artificial intelligence. It is a popular resource for research in cognitive science and psychology, because the Elo rating system provides an objective measure of human skill. Regan's work follows this scientific tradition. He has processed over 200,000 reference games played by players ranging in Elo from 1600 to 2800, using Rybka 3 at depth 13 in single-line mode. Single-line mode is a bit less accurate than multi-line mode, but it runs roughly 20 times faster. These reference games provide a

ants played on boards larger than 8x8 are), but it shares two characteristics: 1) it is practically impossible to prove a solution—for example, to prove a win or draw for White from the initial position; and 2) we can quickly verify a solution—whether or not a particular chess position is a checkmate. The main difference between chess and NP-types is that the solution to chess is theoretically possible, whereas solutions to NP-type problems currently are not. In one way, however, chess can be marked more difficult than NP-types, because with NP-types one can theoretically verify a solution at the start if there is one. To find the solution to chess, one can only compute deeper and deeper.

Claude Shannon, the father of information theory, in his famous paper “Programming a Computer for Playing Chess,” estimated the number of possible unique chess positions to be roughly 10^{43} . “It’s impossible to unpack the complete game tree,” says Regan. “It’s so large that if those bits were placed in an efficient memory device the size of a room, that room would collapse into a black hole.” Regan classifies chess as a Deep problem, “One where I can describe the complete set of rules in a small amount of information, but where unpacking the information will take a long time.”

Chess engines continue to improve at about 20 Elo points per year. If Regan’s estimate of perfect play at 3600 Elo is true, then they will arrive there within a few decades. Regan believes they already play perfectly on occasion, if given enough time to “think.” A chess computer with a good enough algorithm and fast enough processor does not need to store 10^{43} positions to play with the same skill as a computer that does. To understand how such an astonishing feat is possible, consider how it’s possible for a human to play perfect tic-tac-toe without having to store the complete solution to tic-tac-toe. There are 256,168 possible different games of tic-tac-toe, but a little smarts reduces this number to 230 strategically important positions.

IN SEPTEMBER, 2013, HARVARD University hosted the one-day New England Symposium for Statistics in Sports, and Regan decided to attend on a stopover while on his way home from another conference. “I’m not going for the talks so much as to hobnob and buttonhole people,” he wrote to me a few weeks before the event. We met at the bar of the Grafton Street Pub, a crowded restaurant near Harvard Square, for the symposium’s social hour. The din of Saturday night beset normal conversation, and I found Regan leaning into the voice of Eric Van, a 50-something statistician who consulted for the Boston Red Sox between 2005 and 2009. Van was explaining to Regan how the Sox needed to shuffle their lineup to win the World Series, a task Van helped the team achieve in 2007 and one in which they eventually accomplished again a month after this get-together. Regan had come to the conference to form connections, and Van’s attachment to baseball made this one particularly sweet.

But the whole bar scene injects a bit of anxiety into Regan’s body language. He blinks hard at times and chews his gum vigorously. (Later Regan would tell me, “Chess got me comfortable in an adult world. I was able to step right off the boat my first year as a graduate student at Oxford and feel confident.” It was during this time at Oxford when Regan also met his wife.) Regan offered to buy us drinks, in a tone of voice that implied this wasn’t a question he often asked but which he felt was the obligatory thing to do. Nobody

accepted. So he bought a pizza to share, and we moved to a quieter spot.

Van battles attention deficit disorder and narcolepsy, and now spends his time as a private scholar who researches these ailments and works on a theory of consciousness. The conversation turned to the intersection of cognitive science and chess, which naturally led to a discussion about the hypothetical Chinese Room Thought Experiment first proposed by philosopher John Searle.

In this experiment, an English-speaking person sits in a locked room. After a question written in Chinese is slipped under the door, the person follows rules on a flowchart that describes how to write an answer in Chinese. The person then slips the answer back under the door. It would appear to people outside the room that there is an intelligent, native Chinese

A QUICK AND DIRTY HISTORY OF NOTORIOUS CHEATING INCIDENTS

1993 WORLD OPEN (PHILADELPHIA)

An unrated player using the pseudonym “John Von Neumann” scored 4½/9 in the Open section, including a draw against GM Helgi Olafsson. “Von Neumann” was disqualified after he refused a request by a suspicious tournament director to solve a simple chess puzzle.

1999 BOBLINGER OPEN (GERMANY)

55-year-old and 1925-rated Clemens Allwerman scored 7½/9 to win the tournament ahead of multiple titled players. Subsequent analysis using the then-current Fritz engine roused tremendous suspicion, but Allwerman was never disciplined.

