CAN DARWINIAN MECHANISMS MAKE NOVEL DISCOVERIES?: LEARNING FROM DISCOVERIES MADE BY EVOLVING NEURAL NETWORKS

ABSTRACT. Some philosophers suggest that the development of scientific knowledge is a kind of Darwinian process. The process of discovery, however, is one problematic element of this analogy. I compare Herbert Simon's attempt to simulate scientific discovery in a computer program to recent connectionist models that were not designed for that purpose, but which provide useful cases to help evaluate this aspect of the analogy. In contrast to the classic A.I. approach Simon used, "neural networks" contain no explicit protocols, but are generic learning systems built on the model of the interconnections of neurons in the brain. I describe two cases that take the connectionist approach a step further by using genetic algorithms, a form of evolutionary computation that explicitly models Darwinian mechanisms. These cases show that Darwinian mechanisms can make novel discoveries of complex, previously unknown patterns. With some caveats, they lend support to evolutionary epistemology.

KEY WORDS: connectionism, evolutionary epistemology, genetic algorithms, neural networks, scientific discovery

1. INTRODUCTION

Some philosophers of science have suggested that the development of scientific knowledge may be thought of, on analogy with biological evolution, as a kind of Darwinian process. The process of discovery, however, is one potentially problematic element of this proposed model. Until relatively recently, philosophers have taken scientific discovery to be a matter of psychological insight, and so have left the question of how a scientist makes a discovery to psychology, and focused instead on the process of justification. Herbert Simon was one of the first to suggest that there might be an interesting, quasi-formal structure to scientific discovery, and to attempt to simulate scientific discovery in an artificial intelli-
gence computer program. Neither the psychological insight view nor Simon's artificial intelligence view of discovery fit well with the evolutionary model. In this paper I contrast these with a more promising candidate. Recent research on connectionist models that were evolved using genetic algorithms, though conducted for other purposes, provide useful cases to help evaluate this aspect of the evolutionary analogy by throwing light on the possibility of Darwinian discovery. Before describing the models themselves and discussing their suggestive implications, let me begin with a brief review of the relevant ideas of evolutionary epistemology.

2. EVOLUTIONARY EPISTEMOLOGY

In 1857, Herbert Spencer proposed one of the earliest versions of evolutionary epistemology, pointing out that scientific knowledge does not appear fully formed, but develops incrementally with new knowledge based on antecedents and itself serving as a basis for further development. Science, Spencer said, began with a period in which it could not be distinguished from art and religion, and only gradually progressed into its own different kinds of investigation. Today, we would say that Spencer's is a weak, figurative form of evolutionary epistemology, for he does not claim that the development results from specifically Darwinian mechanisms. T. H. Huxley proposed a more robust version of the view in an 1893 essay, in which he pointed out that the essence of science is a critical spirit, and that science progressed by means of a critical selective process. Scientific hypotheses on this view are like organisms that must compete for survival. As Huxley put it: "The struggle for existence holds as much in the intellectual as in the physical world. A theory is a species of thinking, and its right to exist is coextensive with its power of resisting extinction by its rivals." (Huxley, 1893, p. 229)

Bringing in selection in this way starts to make the comparison with evolution more compelling. Stephen Toulmin (1967, 1972, 1981) drew out many more common elements – pointing out, for instance, that scientific concepts can be seen in terms of changing ratios of slightly different theoretical variations in a population of scientists, and that the fitness of these depend upon how well they work in the intellectual environment in solving problems, approx-
imating objective reality, and so on – arguing that the evolutionary model is more than a suggestive analogy, but that it also may be cited to explain aspects of scientific development. Karl Popper (1959, 1968 (1962), 1972), Donald Campbell (1973, 1974a, 1974b), and, more recently, David Hull (1982, 1990) and Michael Ruse (1985, 1986) are among those who have made contributions to evolutionary epistemology.

Notable critics of evolutionary epistemology include L.J. Cohen (1973) and Paul Thagard (1980), who have argued that there are too many significant differences between scientific development and evolution for the latter to be a good model of the former. Darwinism, they say, is a poor analogy that tells us little about the development of scientific knowledge. This is not the place to assess all their criticisms, some of which are quite telling. Here I am just interested in what looks to be the weakest element of the analogy, namely, that having to do with the origin of variations. This is the element of the analogy that corresponds to scientific discovery; how do new hypotheses arise and enter the competition?

