Chess and Antirealism

Samuel Kahn

In this paper I make a novel argument for scientific antirealism. My argument is as follows: (1) the best human chess players would lose to the best computer chess programs; (2) if the best human chess players would lose to the best computer chess programs, then there is good reason to think that the best human chess players do not understand how to make winning moves; (3) if there is good reason to think that the best human chess players do not understand how to make winning moves, then there is good reason to think that the best human theories about unobservables are wrong; therefore, (4) there is good reason to think that the best human theories about unobservables are wrong. The paper is divided into three sections. In the first, I outline the backdrop for my argument. In the second, I explain my argument. In the third, I consider some objections.

Keywords: philosophy of artificial intelligence; philosophy of chess; philosophy of science; scientific realism; scientific antirealism; unobservables

Section 1 Terms and Background

I take the debate about scientific anti/realism to be about the following proposition:

REAL There can be justified beliefs about unobservables.

Realists say that REAL is true whereas antirealists say that it is false.[1] There are four aspects of this debate that require remark.

First, some philosophers question whether the un/observable distinction is well-defined.[2] I cannot settle this here; for present purposes I assume there is some such distinction (even if not a sharp one). I take subatomic particles like protons and neutrons to be paradigmatic unobservables, and I say that a physical entity E is unobservable if a human with statistically normal sensory capacities would not be able to distinguish E from other entities without the use of technology.[3]

Second, the anti/realism debate cannot be settled by appeal to radical skepticism. The contemporary scientific anti/realism debate grew out of debate about how and whether theory terms can have genuine meaning and reference.[4] The debate now focuses on theoretical (unobservable) entities (rather than terms) and belief and justification (rather than meaning and reference). But because of its provenance, the anti/realism debate is orthogonal to the radical skepticism debate.[5] That is, although they disagree about REAL, the antirealist and the realist agree about (or agree to accept for the purposes of debate) the following proposition:

OBS There can be justified beliefs about observables.

---

[1] (Chakravartty, 2017). See also (Boyd, 2002): "Scientific realists hold that the characteristic product of successful scientific research is knowledge of largely theory-independent phenomena and that such knowledge is possible (indeed actual) even in those cases in which the relevant phenomena are not, in any non-question-begging sense, observable."

[2] (Maxwell, 1962) is the *locus classicus* for this. For more recent (and more sympathetic) discussion of the distinction see (Muller, 2004) and (Muller, 2005).

[3] Further refinement is needed in order to clarify that (a) mathematical and linguistic "entities," although perhaps unobservable in a strict sense, do not fall within the purview of REAL, and (b) beliefs like "a neutron is unobservable," "I cannot have justified beliefs about neutrons," "I am not observing a neutron right now," "a neutron is not the size of a marble," "a neutron is not a square circle," and "if one neutron is added to another neutron, then there will be two neutrons," do not fall within the purview of REAL. For present purposes I am operating on the nontrivial assumption that this refinement is possible. (Some might maintain that making my condition a biconditional and dropping the word 'physical' will suffice for (a). But it certainly will not suffice for (b).)

[4] (Wedberg, 1975) and (Carnap, 1975).

[5] This also helps to explain why REAL is in terms of unobservables rather than the (merely) unobserved.

This is the author's manuscript of the article published in final edited form as:

Kahn, S. (2023). Chess and Antirealism. Asian Journal of Philosophy, 2(2), 76. https://doi.org/10.1007/s44204-023-00118-7

The radical skeptic, in rejecting REAL *and* OBS, would not be engaging in the scientific anti/realist debate; a successful antirealist argument will not impinge on OBS.[6]

Third, there are no justified beliefs without a believer. This is a nontrivial aspect of the anti/realism debate that I think has been neglected.[7] The debate is not about whether an omniscient being would have knowledge about unobservables: plainly they would (or they would not be omniscient), and if justified belief is part of knowledge, then they would have justified belief too. For present purposes I say that REAL is true if but only if the believer is a human with statistically normal mental and sensory capacities.

Fourth and finally, what counts as justification is contentious. The scientific anti/realism debate is supposed to be independent of questions about internalism and externalism in the sense that if, for example, reliabilism is correct, and if normal humans have reliable mechanisms wherewith they can form beliefs about unobservables, then REAL is true even if this justification is not (internally) accessible.[8] But the scientific anti/realism debate is not supposed to be neutral about things like testimonial justification: testimony might be a genuine justifier in general, but the fact that a student can have justified beliefs about unobservables on the basis of authoritative testimony would not be taken to show that REAL is true unless it also could be shown that the original affiant has *non*-testimonial justification.[9]

Section 2 My Argument

I call my argument the chess argument. It is as follows:

1. The best human chess players cannot win against the best computer chess programs.
2. If (1), then there is good reason to think that the best human chess players do not understand how to make winning moves.
3. If the consequent of (2) is true, then there is good reason to think that the best human theories about unobservables are wrong.
4. Therefore, there is good reason to think that the best human theories about unobservables are wrong.

It will be noted that there is a gap between (4) and the negation of REAL. I cannot try to bridge that gap here. To do so would require me to take a stand on many of the contentious issues remarked upon in section 1 (and others besides). That would distract from my main goal, which is to introduce a genuinely novel line of reasoning into the debate: I am trying to make a novel argument for antirealism, but that is not the same as making a decisive case for antirealism. Because of this, and because the conclusion of the chess argument is notable even with this gap unbridged, I think that this strategy is warranted.

The chess argument is deductively valid. So, it will suffice to defend each of the premises. But, before I do that, I want to discuss, briefly, some of the historical background for this argument: I am not the first to make a connection between chess and science in the philosophy of science in general or in the anti/realism debate in particular.

---

[6] Thus, the only way for a radical skeptic to take part in the anti/realist debate would be to explain why OBS at least seems plausible.

[7] I have tried to make some inroads into this issue in (Kahn, 2020).

[8] (Newman, 2010).

[9] It is usually assumed that the original affiant also must be a human with statistically normal sensory and mental capacities. But, there is room to question this assumption. For example, if there is testimonial belief about unobservables that is alleged to be from a divine being; if the divine being's ability to discern such things and willingness to impart them can be justified; and if the allegation that the affiant really was this divine being can be justified--would this suffice for REAL? The question, as far as I am aware, has not been discussed. Of course, antirealists might try to brush aside this kind of case, perhaps on the grounds that it rests on non-naturalist assumptions. But it seems to me that charting out such possibilities is important.

