JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2001, **76,** 351–361 NUMBER 3 (NOVEMBER)

UNIVERSAL PLOTKINISM: A REVIEW OF HENRY PLOTKIN'S DARWIN MACHINES AND THE NATURE OF KNOWLEDGE

C. D. L. WYNNE

UNIVERSITY OF WESTERN AUSTRALIA

Plotkin's *Darwin Machines and the Nature of Knowledge* (1993) is a major contribution to the field of evolutionary epistemology and universal Darwinism. Evolutionary epistemology is the idea that evolution is a knowledge-gaining process. Universal Darwinism holds that processes of variation and selection can be observed at different levels from the primary level of biological evolution (where genes code for phenotypes) through to individual learning and culture (where the units of variation and selection are not so clear cut). Although antithetical to behaviorism, large parts of Plotkin's thesis can be recast in nonmentalistic terms and exploited by behavior analysts. In particular, Plotkin's arguments for a strong commonality of process between biological evolution and individual learning offer directions for progress on questions that have long interested behavior analysts, such as: Why do some organisms learn? How did learning evolve? What is the relation between behavior and evolution? Although the paths of connection between evolution and individual behavior that Plotkin sketches are not yet fully clear of confusion, his is undoubtedly a very stimulating direction to explore.

Key words: evolution, epistemology, universal Darwinism

> So it is merely a human conceit to think that knowledge is something that is both unique to our species and located only in our heads. Knowledge is a pervasive characteristic of all of life. It exists in all adaptations in all living creatures. (Plotkin, 1993, p. 229)

"Why study animals?" It is a fair question, and one that I get asked by many different people—from academic colleagues to random individuals on planes—and that consequently can be answered at many different levels. Surely many readers are asked the same question under similar circumstances, and with growing frustration by department chairs who have to find funds to cover the ever-increasing costs of compliance with animal welfare directives.

For me, the answer to this question is bound up at several levels with the name Henry Plotkin. I distinctly remember as an undergraduate finding Plotkin the only lecturer in psychology who could hold my attention with his compelling arguments about evolution and behavior. There was nothing showy about his style, but Plotkin's serious thoughtfulness was inspirational.

In 1993, with the publication of *Darwin Machines and the Nature of Knowledge,* Plotkin went from being a philosophical evolutionary psychologist known to a few intimates to a public figure in the ongoing debates about the role of behavior in evolution and the relation between human and animal psychology.

Plotkin is no friend of behavior analysis. His text is strewn liberally with mentalistic terms; he is critical of the idea that there could be interesting commonalities in learning abilities between arbitrarily chosen species; even his definitions of behavior and the importance of rationality are not likely to sit well with behaviorists. And yet there is much here that addresses problems that behavior analysts have struggled with for decades. Where does behavior come from, and what is its role in evolution? Why do some animals learn? What are the limits to learning? And what is the relation between evolution and culture? These are just a few of the issues that Plotkin grapples with—and, in my opinion, makes significant progress on—in this well-written book. The concepts may be tricky and the thoughts deep, but the writing is careful, with gentle explanations of unfamiliar territory.

In this essay I have two aims. First, I wish

Plotkin, H. (1993). *Darwin machines and the nature of knowledge.*Cambridge, MA: Harvard University Press.

My thanks to James Chisholm, Kevin Durkin, John Dunn, Cecilia Heyes, and John Staddon for helpful comments on earlier drafts.

Correspondence may be addressed to Clive Wynne, Department of Psychology, University of Western Australia, Nedlands, Western Australia 6907, Australia (E-mail: clive@psy.uwa.edu.au).

to introduce Plotkin's form of universal Darwinism to a behavior-analytic audience. Plotkin may not be sympathetic to behavior analysis but there is much that behavior analysts can gain from Plotkin. Second, I aim to identify areas in which more work is needed. Universal Darwinism in its Plotkinian form is a powerful and exciting thesis, but it is not a complete theory. I want to encourage others to push the idea forward.

PLOTKIN'S THESIS

Plotkin's thesis can be summarized in three statements: (a) Evolution can be characterized as a process, independent of the structures and mechanisms in which that process is instantiated. (b) This process is an epistemological or knowledge-gaining one. (c) The oldest, most basic form of the evolutionary process is instantiated in genes and phenotypes. This ''primary heuristic'' has spawned other systems—in particular individual learning, immune responsiveness, and cultural transmission of knowledge—that operate according to the same process and are constrained and informed by the primary heuristic to be both knowledge gaining and adaptive.