One problem is that many novel scientific hypotheses do not seem to be chance discoveries but rather, as Michel Ruse put it, appear to be "moulded or directed according to certain perceived ends of the scientist..." (Ruse, 1986, p. 57) Variations according to the Darwinian model arise at random, without foresight as to their possible utility. There are two main ways to try to overcome this prima facie disanalogy. One might try to show that organic variations somehow simulate directed variation, or else that the introduction of new ideas in science can indeed be seen to fit with the standard Darwinian picture of random variation. Karl Popper (1972) argued for both possibilities, though the second approach is most characteristic of his view. Popper built his evolutionary epistemology upon his falsificationist base – scientists, he said, put forward bold conjectures and then attempt to refute them. The growth of scientific knowledge occurs by this selection process, falsification eliminating those hypotheses that are comparatively unfit.

Despite its title, Popper's most important book – The Logic of Discovery (1959) – says almost nothing about the process of discovery. As for a "logic" of discovery, Popper's view seemed to be that there isn't one. A scientist may discover what will turn
out to be a breakthrough hypothesis simply by accident, or by a lucky guess, by the insight of genius, or in a dream, for all it matters to science. Logic, or rational procedure, enters the picture only in the process of testing that hypothesis. Many philosophers of science have criticized Popper’s falsificationist framework, arguing that science can do more than just show that a hypothesis is false (or that it has survived attempts to falsify it), and can provide hypotheses with positive confirmation. Nevertheless, confirmationists and falsificationists alike have generally agreed that there is no logic of discovery. This standard view has held that there is a logical gulf between the “context of discovery” and the “context of justification.” (As widely accepted a distinction as this is, few seem to have noted that in rejecting a logic of discovery one necessarily comes very close to accepting the Darwinian view that discovery is largely a matter of luck.) Herbert Simon was one of the first to challenge and develop an alternative to the view, suggesting that inductive generalizations may be confirmed in the same way that they are discovered, and that the process of scientific discovery does have a logical mechanism that can govern it that can be modeled on a computer. More importantly, for our interest, Simon and his colleagues held that discovery could not be a random process, but had to be one that involved purpose and reason.

[W]e cannot believe either that the events leading up to the [moment of discovery] are entirely random and chaotic or that they require genius that can be understood only by congenial minds. We believe that finding order in the world must itself be a process impregnated with purpose and reason. We believe that the process of discovery can be described and modeled, and that there are better and worse routes to discover – more and less efficient paths. (Langley et al., 1987, pp. 3–4)

3. MODELING DISCOVERY

Simon’s key idea is that the process of discovery may be thought of as a special kind of problem-solving, and so should be governed by general problem-solving mechanisms. If one then considers problem-solving abstractly, as a search through a space of possibilities, then it may be possible to develop a kind of logic of discovery in the weak sense of there being heuristics that could narrow the search space. Simon and his colleagues pioneered research along
these lines, modeling scientific discovery with a computer program they developed called "BACON," which incorporated heuristics that allowed it to work through a data set it was provided, trying to discover significant mathematical regularities. BACON's heuristics included ones like the following:

- If the values of a term are constant, then infer that the term always has that value.
- If the values of two numerical terms increase together, then consider their ratio.
- If the values of one term increase as those of another decrease, then consider their product. (Langley et al., 1987, p. 66)

With the help of these and other such "productions," BACON was able to discover, for instance, Kepler's 3rd law of planetary motion and the concept of gravitational mass. This was impressive, provocative work, and has led many philosophers to reevaluate their earlier dismissal of scientific discovery as a reasonable topic of philosophical investigation. A few critics argued, however, that BACON's accomplishments were not really "discoveries," but that these had inadvertently been designed-in by the programmers, perhaps in BACON's heuristics or the presented data set. Certainly it is true that BACON's discoveries were not novel, and they were, of course, known to the researchers in advance. These questions about whether we should count these as true discoveries, how BACON actually made them, and how its procedures compares to those used when scientists make discoveries are relevant to our assessment of the Darwinian model of discovery.

Evolutionary epistemologist Donald Campbell, who mounted the most sustained defense of the thesis that a blind-variation-and-selective-retention process is fundamental to all genuine increases in knowledge, admitted that many discoveries made now in science are not random, but are based upon following heuristic principles. He argued that the evolutionary model nevertheless applies to scientific development even at the point of the source of variations in that any heuristics that scientists may use, originally appeared themselves only through a trial and error process. According to Campbell, truly fresh ideas require a random search: "If one is expanding knowledge beyond what one knows, one has no choice but to explore without
the benefit of wisdom (gropingly, blindly, stupidly, haphazardly)"
(Campbell 1974b).