In *The Logic of Scientific Discovery*, Popper argues that methodological rules, in empirical science, are conventional, like the rules of chess:

> Methodological rules are here regarded as conventions. They might be described as the rules of the game of empirical science. They differ from the rules of pure logic rather as do the rules of chess, which few would regard as part of pure logic: seeing that the rules of pure logic govern transformations of linguistic formulae, the result of an inquiry into the rules of chess could perhaps be entitled 'The Logic of Chess', but hardly 'Logic' pure and simple…Just as chess might be defined by the rules proper to it, so empirical science may be defined by means of its methodological rules.[10]

In *The Structure of Scientific Revolutions*, Kuhn appeals to chess in order to make a point that is, in some ways, similar to Popper's. Kuhn thinks the rules of chess are analogous to the general paradigm that is in place during periods of normal science (which might include, although it is not limited to, Popperian methodological rules). More precisely, according to Kuhn, scientists who try to solve the scientific problems that arise during periods of normal science are like chess players. Kuhn argues that the scientific problem is analogous to a chess puzzle and the scientific paradigm is analogous to the rules of chess:

> Insofar as he is engaged in normal science, the research worker is a solver of puzzles, not a tester of paradigms. Though he may, during the search for a particular puzzle's solution, try out a number of alternative approaches, rejecting those that fail to yield the desired result, he is not testing the *paradigm* when he does so. Instead he is like the chess player who, with a problem stated and the board physically or mentally before him, tries out various alternative moves in the search for a solution. These trial attempts, whether by the chess player or by the scientist, are trials only of themselves, not of the rules of the game.[11]

Van Fraassen makes a very different use of chess in *The Scientific Image*. Unlike Popper and Kuhn, who focus primarily on the rules of chess, van Fraassen focuses on the complicated human psychology of individual chess players. Van Fraassen uses chess to illustrate how the aim of science, writ large, might be different from individual scientists' motives:

> The aim of science is of course not to be identified with individual scientists' motives. The aim of the game of chess is to checkmate your opponent; but the motive for playing may be fame, gold, and glory.[12]

There are two things that are notable about these appeals to chess. First, in all three of them, we can substitute other activities in for chess. Indeed, Kuhn himself mentions crosswords, jigsaw puzzles, and more, and

---

[10] (Popper, 2002, 32).

[11] (Kuhn, 1996, 144-145). Kuhn explains some of the specific parallels between normal puzzles and scientific puzzles in chapter IV, mentioning chess puzzles explicitly on page 38. For helpful discussion, see (Bird, 2000, 39, 88, and 101).

[12] (Van Fraassen, 1980, 8). As a former amateur competitive chess player, I feel compelled to comment that the motives of fame, gold, and glory seem (sadly?) out of place in this context. (There is, I think, also room to cavil with the claim that the aim of chess is to checkmate your opponent. But, neither of these points impugns van Fraassen's argument.)

commentators on van Fraassen have discussed his point using activities like pinball rather than chess.[13] As will be seen below, the same is not true of the chess argument: if there are other activities that could be substituted in for chess in the chess argument, they are fewer and further between, at least at present.[14]

Second, I want to flag the fact that the claims made by Popper, Kuhn, and van Fraassen in these passages are controversial. I am not engaging with that controversy because it is not important for my purposes.[15] I merely want to note some of the historical precursors of the chess argument. However, the absence of critical discussion of these precursors should not be interpreted as endorsement.

I turn now to a defense of premise (1) of my argument, the premise which says that the best human chess players cannot win against the best computer chess programs.

There are multiple powerful chess engines, including StockFish, Komodo, Fire, and Houdini. The most recent news (as of the writing of this article) is that Alphabet's chess engine, AlphaZero, is more powerful than StockFish, even when the latter is supplemented with what is called an "opening book" and even when the former is given significantly less time per move.[16]

One of the reasons this is notable, other than the fact that StockFish is one of the most powerful chess engines in the world, is that AlphaZero's programming is radically different from that of other chess engines. Other chess engines are programmed in terms of points associated with pieces and positions: their superiority over humans derives from their ability to perform very quickly and to remember the results of large numbers of calculations. For example, suppose that it is white to play. The engine will examine all possible moves that white can make; then it will examine all possible responses that black can make to each of these possibilities; then it will examine all possible responses white can make to each of these possibilities; and so on to a depth that is constrained only by computer power and time.[17] Thus, whereas human chess players rely on insight and understanding, chess engines (prior to AlphaZero) relied on a brute force approach.

AlphaZero still has this kind of computing power and speed. But, rather than being programmed in terms of points associated with pieces and positions, AlphaZero was programmed with the rules of how to move pieces and then a machine learning algorithm (called neural networking) that enabled it to learn strategy by playing

---

[13] For example:

> In the technical sense, 'the aim of science' means, to a first approximation, what we *should* expect science to achieve. Van Fraassen (1980, pp. 8–9) explains this by asking us to compare the aims of chess with those of chess players. Similarly, consider pinball. The aim of the game is to amass as high a score as possible (when 'possible' is suitably interpreted), although one might play simply because one enjoys the experience, or to distract oneself from troubling thoughts. These personal aims are at best auxiliary to the proper aim, and only count as auxiliary if they *contribute* to fulfilling the proper aim. (Rowbottom, 2011, 126)

[14] One might think that Go could be substituted for chess in the chess argument. The substitution seems especially promising because there is a neural network Go program, AlphaGo, that is the correlate of AlphaZero.

However, there are some important differences between chess and Go. Given current computing power, the brute force method is impracticable for an AI Go program (the chains are too long, and they have too many branches). Moreover, although a neural network Go program was developed--and seemed, in 2016, to establish the age of AI Go-dominance--in 2023 humans successfully deployed a winning strategy against top-level AI Go programs (Waters, 2023). So (at least at the time of this writing), if Go is substituted in for chess, then premise (1) of the chess argument would be false.

For useful discussion of other AI systems in human-computer gaming, see (Yin *et al.*, 2023).

[15] For discussion, see (Rosen, 1994, 146-147); (van Fraassen, 1994, section 4); (Rowbottom, 2011, 109-110, 122, and 126-127); (Rowbottom, 2014, section 2); and (Asay, 2019, section 3.3).