The first premise sets the stage for ''universal Darwinism''—Dawkins' (1983) phrase for the idea that a Darwinian evolutionary process can operate in other instances than just evolution proper. Although the name is fairly new, this idea has been around since Darwin (1859), Huxley (1874/1893), and James (1897) first explored the implications of evolution by natural selection in the 19th century. The most important protagonist of universal Darwinism in the 20th century has been Campbell (e.g., 1974). Dennett (1995) likens it to ''universal acid.''

For Plotkin, it seems that the precise specification of this mechanism is not centrally important. He notes several different ways that Darwinian evolution can be construed. For example, it can be viewed as implying three principles: (a) Phenotypic variation exists (individuals differ in structure and function); (b) these variants show differential fitness (different phenotypes have different rates of survival and reproduction); and (c) fitness is heritable (traits that contribute to fitness of parents will be found in the offspring). This can also be expressed (as by Lewontin, 1970) as a replicator-interactor-lineage. There are replicators, there is interaction with the environment, and there is a lineage of these replicators. Or it may be summarized simply as blind variation and selective retention (Campbell, 1974). A range of variants is generated (blind in the sense that the system of generation is uninformed as to which variants are likely to survive), and these variants are then selectively retained to become the originators of the next phase of variation. Yet another alternative is to summarize Darwinian evolution as a ''g-t-r'' heuristic—a system that generates, tests, and then regenerates. For Plotkin, these are alternative summations of the same principle, and the important point about that principle is that it says nothing about the substrate in which it is embedded or the mechanism at any deeper level by which these operations are achieved. Dennett (1995) adopts a similar approach.

Plotkin's second principle—that the evolutionary process is an epistemological or knowledge-gaining one—stems from Lorenz (1977). There are two sides to this. One is to argue that the process of biological adaptation is a knowledge-gaining one: ''This informing relationship between parts of organisms and their world is knowledge, or biological knowledge if one prefers'' (Plotkin, p. xv). ''Knowledge, as commonly understood, and adaptation are closely related . . . adaptation and knowledge are one and the same thing. Adaptations *are* knowledge'' (p. 116). What Plotkin means here is that the process of adaptation, as it comes to produce organisms that reflect in their structure aspects of the organization of the outside world, can be said to be gaining knowledge of that outside world. Goethe (quoted in Lorenz, 1977) said, ''Wär nicht das Auge Sonnenhaft / Die Sonne könnte es nie erblicken'' (If the eye were not sun-like / It could never see the sun). In what sense does the structure of the eye indicate knowledge of the sun (or more precisely, of the qualities of sunlight at the earth's surface)? In the sense that a Martian, with nothing to work on but an earthling's eye, could deduce things about the nature of sunlight on earth. Even plants embody in their structure features of the niche they inhabit. Thus the solid trunk of a tree implies the force of gravity and the thin

leaves of Australian eucalypts imply the need to conserve water.

Conversely, Plotkin argues that what we ordinarily mean by *knowledge*—the individual human capacity to know by the senses as in ''I know the sun is shining today''—is itself a form of biological adaptation. (Other European languages draw a clear distinction between knowing by the senses—*wissen* in German—and knowing in the sense of being acquainted with—*kennen* in German.) Here Plotkin uses mentalistic terminology, but his point is not essentially mentalistic. Plotkin's argument for knowledge as adaptation is twofold. First, all individual knowledge (read adaptive behavior) is generated by mechanisms that are exposed to natural selection of phenotypes and inheritance of genes (what Plotkin calls the primary heuristic) and, as such, can be viewed as adaptations. Second, and this is a distinct idea, these mechanisms of individual learning themselves operate using an analogous selectionist process—a process of blind variation and selective retention, a replicator-interactor-lineage, or a g-t-r heuristic, or however we care to characterize it. ''Human knowledge is just one kind of a much wider biological knowledge. When we come to know something, we have performed an act that is as biological as when we digest something'' (Plotkin, p. xvi). ''Behavior is a particular kind of knowledge, even though that behavior contains no necessary element of thought, reflection or memory'' (p. 120).

So this leads to Plotkin's third principle—that there is a nested hierarchy of selectionist processes. Evolution as conventionally understood involves the selection among phenotypes and a lineage of genes. But, Plotkin argues, when the environment is unstable on a time scale that this primary heuristic cannot adapt to (crucial here is the ''generational down time,'' which is the time between the production of a new individual—conception—and that individual being ready to produce offspring itself), then secondary heuristics may evolve that enable the individual to adapt its behavior (to gain knowledge) within its own lifespan. ''The secondary heuristic is functionally always tucked under the wing, so to speak, of the primary heuristic'' (p. 161).