In large part, this point of contention turns simply upon what we count as being truly novel. There can be no doubt that scientists often use heuristics to guide their searches for solutions to problems, as do we all in ordinary situations. A heuristic is simply a general pattern, and we search for a solution to a puzzling phenomenon or data set first by running through the array of patterns that we have previously learned in other situations to see whether one or another might apply to this new situation. (BACON is doing just this when it compars the data it is given to the mathematical patterns that constitute its heuristic productions.) In one sense we would naturally say that we have discovered something new if we find that one of these patterns fits the new case.

In another sense, however, one could argue that this very fact – that a previously known pattern applies – means that it was not a truly novel idea that was discovered. Indeed, it is hard to see how rational problem-solving could be otherwise, since one can reason about something only by means of a store of patterns that is already known. Campbell’s point about what is required when we are “expanding knowledge beyond what one knows” is correct in that, if a truly novel idea in this stronger sense is needed, one has by definition no prior applicable pattern that will work as a guide. In such a case there is no alternative left except random variation. (Again, the notion of randomness that is relevant here is the same one that holds in the Darwinian mechanism.) Even when one has a list of heuristics, one typically must apply them blindly; BACON, for example, simply runs through its list of protocols in turn. This is not to say that one could not also have higher-level heuristics (meta-productions that could guide how one tries out heuristics; for example, in an order that has proven to be effective in the past), but even if there are multiple levels, at some point in the process one will have had to rely upon a random choice procedure.

If this is correct, then the disagreement about the import of blind variation versus heuristics as the basis of discovery may just come down to how much of science is novel in the basic sense, and how much of scientific discovery simply involves application of previously known general patterns to new or similar cases. Kepler’s
achievement was a more striking discovery in its day, since it was a radical departure from accepted patterns to suggest that planetary motion was not based on the circle but rather on the ellipse. Today such patterns would be more quickly recognized and not be heralded as quite so revolutionary. Indeed, Simon noted that, in an informal survey, when he showed scientists the data (without saying what it was) and asked what they made of it, most scientists quickly “discovered” Kepler's third law in it, essentially as BACON did. If most “normal” scientific discoveries occur in this way, rather than being revolutionary additions of a new basic kind of pattern, then the scope of the full Darwinian mechanism in science may be rather limited.

Our purpose here, however, is not to inquire about how widespread might be use of the Darwinian mechanism, but rather to consider whether it could work at all, so that evolutionary epistemology can at least get off the ground. Can we show that Darwinian processes – that is to say, natural selection acting upon undirected heritable variations – really are capable of making non-trivial novel discoveries? We could assess this by looking directly at how Darwinian mechanisms work in the evolution of biological organisms, but it may be more revealing to show this with a different kind of computer model.

4. EVOLVING CONNECTIONIST NETWORKS

In contrast to the classic artificial intelligence approach used in BACON, connectionist neural networks contain no explicit productions or heuristics, but are generic learning systems. Neural networks are parallel distributed processing computers built on the model of the interconnections of neurons in the brain. In a connectionist network, differential probabilistic weights of units in a hidden layer determine the pattern of firings in an output layer, given a pattern of firings received from an input layer. Such networks can be trained to be able to recognize patterns and complex analogies even through noise, or with only partial information. (Churchland, 1996). The cases I am interested in here work slightly differently in that they make use of genetic algorithms.
Genetic algorithms are a form of evolutionary computation, which explicitly models Darwinian mechanisms. In evolutionary computation, artificial populations (of, say, computer programs) are allowed to randomly vary and then are automatically culled by a selection function, with the survivors reproducing to form the subsequent generation as the process is repeated. This been done with traditional programming, but the distributed nature of neural networks makes them especially good candidates for evolutionary computation. I’ll describe two cases drawn from recent research on evolving neural networks by Risto Miikkulainen and David Moriarty in the computer science department at the University of Texas at Austin, that are illuminating for our question regarding discovery.