[16] (Lee, 2017).

[17] This is an oversimplification. The engines generally have ways of discounting some possible moves. Moreover, they do not examine all considered chains of moves to the same depth. But these details are, I believe, unimportant for present purposes.

millions of games against itself.[18] According to some, then, AlphaZero combines insight with brute force in a way never seen before.[19] This will become relevant momentarily.

Regardless, the superiority of chess engines over humans arguably dates back to 1996, when IBM's Deep Blue beat Gary Kasparov, the then best ranked chess player in the world, in a match. That was more than 25 years ago, and chess programming and computer processing power have increased exponentially since then, even discounting the innovative technique used with AlphaZero's programming.[20]

I turn now to premise (2), the premise that, if the best human chess players cannot win against the best computer chess programs, then there is good reason to think that the best human chess players do not understand how to make winning moves

I want to argue for premise (2) in two separate ways. But, before I do that, I need to say something about what a winning move is.

End-game puzzles often ask you to find moves that guarantee, typically in a three- or four-move chain, that you will checkmate your opponent. These are, perforce, winning moves. Along the same lines, the holy grail of chess, the "perfect chess game," begins with an opening that guarantees, no matter what move your opponent makes at any point, that you will win. It is not known (as of the time of this writing) whether this is possible. But, hopefully it is clear that, if the perfect chess game is possible, then these moves, like those in the end-game puzzles just alluded to, would count as winning moves.[21] However, I do not need such a strong concept for the sake of my argument: I think I can use a weaker concept, a concept such that there might be other kinds of winning moves, moves that are winning even though they do not guarantee a win, as well. Let me explain.

In poker, people sometimes contrast "playing the person" with "playing the odds." Proponents of the former approach use psychological tricks to beat their opponents rather than, say, counting cards.[22] Similar strategies might be used in chess, especially when one person knows their opponent well. In that sense, we might talk about a winning-move-against-X. But, that is not what I have in mind here when I talk about winning moves in chess.

A winning move in chess, for present purposes, is the start of a chain that marches, if not inevitably, then at least with ever more inevitability, toward a win. Most who have played chess competitively but only at an amateur level probably know the feeling of heightening dread, elicited as a game against an obviously superior opponent unfolds: it is clear that you are going to lose; it is clear that you can do nothing to avert the loss; but it is unclear why your position keeps getting incrementally worse–the comprehension lies just beyond (or perhaps far beyond) the horizon. This account of a winning move in chess is perhaps not as precise (or as determined) as some might like. But–and this is an important point to which I return below–given the current state of our knowledge, this, I think, might be the best we can do. Because the perfect chess game has not been discovered, our only litmus test for winning moves is looking at how games unfold, at which moves actually lead to wins. And this is why premise (1) is so important; it is our best evidence that computer chess programs are playing winning moves. I am going to elaborate on this below. But, for now, let me turn to the two arguments I want to give for premise (2).

---

[18] (Silver *et al.*, 2018).

[19] According to one grandmaster, it is as if "a superior species landed on earth and showed us how they played chess" (Lee, 2017).

[20] (Greenemeier, 2017).

[21] There is some room for argument here. As some use the term, the perfect chess game guarantees not losing rather than winning. In that case, the moves in the perfect chess game (which, to be clear, is not a single game but rather a set of possible games) would not be winning moves; they would be not-losing movings. But, these details are, I think, irrelevant for present purposes.

[22] Some might think that talk of "playing the person" is out of place when talking about AI chess. But, as recent events in the world of AI Go reveal, it is at least unclear that this is so (see note 14 above). The way in which humans won against AI Go programs was by deploying a strategy that effectively "distracted" the computer from an encirclement, a distraction that was based on a weakness in the software algorithms and that probably would not have worked against a similarly skilled human.

First, consider chess engines that use the brute force approach. There are two considerations that tell in favor of premise (2) on the brute force approach: (i) human "insight and understanding" and computer brute force lead to different moves; and (ii) humans cannot perform the kinds of analysis done by brute force chess engines.

Consideration (i) can be motivated in the same way that premise (1) was motivated. That is, if consideration (i) were false, then humans would win against chess engines, or at least draw. But chess engines consistently beat the best humans. So, consideration (i) is not false.

Consideration (ii) can be motivated by reflection on human cognitive limitations. Humans are not able to match the best chess engines either in breadth or depth of analysis.

Now, because human insight and understanding do not track computer brute force (consideration (i)), and because humans are not able to match computer brute force (consideration (ii)), humans are not able to explain or understand why a chess engine moves in the way that it does. But, from premise (1) we know that chess engines make winning moves. So, from premise (1) it follows that humans are not able to explain or understand winning moves, and that is all we need for premise (2).

However, I think that AlphaZero opens up a second and even more powerful way of arguing for premise (2). Consideration (i) still holds in the case of AlphaZero. But, now consideration (ii) must be replaced. I propose consideration (ii*): the kinds of analyses performed by AlphaZero are entirely opaque to humans. This can be motivated by appeal not only to the complexity of the calculations (both in breadth and depth as before) but also by appeal to the fact that we do not know even how the different positions are "scored" by AlphaZero (and, indeed, they might have been scored differently had AlphaZero played a different number of initial learning games). This, I think, makes the case for premise (2) stronger.

I turn now to premise (3), the premise that, if there is good reason to think that the best human chess players do not understand how to make winning moves, then there is good reason to think that the best human theories about unobservables are wrong.

In support of premise (3) I note that, for one thing, the universe is far more complex than a game of chess and, for another, so are the theoretical possibilities that can be explored in the inference from observables to unobservables.[23] I explain this in greater depth below. But, the point for now is that it is precisely the complexity of the game of chess that prevents us from understanding how to make winning moves. We can see this by comparing chess to tic-tac-toe or checkers: humans can master both of these latter games and, therefore, the best computers only can match our play. Not so with chess. But, now we can note that coming up with true theories about unobservables is more difficult than understanding how to make winning moves in chess. Therefore, it is *prima facie* implausible that we have "winning" theories about the universe. That is, it is *prima facie* implausible, given that the universe and the correlative logical space of theory are so much more complex than a game of chess, and given that we do not understand how to make winning moves in chess, that we successfully have plumbed the unobservable universe, either by means of "insight" or by means of brute force calculation. Let me try to make this more perspicuous with an example.