Individual learning programs (the secondary heuristic) are a consequence of, and driven and constrained by, the primary heuristic. There can be no tabulae rasae, no blank slates ready to be written on by whatever life may throw their way. Rather, all learning in all species is highly focused by evolution towards problems that are likely to come along in the life of a particular species. ''Species-typical intelligence rather than some identical intelligence that spills over from one species to the next is what I am arguing for'' (p. 165). In turn, the secondary heuristic of learning by individual organisms in their own lifetimes can give rise to cultural knowledge by the exchange of knowledge between individuals. This is the tertiary heuristic, the realm of the meme.

Universal Plotkinism is thus universal Darwinism at three nested levels: the primary level of genes and phenotypes, the secondary level of individual learning, and the tertiary level of cultural knowledge. The potential of Plotkin's approach lies in the process that glues these three levels together: the universal selectionist process.

IS LEARNING ALWAYS AND ONLY A SELECTIONIST PROCESS?

If we could accept that knowledge gain (learning, adaptive behavior) were only ever a selectionist process, then this would be a very exciting development for at least two reasons. First, at the most abstract theoretical level it would be a great advance in our understanding of learning if we could be sure that wherever it might appear it always had to have the same deep structure, any appearances to the contrary notwithstanding. Second, it would truly be a substantial advance in our understanding of the relation between biological evolution and individual learning if we could say that they are bound together in this relation of closely similar processes. So what are Plotkin's arguments for this important claim? Is there a risk here that the essentials of one process might get thrown out in order to force a similarity with the other?

Given that I am sympathetic to Plotkin's position, I am disappointed to find that his three arguments to support the idea that individual learning should be selectionist are not strong. His first argument is simply that one position or another has to be taken on this issue, so why not adopt the selectionist

position. Plotkin's second argument is that creativity demands selectionist processes. The alternative, what he terms instructionalist processes, cannot lead to creativity:

> Creativity cannot occur if change is slavishly tracked by instructionalist devices. So what we see here is that while selection can mimic instruction, the reverse is never true. Instructional processes can never lead to creativity. To go beyond experience requires the generation of something from inside the knower, and only an intelligence driven by selectional machinery can do that. (p. 172)

There may be something in this second argument (although Plotkin does not develop it with any review of creativity and novelty in behavior), but it is hardly enough in itself to carry the weight that the thesis of process identity between biological evolution and learning requires.

His third argument is an appeal to the elegance and attraction of the idea:

> The third reason . . . is one of parsimony and simplicity. If the primary heuristic works by selectional processes . . .; if . . . culture works by selectional processes . . .; and if . . . the immune system works by selectional processes . . .; then why should one be so perverse as to back a different horse when it comes to intelligence? (p. 172)

Now parsimony is a good ground to prefer one theory over another when both explain the observed facts, but on its own it is not sufficient to convince us that a theory is adequate.

Surely what is needed is a review of different forms of learning in humans and other species (and why not in artificial intelligence?) to demonstrate a deep similarity that lurks beneath the surface diversity. But here Plotkin seems to be hindered by a distaste for general process learning theories, which is ironic given that his great achievement may be the grandest and most general of all general process learning theories.

> Rat intelligence must be understood in the context of rat genes, and human intelligence can only be understood in the context of human genes. Insofar as rat genes are different from human genes, then so too is rat intelligence different from human intelligence. (p. 165)

The problem here lies in the ''insofar as'' phrase; even if we knew how many differences there are between the rat genome and the human genome (something we are probably not that far from knowing), we have no way in the foreseeable future to estimate how much or what kind of difference in ''intelligence'' should result from that difference in genes. The fallacy of computing differences in intelligence from differences in genes is most commonly seen in arguments that chimpanzees must be self-aware because they share 98.4% of their genes with *Homo sapiens* and *Homo sapiens* is self-aware (see Wynne, 2001, for a fuller discussion). But in any case, Plotkin hedges his bets on this point:

> However, the notion of multiple intelligence*s* rather than some single intelligence does *not* imply that intelligence operates through necessarily different processes in different species. Quite the contrary: it is much more likely that the process of intelligence is usually the same across species, especially species within some restricted taxon such as a class (mammals or birds) or supraclass (vertebrates). (p. 165)

So he is arguing both that learning mechanisms should differ between species in principle, but that in practice similarities are to be expected across groups as broad as classes.

Surely what is needed to make his argument compelling is an assessment of different kinds of learning. Some may be plainly selectionist or easily construed as such; others, the interesting ones for Plotkin's theory, would not appear to be selectionist on superficial inspection but, if Plotkin is right, would reveal their selectionist nature on deeper investigation.