2006 SUBROTO MUKERJEE MEMORIAL OPEN (INDIA)

1933-rated Umaket Sharma caught with a Bluetooth device in his hat. Sharma was suspended 10 years by the All India Chess Federation.

2006 WORLD OPEN (PHILADELPHIA)

1974-rated Steven Rosenberg disqualified for using a Phonito, a hearing aid-sized device that slips into the ear, and which can send and receive wireless communications.

2010 FIDE OLYMPIAD (KHANTY-MANSIYSK)

19-year-old GM Sébastien Feller, GM Arnaud Hauchard and IM Cyril Marzolo caught performing an elaborate move-relaying scheme. Marzolo, who would be home at his computer, would text Hauchard in the playing hall, who would then communicate moves to Feller, based on where he was standing. According to Wikipedia, the French Chess Federation’s suspension of these three players was revoked, but Feller is currently serving a 33-month suspension from FIDE.

2011 GERMAN CHAMPIONSHIP

FM Christoph Natsidis used an engine running on his smartphone, while in the bathroom. Natsidis was disqualified from the tournament.

2012 ZADAR OPEN (CROATIA), 2013 BLADOEVGRAD OPEN (BULGARIA), NAVALMORAL OPEN (SPAIN)

26-year-old Borislav Ivanov scored 6/9 at the Zadar Open, including wins over four grandmasters. When Ivanov refused inspection at Bladoevgrad and Navalmoral after suspicious behavior and wins over multiple grandmasters, he was forfeited both times. Eventually Ivanov was suspended for four months by the Bulgarian Chess Federation.

Figure 5

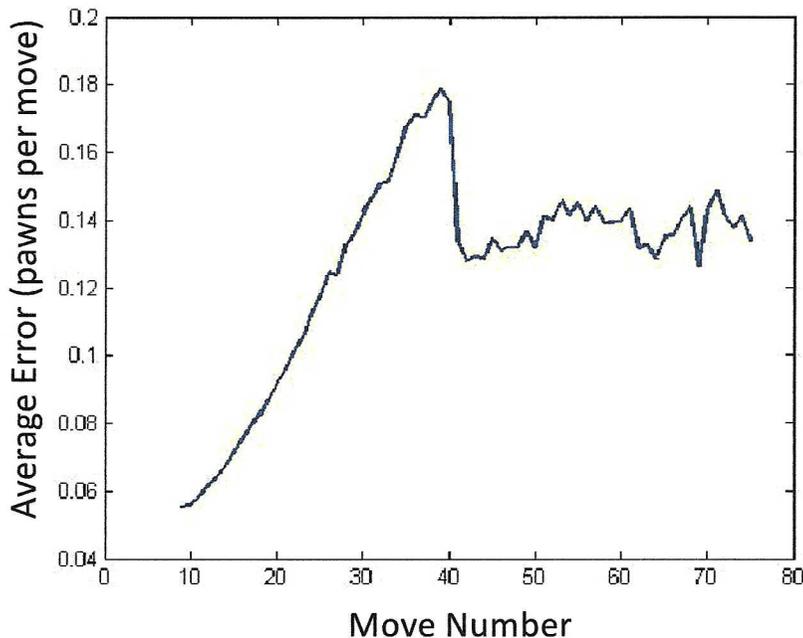


Figure based on 3.6 million moves from all Category 11 and higher tournament games between 1971 and 2011.