Scientists want to come up with a theory of quantum gravity. This involves proposing laws rather than finding analytic solutions to laws. As Kuhn points out in the block quotation reproduced above, articulating such a theory involves exploring possibility space, just as a chess player would. But, there are three things that are of note here. First, the possibility space is not as well-defined for articulating a theory, such as a theory of quantum gravity, as it is for chess. In chess, there are precise rules that define which moves are legitimate, and every chess player knows what they are. By way of contrast, although there are rules that define what would count as an inappropriate theory of quantum gravity, these latter rules seem much more nebulous, and nobody seems to know exactly what they are. Second, the possibility space for quantum gravity is more complex than it is for chess. There are far more

[23] Chess programmers sometimes point out that the Shannon number, an estimate (by Claude Shannon) of the game-tree complexity of chess, is (significantly) higher than the number of atoms in the known universe (according to our best theories). However, this is not a meaningful comparison: the game-tree complexity of 32 particles in a confined space is (significantly) higher than that of a chess game, and there are far more than 32 particles in the universe (and the universe has been and will continue to be around for a far larger number of "moves" than any chess game (on account of the 50-move rule. I note in passing, what is relevant to my argument in the main text, that the 50 move rule has been disproved by computers, although it is retained in official human competitions on the grounds that the counterexamples are, for human play, extremely unrealistic)).

possible theories, even far more possible theories that would conform to all accepted observations (overlooking, for the moment, complications associated with disputed and disputable observations—complications that only make my argument for premise (3) stronger), than there are possible chains of moves that need to be explored in a chess game. Indeed, it is precisely the fact that the possible chains that need to be explored in a chess game are limited that made a brute-force software approach tractable.[24] Third, the criteria of success, as defined in the present context, are much more stringent for a theory of quantum gravity than they are for chess. A quantum theory of gravity will explain events all across the universe and all across time, and whether it is successful in this will be difficult to ascertain even in a best case scenario. A winning move in chess, by way of contrast, will be part of a comparatively limited chain that ends in a state that is easily ascertainable on all sides: checkmate. Thus, it seems to me that if there is good reason to think that the best human chess players do not understand how to make winning moves, then there is good reason to think that the best human theories about unobservables are wrong.

This concludes my initial articulation of the chess argument. In the next section I expand it by defending it from objections.

Section 3 Some Objections

I think premise (1) is relatively uncontroversial. But, I think many will object to the other premises in the chess argument. So, in the following, I am going to confront various objections to premises (2) and (3), many of which have been raised by actual interlocutors.

Subsection 3.1 The Human Programmers Objection

Some might object to the claim that we do not understand how to make winning moves in chess. To see how this objection gets off the ground, we can begin by noting that chess engines are programmed by humans. Moreover, the points systems (whereby points are assigned to different pieces and positions) are determined by humans. So, the fact that chess engines play winning moves is actually direct evidence that humans (in particular, the humans who devised the points systems) do understand how to make winning moves.

Two considerations give more force to this objection. First, each of the individual calculations performed by a chess engine is relatively simple and straightforward. In principle, humans could carry out the calculations performed by these chess engines (given enough scratch paper and enough time). So, a human *could* understand the winning moves proposed by a chess engine and, indeed, could carry out the analysis one or two moves deeper in order to best the engine. We just would have to make the handicaps suitable. Perhaps the computer has only 1 microsecond to move whereas the human (or arbitrarily large team of humans) has 1 year.

Second, in addition to beating humans at chess, computers can outperform humans at the multiplication of large numbers. But, I do not think anyone would be tempted on these grounds to say that humans do not understand multiplication. So, we should not conclude that humans do not understand how to make winning moves at chess: premise (2) is false, or at least not well grounded.

I would like to say three things in response to this.

First, being able to perform a calculation in principle is not the same thing as actually understanding it. Humphreys makes this point well in a slightly different (but related) context:

> An average human can calculate at the rate of 10^{-2} floating point operations (flops) per second. (One flop is the equivalent of adding or multiplying two thirteen-digit numbers.) Current supercomputers operate at teraflop speeds: 10^{14} times faster than humans. A teraflop machine operating for three hours can perform a calculation it would take a human the age of the universe to complete and current supercomputer speeds will seem laughable a decade from now. This

[24] See note 14 above.

> extrapolation of our computational abilities takes us to a region where the quantitatively different becomes the qualitatively different…Speed matters.[25]

Precisely the fact that the performance of this calculation would take so long for a human to perform (even with a team) further undermines the plausibility of comprehension: by the time the team had finished the calculation, they would have forgotten the beginning. Moreover, giving the computer less time (as a handicap) is a way to ensure that *it* does not come up with a winning move, so beating it in those conditions does not evince genuine comprehension of winning moves in chess. At best, we can give the computer a reasonable amount of time; assume the computer did not have any glitches; and assert that the computer's move *is* a winning one based on the authority of (a) the algorithm and (b) the computer. This leads me to the second thing I would like to say.

Given that advances in chess engines are still going on (as evidenced by AlphaZero), I am not confident that we are justified in asserting either (a) that we understand how to program a computer to make winning moves or (b) that the computer makes winning moves at all. For all that we know, the computers might be making losing moves, it is just that they are making (far) better losing moves than humans.[26]

However, even if we are justified in asserting (a) and (b) , that is not the same as understanding why a particular move is a winning move. But, precisely that is what is at stake in the chess argument, and the fact that we do not understand that is unimpugned by virtue of the fact that the winning move is calculated using a brute force approach. Understanding why a chess move is a winning move involves being able to explain why it is so—how it confers an advantage, in terms of pieces, strategy, or position, or how it undermines an opponent, again in terms of pieces, strategy, or position—and all of this is opaque at the level of arithmetic. This is why, when there are too many calculations for a human to perform, even a brute force chess engine like Deep Blue provides evidence that humans do not understand how to make winning moves. Further evidence for this is revealed by reflection on the fact that the programmers who write the code for chess engines and who understand, in broad strokes, not only the ordered pattern of calculations that the computer will follow, but also how to do any of those individual calculations, nonetheless might not have a deep understanding of chess and, indeed, might not stand a chance in a match against even a moderately good chess player, let alone the grandmasters against whom their engines are tested.