The most obviously evolutionary form of learning I can think of is the form of artificial intelligence known as genetic algorithms (Holland, 1992; Koza, 1992, 1998). In a genetic algorithm a digital computer is programmed in a manner directly inspired by the activity of genes and phenotypes. Optimal solutions to problems are found by starting with an initial pool of programs that vary in some way. Each of these programs is tried on the problem to be solved. The least successful programs are discarded and a new set of putative solutions is created by recombination of the remaining program set. Mutation also takes place by random alteration of pieces of the program code. (Mutation turns out to be relatively unimportant.) This process repeats

through many iterations until a sufficiently ''fit'' solution to the problem has been found. The genetic algorithm approach may come up with solutions that are smarter (more compact, faster; whatever the dimensions of the problem and its solution may be) than those that any human programmer had been able to think of. Here the analogy with biological evolution is explicit—the case for commonality of process is easily made.

But it is not always so. Another form of artificial intelligence informed by biological processes is the neural network approach (Rummelhart et al., 1986). In a neural network, a digital computer is programmed so as to emulate the behavior of a number of interconnected neuron-like elements. Crucial to these elements is that each possesses a level of activity that it can propagate to the other elements with which it is interconnected. Problem solving in this environment proceeds by the interconnection weights between the units being progressively altered according to certain rules. These rules are designed to reduce the difference between the output of the network and the desired output. The desired output represents the desired solution to the problem that has been set. Neural networks of this kind can be highly successful in producing an optimal solution to problems in many domains, including categorization and conditioning (e.g., Grossberg, 1998; Kehoe, 1988; Schmajuk, 1997; Schmajuk, Urry, & Zanutto, 1998). Where are the variation and selection here which Plotkin views as the essentials of any learning process? This is not so simple. Perhaps the variation lies in the initial random weights of the interconnections between units that are applied to the network to get it started; and perhaps the rule that updates these weights might then be considered as a process of selection (in which case it seems that we have a process of one phase of variation followed by many phases of selection without any more variation). And what about the case in which the network starts out with all connection weights equal? Is this a special case of variation—one in which variation equals zero? Or should we perhaps view the changing connection weights as the variation, in which case, where is the selection? Alternatively, the changing output of the system might be the relevant variation, and perhaps the desired output and the rules that change the interconnection weights should be construed as the selection mechanisms. The problems involved in viewing neural networks in selectionist terms may not be insurmountable, but neither is the solution self-evident. And perhaps (although I think it lays out the issue in conveniently stark terms) artificial intelligence does not have to show the selectionistic properties that all forms of natural intelligence are supposed to.

One form of natural learning that is close to the hearts of readers of this journal is operant conditioning. The selectionist nature of operant conditioning has been recognized before, not least by Skinner himself (1981) and by Staddon (1979, 1983), as well as others. Plotkin also gives an account of instrumental learning. The analogy to evolution appears to be obvious. An individual confronted with a problem generates a set of variant responses; these are tested against the problem and the most successful form the basis for a new set of variant responses. As conditioning proceeds the variability of behavior usually declines until a single most efficient solution behavior (or a small set) remains. The similarities between shaping and Plotkin's primary heuristic of biological evolution are self-evident. There is variation followed by selection leading to a gradually changing population of variants.

But does this superficial similarity hide deeper differences? Hull, Langman, and Glenn (in press) have pointed out that aside from the similarities, there are also differences in process between operant conditioning and primary biological evolution. Thus, in biological evolution there are many variants concurrently operating in the outside world in direct competition with each other; in operant conditioning, only one behavior can be tested against the environment at a time. The units that vary in operant conditioning are also not obvious. Finding any units at all in a continuous stream of behavior is hard enough, never mind the problem of how these units are to be seen as stable (or changing only relatively slowly) in the clearly plastic output of behavior. And, if we found some units of operant behavior, in what sense is behavioral variation ''blind''? Even trial-and-error learning is not random.

And what about Pavlovian conditioning?

Can variation and selection be identified there? Here Plotkin argues that we should look for hidden internal generation of variants and their selection. He takes the example of an infant coming to associate its mother's facial features with her smell:

> The brain of the infant may internally generate multiple representations of the mother in which are mixed components of the actual mother as well as features of the world that are not the mother. . . . These multiple representations of the mother are tested perceptually against the appearance of the mother on many occasions, as well as being acted upon to test the accuracy of the representation. . . . The infant *creates* a representation of its mother by generating, possibly sequentially and possibly simultaneously, numerous brain states representing her, and successively selecting out and reducing these brain states until only one composite brain state is left. This, of course, is a selectionist process. (p. 167)

Pavlovian conditioning might work like this, but I am not aware of any studies demonstrating that it does. We do understand a little of the brain mechanisms underlying classically conditioned eye blinks, for example, and there is nothing to suggest the initial variability that Plotkin proposes (e.g., Krupa, Thompson, & Thompson, 1993). Perhaps here, as in the case of neural networks considered above, Plotkin would argue that the variation side of the processes could be reduced to a minimal level but still count as variation.