Figure 6

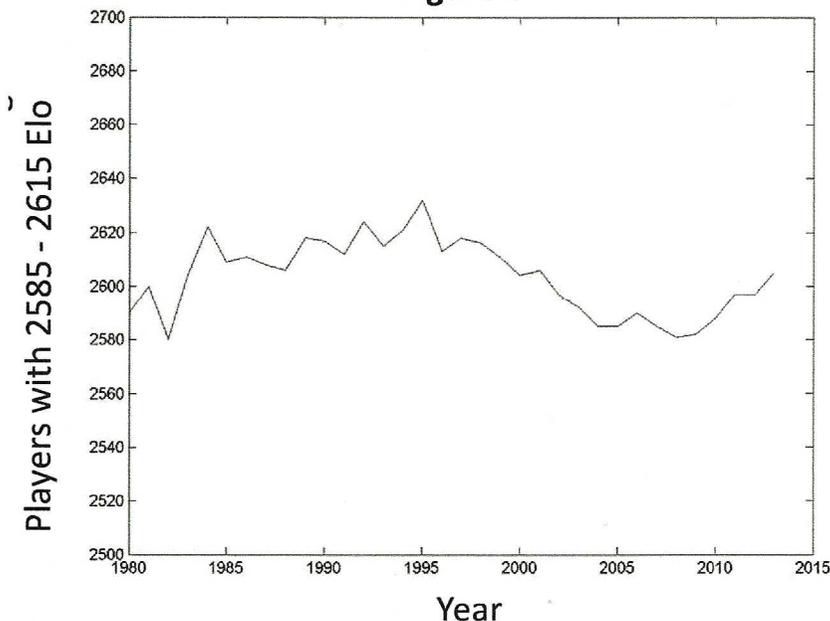


Figure based on 120,000 positions between players with Elo 2585-2615 played between 1979 and 2014.

"What's the rating of perfect play?
My model says it's 3600."

rich set of data with which to create all sorts of chess-based applications.

In 2012, FIDE sold the marketing and licensing rights of professional chess to AGON, a company run by Andrew Paulson. According to the *New York Times*, "[Paulson] wants to turn chess into the next mass-market spectator sport." Paulson plans to supplement Internet coverage of major competitions with something he calls ChessCasting, a broadcast of not only moves, commentary, video, and live engine evaluations, but also biometrics such as a player's pulse, eye movements, blood pressure, and sweat output. Regan's work adds many non-invasive statistics to this list. "The greatest immediate impact on the professional chess world that I think I'm going to have, besides my anti-cheating work, is that I'm going to come up with a statistic called 'Challenge Created,' which is going to be an objective way to single out the players who create difficult problems for their opponents." The greatest over-the-board practical problems are not always caused by the objectively best moves, and Regan's metric can quantify this distinction.

Other statistics that emerge from Regan's IPR calculation include ways to visualize the degradation of move quality during time pressure (in **Figure 5**, notice how error increases as the move number approaches 40, the standard time control) and a way to normalize the different chess rating systems of the world. Amateur players constantly wonder how, say, their *Chess.com* rating compares to their national federation's rating. IPRs provide a way to standardize this procedure. In some ways, IPRs are even more accurate than traditional ratings, because they're calculated on a per-move basis rather than on a per-game basis. One bad tournament could sink a traditional rating, but if this bad tournament was the effect of only, say, three isolated bad moves, then such bad luck would not detrimentally affect an IPR. Regan does admit, however, that engines bias their evaluations ever so slightly against human-like moves, and this effect "nudges IPRs slightly out of tune." The exact reason for this tiny bias is unclear and it obsesses Regan during his free time.

For the improving player, IPRs can be used as training metrics for different phases of the game. Say a person wants to obtain an objective measure of how well they play middle games out of the Ruy Lopez versus how well they play middle games out of the Scandinavian. All they would need to do is isolate the particular moves and positions of interest, send them through Regan's IPR-generator, and they have a performance metric. This method has been used by Regan to rate historical players. For years, statistician Jeff Sonas has been rating historical players, but Regan's IPR is more objective. Sonas uses

historical game results, which provide information about relative performance only within eras. Only players alive during the same period can play each other. But since Regan's method compares moves to a common standard (the engine), rather than the results of games, he can objectively relate player abilities across eras. What he found was that rating inflation does not exist. Between 1976 and 2009, there has been no significant change in IPR for players at all FIDE ratings. Figure 5 shows, for example, how the IPR for players rated between 2585 and 2615 has remained relatively constant over time. Today's thousands of grandmasters and dozens of players rated over 2700 indicate a legitimate proliferation of skill. Thus one may conclude that Hikaru Nakamura's peak FIDE rating of 2789 beats Bobby Fischer's peak of 2785 for best American chess player of all time, and Magnus Carlsen's peak rating of 2881 places him as the best human chess player of all time. (See Figure 6)

WHY DO WE FAIL TO UNDERSTAND those who cheat? In the journal *The New Atlantis*, Jeremy Ruzansky writes, "Performance-enhancing drugs are a type of cheating that does not merely alter wins and losses or individual records, but transforms the very character of the athlete. ... If our entire goal were to break

pitching records in baseball, we could build pitching machines to pitch perfect games. It is worth asking why we would never do this, why we would never substitute our sportsmen with machines, even though machines could easily achieve superior performance."