Third and finally, the development of AlphaZero, in my mind, completely upends this objection: AlphaZero makes winning moves against humans and against the best brute force chess engines (the best justification we can have at this time, it seems, for asserting that it makes winning moves *sans phrase*), and its process of figuring out what move to make is such as concomitantly to render the analogy between multiplication and chess engines moot and to make more forceful my point about the difference between being able to perform individual steps in a calculation and understanding the whole. So, it seems to me that the human programmers objection does not withstand critical scrutiny.

Subsection 3.2 The Disanalogy Objection

Another objection that someone might make is that chess and science (and, in particular, the attempt to come up with theories that use unobservables to explain and predict observables) are disanalogous, and so premise (3) is false.

---

[25] (Humphreys, 2004, 53-54).

[26] Some have noted in response to this point that, although the StockFish chess engine is able to come up with moves that are winning-against-humans, it clearly does not know how to come up with moves that are winning *sans phrase*, for it is less powerful than AlphaZero. Not only, as pointed out in the main text above, does the mythical perfect chess game seem to elude us (even with AlphaZero), but, more, there is no guarantee that humans will never develop a winning chess strategy against AlphaZero—it is worth remembering, as remarked above (see notes 14 and 22 above), that it took 7 years for humans to topple the AI-established dominance in Go.

All of this might be true. But I think that the uncertainty behind all of it bolsters rather than undermines my position in the chess argument. I return to a related point in subsection 3.5 below.

One disanalogy between chess and science is that a typical chess match involves only two competitors, whereas science is a team effort frequently involving hundreds of people. To put this another way, chess is individual whereas science is collaborative. Even team chess, allowed at many tournaments for younger players, involves merely summing the wins from individual games (in technical language, the team score is an aggregate rather than a totality). Science, by way of contrast, generates results and theories only by having people work together.

A second and related disanalogy between chess and science is that a typical chess game will last less than a few hours. Even postal chess games generally will not last more than a year. It is not unusual for science projects, by way of contrast, to go on for decades at a time, and the tradition of science is cumulative in that newer theories frequently build on the successes of older ones, effectively extending the collaboration noted in the previous paragraph across centuries of practice.

Now I would like to say three things in response to this. First, I grant that there are many ways in which chess and science are dissimilar. I am not convinced that the dissimilarities remarked upon in the previous paragraphs are as profound as an opponent of the chess argument might make them out to be. For example, attempts to understand the game of chess are cumulative and in that sense go on for very long periods of time and involve many people, as anyone who strays too far into the chess section of a bookstore can attest. But, even if the dissimilarities remarked upon in the previous paragraphs are granted, we must ask whether these suffice to debunk premise (3) of the chess argument, and this leads me to the second thing I would like to say: they do not.

What is relevant for the chess argument is whether the particular property that makes it impossible for humans to understand how to make winning moves in chess is shared by the (unobservable) universe and the theoretical tools at our disposal to understand it. That property is complexity, and I think that even an opponent of the chess argument would grant that the universe is (far) more complex than the game of chess if only for the simple fact that every chess game takes place within the universe.

Moreover, at least some disanalogies between chess and science support, rather than undercut, the chess argument. For example, consider the following, noted at the end of section 2 of this paper. The rules of chess are (more or less) static, known, and inviolable. But, it seems unlikely that there is a fixed scientific method—and, if there is one, it is notable that previous attempts to articulate it have proven to be, at best, contentious. In addition, even if we cannot determine whether a given move in chess is a winning move, we can determine who has won, and who has lost, a given chess game. But, the same does not seem to be true of science: the issues associated with confirmation and the problem of induction are well known, as are the issues associated with disconfirmation and the Duhem-Quine problem.

Now, if the main difference between chess and science is that the former has fixed rules and an ascertainable endpoint whereas the latter does not, then the fact that we do not understand how to make winning moves in chess only militates more powerfully in favor of the conclusion of the chess argument, that there is good reason to think that the best human theories about unobservables are wrong. Thus, I think that the disanalogy objection, like the human programmers objection, does not withstand critical scrutiny.

Subsection 3.3 The Antirealism Objections

Recall van Fraassen's use of chess, discussed in section 2 of this paper, to illustrate the difference between the aims of science and the aims of individual scientists. Some might take this to reveal a problem for my argument. That is, van Fraassen thinks that the aim of science is to discover empirically adequate theories, not true ones. But, if that is correct, then premise (3) of the chess argument (if there is good reason to think that the best human chess players do not understand how to make winning moves, then there is good reason to think that the best human theories about unobservables are wrong) seems to depend on a realist understanding of science, one that an antirealist will reject. Some might argue on these grounds that the chess argument is not available to the antirealist.

Now, I do not think that premise (3) of the chess argument depends on a realist understanding of science. The idea behind the consequent of premise (3) is *not* supposed to be: there is good reason to believe that the theories that successfully refer to unobservable entities make mistakes. Rather, the idea is supposed to be: there is good reason to think that the best theories that we have that posit unobservables, which may or may not exist, are wrong and, in particular, that this error enters in when the theories cross the line from observables to unobservables.

However, this only raises another problem. Suppose that the chess argument works and, therefore, that we have good reason to believe that our theories about unobservables are wrong on account of the complexity of the universe and the correlate complexity of the logical space of theories about unobservables. If so, then there seems to be every reason to believe that this complexity also will entail that it is impossible to arrive at justified empirically adequate theories. After all, empirical adequacy is a very demanding standard, requiring that a theory account not merely for all extant observational evidence but also for all possible observational evidence. So, some might object that the chess argument threatens to go too far: it might undermine both realism *and* antirealism.

In response to this, let me begin by pointing out that the chess argument does not impugn our ability to make true claims about actually observed observables, nor does it impugn our ability to make true claims about unobserved observables. For example, the chess argument is not based on induction regarding our ability to justify presently unobserved but observable phenomena.