In summary then, although I share Plotkin's intuition that such a neat idea as a commonality of algorithm among all learning processes *should* be true, Plotkin has not marshaled the evidence we need to be convinced that it is. Some forms of individual learning share a superficial similarity with biological evolution (operant conditioning, genetic algorithms), although even here there are questions that can be asked about just how compelling that similarity is. For other common forms of learning, however (e.g., Pavlovian conditioning and neural networks), the analogy is much less clear, and substantial special pleading is required.

Plotkin's analysis of the possibility that all forms of individual learning can be cast as variation and selection algorithms is limited by his reluctance to be specific about the exact process he is talking about. His argument seems to be that the different formulations of the variation and selection idea are essentially the same process described in slightly different ways. But any attempt to show that the many forms of individual learning are all consistent with a variation and selection process demands a clear definition of what the essentials of such a process are.

BEHAVIOR

Plotkin's internalization of knowledge in mentalistic terms might be seen as explicitly disregarding behavior (e.g., p. ix). But, although he is happy to use mentalistic terminology, Plotkin does operationalize these ''incorporations'' as physical and biological states.

Surprisingly, given his lack of sympathy with operant psychology, Plotkin adopts an extremely ends-oriented definition of behavior. To qualify as ''behavior'' in Plotkin's terms, an action has to ''be directed towards effecting change in the world outside the behaving creature'' (p. 105). Plotkin uses coughing as an example. When an individual coughs as a protective reaction to noxious material introduced into the air passages this is *not* behavior, whereas coughing to gain attention or sympathy *is* behavior. Similarly, the action of a male dog in lifting its hind leg to urinate counts as behavior because the animal is marking its territory, whereas the action of a female dog in urinating without lifting a hind leg is *not* behavior because the aim of marking a territory is absent.

Plotkin appeals to Piaget for this extremely operant definition of behavior. But there are echoes too of Rachlin's (1994) teleological behaviorism in the demand that behavior be viewed in terms of its ultimate adaptive consequences. Sympathetic as this definition may be to behavior analysts, it does leave a residue of nonoperant behavior that is not behavior at all on Plotkin's definition. What can this nonbehavior be? It is just action—movement without any aim or purpose, at least in the present. This offends my sense of the use of the word *behavior,* in that operant behavior is a subset of behavior, but an action can still qualify as behavior even if it is not operant. It also seems to lead to a messy need to iden-

tify the ends of a behavior in the here and now if it is to qualify as behavior and not mere action. Furthermore, it detracts from Plotkin's message that there will always be an ultimate purpose to any adaptation, including behavior.

Plotkin's distinction between instinctive and rational behavior seems to be even more problematic. I do not mind his revival of the term *instinct* for relatively inflexible species-specific action patterns. This ties them neatly into Plotkin's primary heuristic of biological evolution and makes clear that they are the products of an evolutionary process of variation and selection. The problems start with the definition of all the rest of behavior—behavior that adapts through modification controlled by the brain in an individual's lifetime—as ''rationality.''

Plotkin's insistence on noninstinctive behavior being labeled as rational (''the product of reason, intelligence, learning and memory—in short, rationality,'' p. 125) is not helpful. Surely, even if one accepts that rationality is capable of an operational definition, there can be aspects of goal-directed behavior that are neither instinctive on the one hand nor rational or thoughtful by any useful definition on the other. Pavlovian conditioning can often proceed without conscious awareness in humans (e.g., Oehman, 1988) and therefore can surely not be considered thoughtful, yet it is a clear example of adaptive behavior modified through experience during the lifetime of the individual. Much learning by humans and other species may not be instinct, but neither is it rational in any of the senses of that word. Plotkin also fails to consider that even behavior that we might want to label rational for its complexity and ingenious goal-directed nature might be the product of simpler, more mechanical processes. Braitenberg's (1984) delightful little vehicles are one example of an approach to building behavioral complexity out of the simplest components. My own attempts to model transitive inferential reasoning using simple associative mechanisms are another (Wynne, 1998). Perhaps it is the choice of label that is causing problems—if Plotkin had stuck to calling noninstinctive behavior ''individual learning,'' my concerns would have been lessened. The discussions of action, instinct, and rationality are made unnecessarily confusing by definitions of terms in which only the second of the three is being used in a sense that would be familiar to a reader.