Ruzansky's answer is that we value statistics only as the result of superior human performance. Countless athletes and chess players, including Bobby Fischer, have compared sports to life. "Chess is life," the former American world champion said. Sports provide society with a metaphor for the competition inherent in life, and this metaphor works only when a living person competes—or, in chess, when a living mind contemplates the complexities of the moves.

Yet cheaters look upon their act as its own kind of sport. In *The Journal of Personality and Social Psychology*, researchers found that cheaters enjoy the high of getting away with their wrongdoing, even if they know others are aware of it. Boris Ivanov, for example, continued to cheat after he was caught but before he was suspended. Behavior like Ivanov's poses a great threat to tournament chess, because it doesn't take much risk to reap reward. Faced with a complex calculation, a player could sneak their smartphone into the bathroom for one move and cheat for only a single critical position. Former World Champion Viswanathan Anand said that one bit per game, one yes-no answer about whether a sacrifice is sound, could be worth 150 rating points.

"I think this is a reliable estimate," says Regan. "An isolated move is almost uncatchable using my regular methods."

But selective-move cheaters would be doing it on critical moves, and Regan has untested tricks for these cases. "If you're given even just a few moves, where each time there are, say, four equal choices, then the probabilities of matching these moves become statistically significant. Another way is for an arbiter to give me a game and tell me how many suspect moves, and then I'll try to tell him which moves, like a police lineup. We have to know which moves to look at, however, and, importantly—this is the vital part—there has to be a criterion for identifying these moves independent of the fact they match."

Although none of these selective-move techniques have yet to be discussed with the ACP-FIDE anti-cheating committee, Regan has confidence they'll work. But he keeps his optimism restrained. He doesn't look forward to the leapfrogging effect bound to happen between cheaters and the people who catch them, a phenomenon that has invariably plagued other sports. Other challenges remain, too. A new "depth"

parameter to model the number of plies a player evaluates is being researched to join 's' and 'c'; the standard engine is being converted from Rybka 3 to Houdini; and the ever-present but minimal anti-human bias in engine scores must be cancelled.

In 2012, Regan lost an exhibition match to a Lego-built robot running the Houdini engine, equipped with an arm that moved the pieces on a real board and a camera that could interpret the position. The experience made an awesome impression on him. "Is technology going to be so ubiquitous that we'll not be able to police it anymore?" he asks while he, his wife, and I eat dinner at a local Thai restaurant. Regan slumps over his food, looking depressed about the need to even ask the question. "Houdini won using only six seconds per move," he says. The exhibition reminds Regan that his calling has carved valuable time from his research and family. "He's obsessed," says his wife, who sits across the table. Then she adds, "But you've got to be obsessed to be good." Regan ignores the flattery, his attention held by an emerging thought. Finally he springs forward in his chair, smiling. "By the way," he says. "This project was run by a person whose mother and my mother share a best friend back in New Jersey." ♣

See Dr. Regan's website www.cse.buffalo.edu/~regan/chess/ for more of his work.

USCF MEMBERS WEIGH IN ON CHEATING

TIM JUST, NATIONAL TOURNAMENT DIRECTOR

The perception of an opportunity to cheat far outweighs the actual act of cheating. Honest players want those opportunity doors closed. The challenge is how to do that without invoking the "law of unintended consequences." And do it without raising the costs of running chess tournaments. In the "bathroom scenario" totally locking the restroom door solves the problem, but there are all sorts of unintended consequences to that anti-cheating method. Or do we inspect every player wanting to use the bathroom? How about forcing players to hand over their electronic gizmos before they can use the bathroom? Again, unintended consequences will rear their ugly heads. The challenge is eliminating as many opportunities for cheating as possible while still respecting the rights of all players at a reasonable cost. What individual rights are players willing to give up to close the cheating opportunity door? How much money are players willing to pay to close the opportunities door? What conveniences are wood pushers willing to let go of to limit cheating opportunities? For someone convicted of cheating the penalty should be banishment from the USCF.

GM ROBERT HESS

In general, what the Chess Club and Scholastic Center of Saint Louis does is straightforward and good. They wand people before they enter the playing area and you are not allowed to bring your phone upstairs. It seems pretty simple. Also, I think cheating is typically less of an issue in elite events, so at the tournaments where cheating is to occur, generally there is less security.

FRANK BERRY, INTERNATIONAL ARBITER

In an Open event there is nothing you can do. In invitational events—according to IM Jack Peters—you have to be aware enough to invite only players with ethical reputations.