Moreover, van Fraassen's constructive empiricism is not the only antirealist game in town. Recall from the introduction that, whereas the antirealist, unlike the realist, maintains that there cannot be justified beliefs about unobservables, the antirealist, unlike the radical skeptic, concedes that there can be justified beliefs about observables. From this it may be seen that, between realism and radical skepticism, there is plenty of space for different kinds of antirealism other than the constructive empiricist empirical adequacy. An antirealist might assert that we can have justified beliefs about observables, but not about unobservables; she might (or, indeed, might not) assert that we can have justified beliefs about some unobserved observables; and she might remain silent on or reject the possibility of justified empirically adequate theories. So, even if the chess argument suggests that any claims to empirical adequacy are unjustified—and I am not saying that it does—this would not undermine antirealism or militate in favor of radical skepticism: to assert that we never will be justified in believing that a theory can account for all possible observables is a far cry from saying that we do not have any justified beliefs about observables (whether actually observed or not).

Subsection 3.4 The Complexity Objections

In the penultimate paragraph of section 2, I defended premise (3) of the chess argument (if the best human chess players do not know how to make winning moves, then there is good reason to think that the best human theories about unobservables are wrong) on the basis of ideas about complexity: I argued that the complexity of the game of chess is what prevents us from knowing how to make winning moves; that the universe is more complex than the game of chess; and that arriving at true theories about unobservables is more difficult than making winning moves. However, this is controversial, and there are at least two distinct ways of attacking it.

For one thing, there are many alternate views about why we have good reason to believe that our theories about unobservables are mistaken. For example, one possible view is based simply on the fact that unobservables, as such, cannot be experienced directly: they are too small, too distant, too X to be observed and, in the absence of observation, no theory can be justified. Another possible view is that we can conclude from the history of science (especially the so-called "graveyard" of previously successful but rejected theories) that we cannot have true theories about unobservables: we are not reliable when it comes to theorizing about unobservables. So, complexity is not the only reason someone might have for being skeptical of our ability to arrive at true theories about unobservables. Therefore, there should be some argument for the claim that it is primarily complexity that prevents humans from having true theories about unobservables, rather than, say, considerations regarding the limitations of human experience and historical considerations.

For another thing, while the universe might be more complex than a game of chess, it is not at all clear that coming up with a true theory about unobservables is more complex than, or even as complex as, figuring out a winning move against the best chess computer programs. Indeed, it seems that scientists have achieved remarkable success, such as the prediction made by general relativity that light must be deviated by massive bodies or the discovery of the Higgs boson, and perhaps this success translates into justified theories about unobservables. Thus, further argument is needed in order to infer, from the comparative complexity of the universe, that we have good reason to believe that our best theories about unobservables are, in general, mistaken.

Let me tackle these two objections in the order in which I have posed them: I have one thing to say about the first and two things to say about the second.

I think that the first complexity objection mistakes the goal of the chess argument. The chess argument is not supposed to show that the *only* reason to think that our best theories about unobservables are wrong is bound up with the fact that the best human chess players do not understand how to make winning moves, nor is the chess argument supposed to show that the *only* reason to be an antirealist has to do with the complexity of the universe. Rather, the chess argument is supposed to be one argument among many that an antirealist might deploy in order to justify her position or, conversely, one argument among many that a realist would have to respond to in order to justify his position. Some of these arguments might overlap to some extent. For example, the chess argument relies on ideas about our cognitive limitations, and perhaps other antirealist arguments do so as well. But, that does not seem to be a problem, nor does it impugn the chess argument, provided we remember, once again, that this is not meant to be the only argument in town. Indeed, as noted at the outset of this paper, I am not arguing that the chess argument provides decisive, or *ultima facie*, grounds for antirealism. I think that such a project would require, in addition, grappling with various realist arguments (perhaps most prominently, the no miracles argument), something I have tried to avoid here inasmuch as my goal is at once more modest and more ambitious: not to make a decisive case, but, rather, to introduce a genuinely novel line of argument.

This ties into part of what I would like to say about the second complexity objection. To show that we have good reason to believe that our best theories about unobservables are mistaken is not, in my view, to show that we cannot be justified in believing these theories. While I cannot defend this view in detail here, I think we can have good reason to believe X, and we also can have good reason not to believe X; to figure out what we *ultima facie* ought to believe in such cases can be difficult. For example, suppose that I have two reliable friends, B and C; suppose that I do not have access, at the moment, to any direct evidence about whether it is raining; and suppose that B says that it is raining whereas C says that it is not. On my view, assuming that all else is equal, I have good reason to believe that it is raining, and I have good reason not to believe that it is raining. Thus, as noted in the previous paragraph, to show that we cannot be justified in believing our theories about unobservables, we almost certainly would need, in addition to the chess argument, at the very least, to grapple with traditional realist arguments, and this outstrips my present ambitions. Nonetheless, I think that the second objection does require me to say more about complexity and its role in supporting premise (3) of the chess argument, and I would like to conclude this subsection by doing that.

The observable universe is incredibly complex in a fairly straightforward way: it has a vast number of parts that, from moment to moment, move and interact in a vast number of ways. But, to articulate a true theory about some unobservable part of the universe is to master and move beyond this incredible complexity, theorizing about and exploring explanatory hypotheses. And the point is that, if we are unable to master the incredible complexity of a game of chess, a mastery that would be evinced by understanding how to make winning moves (and is belied by the failure to do so), then we have good reason to suppose that we are unable to master the even more incredible complexity of the observable universe.

The move from observables to unobservables is made, according to realists, with a creative form of inference, like inference to the best explanation (IBE), or abduction. Thus, scientific realists often defend the truth-conduciveness of IBE, and in particular of the non-empirical criteria involved in scientific explanations, such as simplicity, unificatory power, and coherence of theories.[27] Obviously, I cannot provide a full analysis of IBE here, nor, I think, do I need to; the point I want to make is merely that the chess argument can be interpreted as a novel argument against IBE, siding with the antirealist who claims that empirical adequacy is the most that can be justified without IBE.

---

[27] (Psillos, 1999).

Now, as noted in the second complexity objection, some might appeal to the remarkable success that scientists, as compared to chess players, have achieved, in order to try to drive a wedge between this move from chess to theories about unobservables. But, there are two problems here. One is that, in this context, any such appeal looks to be question-begging, for it relies on exactly what needs to be shown: that scientists have been successful in articulating true theories about unobservables. The other is that this sounds like the no miracles argument and, as such, is out of place: once again, my goal here is not to assess different realist arguments in order to show that the chess argument gives decisive reason to be an antirealist; my goal is to show that the chess argument gives good reason to be an antirealist, with an understanding that more work would have to be done in order to give anything approaching a full-throated defense of anti/realism. I conclude that the chess argument is not undermined by the complexity objections.