In the first case, evolving neural networks learned how to play the game of Othello. Othello is a Japanese board game that is a variation of the classic game of Go. It is played on an eight-by-eight grid with two players taking turns placing tiles on the squares of the board around an opening four-tile central configuration. The only legal moves are those into an open space such that it flanks an opponent’s piece or pieces between it and another of one’s own tiles. You must pass if you have no legal move, and play continues until neither player has a legal move. The tiles are black on one face and white on the other, distinguishing the players, and you flip over your opponent’s pieces whenever you flank them horizontally, vertically or diagonally. The goal is to have the most tiles showing one’s own color on the board at the end of the game. The opening and end phases of a game of Othello are fairly simple; the difficulty occurs in the middle game, during which one must strategically place one’s tiles to give one a good position for the endgame. The first computer program that could play Othello at a master level, named Iago, was designed by Rosenbloom in 1982. Like most artificial intelligence game programs, it worked by doing a deep search of possible moves with tables of heuristics that helped eliminate poorer moves. Rosenbloom subsequently developed Bill, which supplemented these search techniques with a Bayesian learning function to help optimize its evaluation function.

In contrast, Moriarty and Miikkulainen’s neural networks did not incorporate any search mechanism or heuristics, but were forced to use only pattern recognition of the current board configuration
to determine its moves. (Moriarty and Miikkulainen, 1995) The configuration of the 64 squares of the board was represented by 128 nodes in the input layer, the minimum needed to specify which of the three possible states (empty, own tile, opponent’s tile) each square of the board was in. The output layer had just 64 nodes, each corresponding directly to a space on the board. The activity of each output node was interpreted as indicating the strength of the network’s suggestion to move to that spot. The structure of the network was encoded in an artificial “chromosome,” with each node in the network defined by two 8-bit integers that coded its value, what other nodes it was connected to, and the weight of each connection. The networks used a marker-based scheme that allowed the architecture of the networks to vary as well as the weights of the units.

The key feature of the system for our interest enters at this point: the use of genetic algorithms to evolve the system. The researchers began with a population of fifty networks, each initially competing against an opponent that only made random moves, and later against Iago’s search program searching three levels down. To ensure different games, the initial board configurations were selected at random from among the 244 possible positions after four moves. The twelve networks of each generation that had fared best in the competition were randomly mated with one another (with two-point crossovers) to produce the next generation of networks, replacing the worst playing networks. In addition, random mutations were introduced each generation.

What was the result? In the course of their evolution against the random mover, the networks quickly discovered what is known as the “positional” game strategy. This is the standard strategy that human players learn to develop. It involves trying to maximize one’s tiles at the edges and corners of the board, while minimizing the number of tiles that one’s opponent has in those positions. The networks learned to do this, while also learning to avoid vulnerable positions. This is an impressive discovery in itself, but it is what happened subsequently that is particularly striking. After they had learned to beat the random player, the networks were pitted against Iago. At first the networks regularly lost, but eventually they evolved so that they were winning most of their games. This occurred with
the network's discovery of a complex Othello strategy, known as "mobility strategy," that is rarely seen outside of advanced tournament play. Unlike positional strategy, which many human players eventually discover independently, mobility is thought to have been discovered by just one tournament player, with others subsequently learning it from his example. Although the board's corners of course remain valuable, mobility strategy works not by systematic acquisition of territory, but rather by placing tiles to limit one's opponent's move options in ways that allow a massive reversal of fortunes in the endgame. The strategy is counter-intuitive and difficult to learn, making the networks' independent discovery of it all the more noteworthy.

5. DISCUSSION

The first question to ask is whether the process at work in these networks is a good model of Darwinian evolution. The answer is clearly that it is. The networks instantiate the key elements of the Darwinian mechanism – natural selection of heritable random variation. The networks vary randomly through both mutation and recombination in the relevant Darwinian sense, namely, that these occur independently of any specific goal. Furthermore, the variation is heritable, again in the appropriate sense that the network "offspring" arise by a copying mechanism from the "parents." Finally, natural selection operates over the course of the replications in that poorer networks are selected out automatically each generation as the result of a competitive process. Nothing about the Darwinian mechanism limits its occurrence to organic beings. As Daniel Dennett (1995, p. 50) has put it, the Darwinian mechanism is simply an algorithmic process, and it may be instantiated in biological organisms, computers, or any number of other forms. Evolving neural networks are truly evolving in the relevant Darwinian sense.

The second question is whether such evolving networks provide a good model of discovery? Here, too, the answer is affirmative. The discoveries of Othello strategies made by the evolving neural networks are not susceptible to the criticism that the discoveries were somehow already "designed in." Of course the programmers did set up the number of input and output nodes so that the networks
could represent the full range of possible positions of the Othello board, but this did not constrain the moves themselves in any way. As we saw, in contrast to BACON, the networks had to learn to play with no previous knowledge of the game; they included no hand-coded rules or heuristics that guided their moves; the strategies that evolved were discovered simply by playing (and winning or losing) the game. Moreover, the connections in the networks were originally given random weights; the starting game positions used as the networks learned were always picked at random; and the mutations and recombinations of the networks that were introduced in each generation were also random.