NESTING HEURISTICS

Some part of Plotkin's problem with traditional general process learning theories (notwithstanding that his own approach is the general process to end all general processes) relates to his belief in the nesting of heuristics. Individual learning, Plotkin's secondary heuristic, ''because it is nested under the primary heuristic, is *always* primed in some way rapidly to gain particular forms of knowledge'' (p. 181). This has three consequences. First, there can be no tabulae rasae in humans or any other species. Plotkin emphasizes notions of preparedness in conditioning: the fact, for example, that rats are very quick to learn about tastes that lead to sickness because this effectively protects them from poisoning (Garcia & Koelling, 1966). This is because in the past there has been strong selection pressure for animals to learn rapidly about strange tastes that are poisonous. He also reviews work on human cognitive biases showing that our own powers of reasoning are a good deal more constrained than we might like to think, and furthermore that the effectiveness of our logic depends greatly on the context within which a problem is posed. Thus, we are highly effective at detecting social cheats but are rather poor at solving problems of equivalent logical complexity posed in terms of selecting vowels and consonants.

Second, individual learning (and culture) must be constrained by Plotkin's primary heuristic—biological evolution. I have already quoted Plotkin arguing that rat intelligence must be different from human intelligence. According to Plotkin, the primary heuristic ''sets up'' learning algorithms in individual animals so that they can learn only about matters that biological evolution has found valuable in the past. The problem of the relation between evolution and individual learning is an important one, and one that has evaded human understanding for a long time. Plotkin's suggestion of a nested relation must be true at some level, but, as a basis for scientific explanation, it seems to proscribe too little. Clearly our ability to learn is not limitless

(nor is that of any other species), but we do not know what those limits are, nor does Plotkin's approach give us any means to predict what they might be. Often individual learning and culture seem to wander far and free from anything of obvious fitness advantage.

Third (to some extent this is a continuation of the second point), Plotkin argues that human knowledge must be domain specific, in the sense of Fodor (1983). Knowledge about different aspects of the world may have evolved independently, and consequently there may be little sharing of knowledge between domains. One strong example of the modularity thesis is face recognition. We recognize faces far more readily than other objects of equivalent complexity, but this ability is limited to properly oriented faces with the characteristic organization of a human face. The ability does not apply even to upside-down faces. Lesions to specific brain regions lead to deficits specifically in this ability and no other aspects of visual perception or memory (Murray, Yong, & Rhodes, 2000). Plotkin believes that this pattern is the norm for individual knowledge, not the exception. The notion of a nesting of knowledge-gaining heuristics sets the scene for the evolution of learning in individual organisms.

EVOLUTION OF THE CAPACITY FOR INDIVIDUAL LEARNING

Instinctive behavior is governed by the primary heuristic of biological evolution. The variation in this behavior comes about through variation in phenotypes caused by varying genes and epigenesis, just like the variation in any other phenotypic trait (Dawkins, 1986). On occasion (according to Plotkin, examples can be found in five of the 25 phyla of animals), this process is unable to track changes in the environment because of their speed and unpredictability (''predictable unpredictability,'' he calls it). In such cases,

> Animals must evolve additional knowledge-gaining devices whose internal states match those features of the world that we are calling short-term stabilities. Such tracking devices would be set in place by the usual evolutionary processes of the primary heuristic and hence would operate within certain limits. But the exact values within those limits that these devices will settle to, and for how long, are not within the power of genes to decide. So devices such as these have a degree of autonomy in their functioning that makes them partially independent of both genes and development. (p. 149)

But if individual learning offers this additional adaptability, why do relatively few animals possess it? Plotkin argues for a cost-benefit analysis to explain the rarity of noninstinctive behavior among animals:

> For an instinctive behavior the instructions that have to be carried by genes, and the instructions that later have to be carried in the brain and the computations that have to be carried out by the brain to produce the behavior . . . will be fewer than those required for that same behavior to be acquired and controlled by learning and thought. (p. 132)

Is there a trap in the intuitive plausibility of this argument? Is it in fact true that instinctive behavior can be generated in the individual at lower cost than individually learned behavior? These arguments have a plausible ring, but they are not, in fact, straightforward. How do we measure these costs? Few if any attempts have been made. It is at least possible that some forms of learned behavior might be coded very efficiently in the brain in such a way that the extra expense of learning compared to instinct may not be a significant factor. These are important questions, and I cannot think of any attempt to grapple with them except McFarland and Bösser (1993).

MEMES AND CULTURE

Individual learning, where the knowledge resides in the brain of the learner, is not the only example of knowledge gain that goes beyond the genes. In humans and a handful of other species, knowledge may be exchanged among individuals in the form of culture. Plotkin is emphatic that culture, just like individual learning, is a product of primary evolution and, as such, must operate ''within certain limits'' (p. 149), even if these cultural practices ''have a degree of autonomy in their functioning that makes them partially independent of both genes and development.'' This hedging must surely be appropriate, but it does leave the account rather toothless. Any particular cultural practice might be adaptive,

or it might be an example of the autonomy that individual and cultural learning processes possess. There seems to be no way to know which is the case.