Subsection 3.5 The Winning Objection

Premise (2) of the chess argument infers from the fact that humans cannot beat the best chess engines at chess to the claim that humans do not understand how to make winning moves. But, some might try to block this inference. In particular, some might argue that human chess players certainly do understand how to make winning moves against human opponents. After all, predicting who is going to win when two humans play against one another is rarely a toss-up. And human chess players also often are able to solve endgame puzzles, or to find a forced mate if the latter is not buried too many moves away—and surely these qualify as understanding how to find winning moves, even if only in limited circumstances. Does this impugn the chess argument?

I do not think that it does, and I want to begin my explanation of why I think this with a thought experiment.

Suppose that two ten-year-olds are playing one another at chess. Suppose, further, that these ten-year-olds are not particularly good at chess. That is not meant to be nasty. The point is merely that they are not prodigies, and, we may suppose, they do not need to be. Finally, suppose that the year is 1972: humans (and, in particular, Bobby Fischer) reign supreme in the world of chess—Deep Blue and AlphaZero are not even on the radar.

Now, if a grandmaster were to observe these two children playing, I think the grandmaster would be justified in concluding that she is able to beat either one of them; that neither of them has a deep understanding of chess; and that, although both of them might have mastered the rules of the game and even might be able to explain these rules to another ten-year-old, neither of them understands how to make winning moves--this despite the fact that one of them is most likely to beat the other and, indeed, might do so consistently.

The point of this thought experiment is, I hope, clear: current AI chess engines are to the grandmaster as the grandmaster is to the children. In fact, the difference between current AI chess engines and the grandmaster, in terms of level of play, is arguably more profound than the difference between the grandmaster and the children, and it manifests in manifold ways: number of games that can be played simultaneously and successively; mastery of bullet, blitz, standard, and other forms of chess; etc.

Some humans might be puzzle-masters in the following sense: they can find winning moves when the game has been highly simplified or in endgame scenarios. Some humans might be chess-masters in the following sense: they can win against most other humans most of the time. But puzzle-masters are not chess-masters, and chess-masters are, as we now know (thanks to our current chess engines), not up against very stiff competition. So, I do not think that the existence of puzzle-masters or chess-masters undermines premise (2) of the chess argument.

Subsection 3.6 The Cooperation Objections

An alternate line of objection against the chess argument pushes on the way that the argument seems to view technology.

Humans are not able to build skyscrapers without the aid of various tools, like cranes, excavators, and rollers. But, we do not say that humans are unable to build skyscrapers for this reason. Of course, cranes out-lift us, excavators out-dig us, and rollers out-roll us. But, that is sort of the point: that is precisely why we use them--and this sort of thing is prevalent in human life in general and in science in particular. Technology can be viewed as a competitor, as in the Hollywood AI-against-the-humans trope, but it also can be viewed as a cooperator, and in the latter role, technology enhances human capabilities rather than anything else. So, why not think that AlphaZero has given us a deeper understanding of chess rather than that it evinces a failure in that regard?

The line of reasoning in the previous paragraph threatens to upend premise (2) of the chess argument. But, I do not think that this line of reasoning works. Using a computer to win at chess is cheating, and the reason for this is that the point of chess is to win on your own two feet, so to speak--on the basis of your own understanding of the game--and winning via a computer is seen, rightly in my view, to bypass that. Technology, in general, might enhance human capabilities. But, in the world of competitive chess, the cooperative role of computers is banned.

Now, some might respond that this is an arbitrary limitation. For one thing, in much the same way that some people propose sports leagues in which performance enhancing drugs are allowed, some might propose chess tournaments that allow players to use computers. For another thing, the concern here is not primarily winning or losing at chess, but, rather, understanding chess, and (the objection continues), in the same way that computers are often used to do calculations in engineering and mathematics but do not seem to threaten our understanding of those fields, we might think that computers can be used to do calculations in chess without undermining our understanding of the game. Finally, this line of reasoning can be extended into an attack on premise (3) as well. Science is not a game, and using a computer when engaged in science is often encouraged: even if computers are viewed as competitors in chess, they are viewed as cooperators in science. Thus (to wrap up the objection), even if computers pose a threat to our understanding of chess, it does not follow that they pose a threat to our understanding of unobservables.

However, there are at least two problems with this line of reasoning. One is that it rests on a false presupposition. Computers *are* used for cooperative purposes in chess just as much as other technology is used for cooperative purposes in other human endeavors. There probably is not a single top-ranked player alive today who does not train using a computer; many established chess teachers use chess engines in order to figure out what moves to recommend to their students; and some correspondence chess, at least at the time of this writing, does allow players to use AI chess programs.

The second and deeper problem with this line of reasoning has to do with opacity. It is not merely that chess engines come up with winning moves more quickly than humans, nor is it merely that humans can use a chess engine to determine a winning move which is then readily explicable. Chess engines, like Deep Blue and, for reasons already articulated, even more clearly, AlphaZero, come up with moves that we can accept as winning (we see the evidence in the fact that they beat us) but that we do not understand as such--their winningness is opaque--and part of the evidence for this is not merely that computer chess programs beat us, beat us consistently, and have been beating us consistently for more than 25 years, but also, and relatedly, that we have not been able to analyze computer chess programs, either ourselves or with the aid of other computers, in order to develop a humanly-master-able strategy for beating them.[28] And, if the arguments above are successful, this generalizes to theorizing about unobservables.

At this point it is worth pausing to note that, although computers are used in science, they are not used to generate theories about unobservables. Of course, this might change in the (near) future. But, at present, the fact that computers enhance our abilities in empirical science is orthogonal to the chess argument. Moreover, this last line of reasoning can be strengthened. To see how, consider some of the paradigmatic ways in which computers are used in science: to help us to make predictions about weather patterns or about how small molecules will behave. There are three things that are noteworthy about this.

---

[28] The contrast here with Go is striking. See note 22 above.

First, the results from these calculations are often mistaken. For example, weather predictions are notoriously unreliable, and the reasons often given for this include known unreliability of models, the incredible complexity of the systems, and the sensitivity of these systems to small changes in initial conditions.[29] Given the foregoing, I want to suggest that this bolsters my general project here, even if, perhaps, only in a small way.