But we can say more. The published paper does not report this fact, but the programmers had themselves originally been unaware of mobility strategy. (Miikkulainen, personal communication). When they saw that the networks had learned to snatch victory from apparent defeat in the last moves of the game, they did not know what was going on. To find out, they sent move summaries to a world champion player to analyze, who identified what the networks were doing. Although the networks' did not consistently pick the best moves that an expert in mobility strategy would, their moves were mostly good ones, and clearly recognizable as being like those that a human player who was beginning to learn mobility would make. Given that the programmers had been previously unaware that such a high-level strategy even existed, there is no way that they could have even inadvertently designed in this striking discovery.

Nor is this an isolated example. In a second case, Moriarty and Miikkulainen were again using genetic algorithms to evolve neural networks, this time to automatically control a robot arm so it would avoid obstacles. (Moriarty and Miikkulainen, 1996) The networks could control the multiply-articulated arm by turning on or off independent motors that were at each of the joints, and after generations of evolution, learned to do this so that the arm could successfully move without bumping into anything most of the time. At one point, the researchers noticed an unusual motion in the robot arm; in certain cases it would have to move in one direction to avoid an obstacle, and it could do so successfully, but only by means of a jerked motion that began in a different direction. It was only when they noticed this odd motion and tried to figure out its cause that
they noticed that one of the joint motors was not working. (Miikkulainen, personal communication). The jerked motion was a way the networks had found as a work-around to overcome the problem. Here was a novel discovery of a movement that overcame a mechanical failure that had been unknown to the researchers. Once again, there seems to be no way that this discovery could have been even inadvertently designed in.

So, the cases are a good model of discovery, but are they a good model of scientific discovery? The answer to this question is less clear, for neither example involves scientific discoveries in a straightforward sense. It would be worthwhile research to see whether evolving neural networks could learn to discover regularities in data sets comparable to those laws discovered by BACON. However, even though the two cases we have discussed may not be canonical scientific discoveries, they do serve as a relevant analogy. In particular, they fit Simon’s original abstract framework that takes scientific discovery to be a species of problem solving. The play strategies the networks discovered in Othello are solutions to the problem of how to win in a particular kind of competitive struggle. The work-around motion discovered by the robot arm controller was a solution to an engineering problem. Thus, if the problem-solving model is appropriate for scientific discovery, these cases should also qualify.

One might object that what the networks are doing should not be considered problem-solving, in that that concept involves deliberate, goal-directed action. But such an objection would just beg the question against Darwinian mechanisms. Admittedly, we may typically think of problem-solving in terms of the intentions of a conscious problem-solver, but there is no good reason to limit the concept in this anthropomorphic manner. Biologists often use the concept of problem-solving when analyzing relationships among organisms in an environment. Indeed, it is common in such biological contexts to speak of evolution as “discovering a solution” to some adaptive problem, so it is completely appropriate to continue to use that terminology in speaking about what evolving neural networks are able to do.
6. CONCLUSIONS

We are now in a position to assess Simon's and colleagues' view as it relates to our original question about scientific discovery and evolutionary epistemology. They are correct that scientists do often make explicit use of heuristics in their attempts to solve scientific puzzles and in this sense there is a kind of logic of discovery that can be described and modeled. The unanalyzable insight of genius is not required. It is also incontestable that for a given problem (though not for problems in general), some heuristic paths to discovering its solution will be more efficient than others. But heuristics themselves have to come from somewhere, and this leaves ample room for the insights of evolutionary epistemology. It could also be that a Larmarckian evolutionary model better characterizes some scientific discoveries, since scientists are able to incorporate information directly from the environment (whereas the central dogma of genetics rules that out for biological organisms). At the most basic level, however, the logic of discovery may have to be a Darwinian logic (which, of course, is not entirely random because of the role of natural selection), and the cases we have examined show that that approach is viable. Simon and his colleagues are wrong in thinking that the process of discovery "must" be impregnated with purpose and reason. Can Darwinian mechanisms, devoid of purpose and reason, make novel discoveries of complex, previously unknown patterns? Yes, they can.

REFERENCES


Lyman Briggs School and Department of Philosophy
Michigan State University