A major problem in the development of a selectionist account of culture is that of identifying the units of replication and selection. Plotkin aligns himself with Dawkins' (1976) notion of the *meme*. Huxley recognized the need for a unit of states of consciousness if a Darwinian analysis is to be applied to thought, and called these units "ideagenous molecules" (1874/1893, p. 239). Dawkins' term may be easier on the tongue, but it is not necessarily much easier to work with.

Plotkin argues that memes may vary and are selected. It is not clear to me that the fact that people may hold a range of opinions on, say, market forces (to take one of Plotkin's examples) is the same as saying that there is a unit of anything that can be said to be varying in the sense required by selectionist theory. As my understanding of market forces changes, it is (at the very least) not intuitively obvious how this maps onto a process of variation and selection. Plotkin acknowledges that memes may blend and show multiple parentages, and that these are differences from the process of primary evolution. Memes also show very high rates of mutation in that they vary frequently at each retelling (Boyd & Richerson, 2000). He also states that the brain is complex and dynamic enough to permit a selectionist interpretation of its operation (calling on Edelman's, 1989, neural Darwinism to back him up). But the brain is large and complex enough that it might be (and has been) conceived of in a great many different ways. We need some evidence and argument to convince us that there really are units of thought and culture that vary and are selected, and that this is a compelling description of the whole of the brain's operation. Although others have argued that we can talk about memes without wishing to imply a very strong analogy between genes and memes (e.g., "genes and memes are both replicators but otherwise they are different"; Blackmore, 1999, p. 66), Plotkin's process-driven approach surely demands that the analogy be spelled out. Plotkin is apparently about to publish a new book on culture and evolution, so perhaps we should reserve opinion until we see it.

WHO'S AFRAID OF UNIVERSAL DARWINISM?

Universal Darwinism is an idea that has been around at least since James (1897), and it recurs about every 20 to 30 years as a motif in the discussion of the relations among culture, individual learning, and biological evolution. Skinner dabbled in it a little himself (Skinner, 1966). Popper had a rather ambivalent attitude to Darwinian evolution (considering it at different times both irrefutable and refuted; Popper, 1972), but also saw an analogy between the kind of learning process he proposed for science and the variation and selection of biological evolution. What is it about universal Darwinism? Why does it engender such excitement among a few, but never really developed as a practical theory of life and learning?

I think at least part of the answer lies in the very universality of the process being claimed. The ultimate theory of everything cannot make many specific predictions. Given the diversity of animal and human lives, the ultimate theory of evolution, behavior, and culture cannot be a very constraining model. Any viable account of the relation between evolution and culture (such as Plotkin's) has to allow culture a fair amount of independence from immediate fitness concerns; otherwise, it would be too easily refuted by the things people do that thwart their own biological fitness.

Another important part of the reason why universal Darwinism has not had the impact it deserves is because its champions, in their legitimate enthusiasm to demonstrate connections between apparently disparate processes (of biological evolution, individual learning, culture, and others), have avoided a strong definition of what the essential things are that these processes have in common. This has made it difficult to critically assess their claims for commonality. There is a danger that this blurring of the details of processes may in fact be so great that evolution, learning, culture, and so forth have no more in common than any other trio of adaptive processes. There is, after all, a branch of analysis devoted to studying any and all dynamic systems in the abstract: *dynamical systems theory*. The possibility cannot be entirely dismissed that systems of biology, learning,

and culture have no special relation at all beyond all being dynamic systems.

My hunch, however, is that Plotkin is right: There is a deep, meaningful, and interesting commonality of process among all forms of adaptation, and learning in individuals is rightly seen as one of these processes. But I think this contention will only convince, and the power of the idea will only come through, once we have a strong definition of what that process is. Plotkin, like other evolutionary epistemologists, emphasizes variation and selection as the core processes of evolution. But what about other aspects of evolution, such as speciation? Speciation is clearly an important process in evolution, but does it matter that there is nothing that can be considered analogous to speciation in operant conditioning? Or, if one can see an analogy to speciation somewhere in operant conditioning, can one move on to Pavlovian conditioning and so on until one finds a form of individual learning or culture that lacks something analogous to speciation? Absent a strong definition, the claims of common process are so many promissory notes still waiting to be cashed into hard currency.

Armed with such a definition, universal Darwinists will be able to contribute a great deal to studies of adaptive behavior, individual learning, cultural knowledge, and the relations these three have with each other and with biological evolution.