Second, it seems to me that, even when a computer is used to do brute force calculations, if the calculations are sufficiently complex, then understanding is, if not completely undermined, at least greatly attenuated. This is a version of Humphrey's claim from above (subsection 3.1) and, as with what I said in the previous paragraph, weather patterns are illustrative. The point of using brute force computer calculations to make weather predictions is that the models are so incredibly complex that no group of humans, let alone any individual human, would be able to carry out the necessary calculations in a timely fashion.[30] When the calculations reach that level of complexity, as Humphrey points out, a difference in degree becomes a difference in kind: we are looking at an opaque data system, one for which, at this point, we cannot even master the masses of input data (largely flowing in automated fashion from satellite systems and other tools) let alone the transformations they undergo. We just have outputs at the end that we take as authoritative and interpret accordingly.

Third and finally, when machine learning is used, as, for example, in computer simulations of protein folding, then, as with the move from Deep Blue to AlphaZero, there is a qualitative jump, a difference in kind that is not related to a difference in degree, and comprehension is no longer on the table: we see the starting points and the endpoints, and we have a general idea of what the model says about the middle at a very abstract level—but this, I think, is not sufficient for understanding. Moreover, this is not merely my idiosyncratic view; it is the published view of experts in the field:

> The [current] situation [regarding the use of deep-learning AI in protein fold prediction] is analogous to interpreting a movie by fast-forwarding to the final scene without first watching the previous two hours; we know how it ends, but we don't know why.[31]

Thus, it seems to me that the chess argument can withstand the cooperation objections.

Conclusion

In this paper I have advanced a novel argument for scientific antirealism, the position that affirms OBS while denying REAL. I called this argument the chess argument because it is based on recent advances in computer chess. After clarifying the debate and what I am arguing for (in section 1), I set out and defended the chess argument (in section 2), and I confronted various objections (in section 3).[32]

---

[29] (Krishnamurthy, 2019).

[30] (Mason, 1973).

[31] (Chen *et al.*, 2023, 2).

[32] I would like to thank the editors and anonymous reviewers' at the *Asian Journal of Philosophy* for their patient, detailed, and constructive feedback; they enabled me to improve the article immeasurably, and they will see their influence throughout.

Bibliography

Asay, Jamin. 2019. "Going local." *Synthese* Vol. 196, No. 2: 587-609.

Bird, Alexander. 2000. *Thomas Kuhn*. Acumen.

Boyd, Richard. 2002. "Scientific Realism," *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2002/entries/scientific-realism/>. Accessed 2019.11.22.

Carnap, Rudolf. 1975. "Observation Language and Theoretical Language." In *Rudolf Carnap, Logical Empiricist* (ed. Jaakko Hintikka), Synthese Library: 75-86.

Chakravartty, Anjan. 2017. "Scientific Realism," *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>. Accessed 2019.11.22.

Chen, Shi-Ji *et al*. 2023. "Protein folds vs. protein folding: Differing questions, different challenges." *Proceedings of the National Academy of Sciences* Vol. 120, No. 1: 1-4. https://doi.org/10.1073/pnas.2214423119

Greenemeier, Larry. 2017. "20 Years After Deep Blue." *Scientific American.* <https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>. Accessed 2019.11.22.

Humphreys, Paul. 2004. *Extending Ourselves*. Oxford University Press.

Kahn, Samuel. 2020. "A Realistic Argument for Scientific Realism." *Philosophia* Vol. 48, No. 5: 1901-1914.

Krishnamurthy, Venkataramanaiah. 2019. "Predictability of Weather and Climate." *Earth and Space Science* Vol. 6, Issue 7: 1043-1056. doi: 10.1029/2019EA000586.

Kuhn, Thomas. 1996. *The Structure of Scientific Revolutions*. The University of Chicago Press.

Lee, Dave. 2017. "Google's 'superhuman' DeepMind AI claims chess crown." *BBC Technology*. Dec. 6. < https://www.bbc.com/news/technology-42251535>. Accessed 2019.11.22.

Mason, Basil John. 1973. "The Application of Computers to Weather Forecasting." *Physics in Technology* Vol. 4, No. 1: 62. doi: 10.1088/0305-4624/4/1/I03

Maxwell, Grover. 1962. "On the Ontological Status of Theoretical Entities." *Minnesota Studies in the Philosophy of Science* Vol. 3: 1-27.

Muller, F.A. 2004. "Can a Constructive Empiricist Adopt the Concept of Observability." *Philosophy of Science* Vol. 71, Issue 4: 637-654.

Muller, F.A. 2005. "The Deep Black Sea." *British Journal of the Philosophy of Science* Vol. 56, Issue 1: 61-99.

Newman, Mark. 2010. "The No-Miracles Argument, Reliabilism, and a Methodological Version of the Generality Problem." *Synthese* Vol. 177: 111-138.

Popper, Karl. 2002. *The Logic of Scientific Discovery*. Routledge Classics.

Psillos, Stathis. 1999. *Scientific Realism*. Routledge.

Rosen, Gideon. 1994. "What is Constructive Empiricism?" *Philosophical Studies* Vol. 74, No. 2: 143-178.

Rowbottom, Darrell. 2011. *Popper's Critical Rationalism*. Routledge.

Rowbottom, Darrell. 2014. "Aimless Science." *Synthese* Vol. 191, No. 6: 1211-1221.

Silver, David, *et al.* 2018. "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go Through Self-Play." *Science* Vol. 362, Issue 6419: 1140-1144.

Van Fraassen, Bas. 1980. *The Scientific Image*. Clarendon Press.

Van Fraassen, Bas. 1994. "Gideon Rosen on Constructive Empiricism." *Philosophical Studies* Vol. 74, No. 2: 174-192.

Waters, Richard. 2023. "Man beats machine at Go in human victory over AI." *ARS Technica*. Feb. 19. < https://arstechnica.com/information-technology/2023/02/man-beats-machine-at-go-in-human-victory-over-ai/>. Accessed 2023.07.24.

Wedberg, Anders. 1975. "How Carnap Built the World in 1928." In *Rudolf Carnap, Logical Empiricist* (ed. Jaakko Hintikka), Synthese Library: 15-54.

Yin, Qi-Yue *et al.* 2023. "AI in Human-computer Gaming." *Machine Intelligence Research* Vol. 20: 299-317. <https://doi.org/10.1007/s11633-022-1384-6>