Plotkin's book is a very exciting step along this path. This is the first book ever to be dedicated solely to universal Darwinism, and it is particularly well written. Plotkin explains complex and abstract concepts at a level that anyone with a curiosity about such matters is sure to understand. If I have lingered in this review on areas in which I think work still needs to be done, it has been solely in the hope of motivating others to pick up those items of unfinished business. My hope is that Plotkin's book will push universal Darwinism back into the mainstream of debate about behavior and evolution—a debate that is central to arguments that I (and I'm sure other readers) appeal to when asked why we study animals.

REFERENCES

Blackmore, S. (1999). *The meme machine*. Oxford, England: Oxford University Press.

Boyd, R., & Richerson, P. J. (2000). Meme theory oversimplifies how culture changes. *Scientific American, 283*(4), 70–71.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). La Salle, IL: Open Court.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John Murray.

Dawkins, R. (1976). *The selfish gene*. Oxford, England: Oxford University Press.

Dawkins, R. (1983). Universal Darwinism. In D. S. Bendall (Ed.), *Evolution from molecules to men* (pp. 403–425). Cambridge, England: Cambridge University Press.

Dawkins, R. (1986). *The extended phenotype*. Oxford, England: Oxford University Press.

Dennett, D. C. (1995). *Darwin's dangerous idea*. London: Penguin.

Edelman, G. M. (1989). *Neural Darwinism: The theory of neuronal group selection*. Oxford, England: Oxford University Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123–124.

Grossberg, S. (1998). Neural substrates of adaptively timed reinforcement, recognition, and motor learning. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 29–85). Hillsdale, NJ: Erlbaum.

Holland, J. H. (1992). Genetic algorithms. *Scientific American, 267* (1), 44–50.

Hull, D. L., Langman, R. E., & Glenn, S. S. (in press). A general account of selection: Biology, immunology and behavior. *Behavioral and Brain Sciences*.

Huxley, T. H. (1893). On the hypothesis that animals are automata, and its history. In *Collected essays* (Vol. 1, pp. 199–250). London: Macmillan. (Original work published 1874)

James, W. (1897). Great men, great thoughts, and the environment. In *The will to believe and other essays in popular philosophy* (pp. 163–189). Cambridge, MA: Harvard University Press.

Kehoe, E. J. (1988). A layered network model of associative learning: Learning-to-learn and configuration. *Psychological Review, 95,* 411–433.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

Koza, J. R. (1998). Using biology to solve a problem in automated machine learning. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 87–126). Hillsdale, NJ: Erlbaum.

Krupa, D. J., Thompson, J. K., & Thompson, R. F. (1993). Localization of a memory trace in the mammalian brain. *Science, 260,* 989–991.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics, 1,* 1–18.

Lorenz, K. (1977). *Behind the mirror*. London: Methuen.

McFarland, D., & Bösser, T. (1993). *Intelligent behavior in animals and robots*. Cambridge, MA: MIT Press.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting

the perception of upside-down faces. *Psychological Science, 11,* 492–496.

Oehman, A. (1988). Nonconscious control of autonomic responses: A role for Pavlovian conditioning? *Biological Psychology, 27,* 113–135.

Plotkin, H. (1993). *Darwin machines and the nature of knowledge.* Cambridge, MA: Harvard University Press.

Popper, K. (1972). *Objective knowledge: An evolutionary approach.* Oxford, England: Oxford University Press.

Rachlin, H. (1994). *Behavior and mind: The roots of modern psychology.* New York: Oxford University Press.

Rummelhart, J. L., McClelland, D. E., & PDP Research Group. (1986). *Parallel distributed processes: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Schmajuk, N. A. (1997). *Animal learning and cognition: A neural network approach.* Cambridge, England: Cambridge University Press.

Schmajuk, N. A., Urry, D. W., & Zanutto, B. S. (1998). The frightening complexity of avoidance: A neural network approach. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 201–238). Hillsdale, NJ: Erlbaum.

Skinner, B. F. (1966). The phylogeny and ontogeny of behavior. *Science, 153,* 1205–1213.

Skinner, B. F. (1981). Selection by consequences. *Science, 213,* 501–504.

Staddon, J. E. R. (1979). Operant behavior as adaptation to constraint. *Journal of Experimental Psychology: General, 108,* 48–67.

Staddon, J. E. R. (1983). *Adaptive behavior and learning.* New York: Cambridge University Press.

Wynne, C. D. L. (1998). A minimal model of transitive inference. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models for action: Mechanism for adaptive behavor* (pp. 269–307). Hillsdale, NJ: Erlbaum.

Wynne, C. D. L. (2001). The soul of the ape. *American Scientist, 89,* 120–122.

Received May 2, 2001
Final acceptance August 7, 2001