How Do Simple Rules 'Fit to Reality' in a Complex World?

MALCOLM R. FORSTER

Department of Philosophy, University of Wisconsin-Madison, 5185 Helen C. White Hall, 600 North Park Street, Madison, WI 53706, USA; Email: mforster@facstaff.wisc.edu

Abstract. The theory of fast and frugal heuristics, developed in a new book called *Simple Heuristics that make Us Smart* (Gigerenzer, Todd, and the ABC Research Group, in press), includes two requirements for rational decision making. One is that decision rules are bounded in their rationality—that rules are frugal in what they take into account, and therefore fast in their operation. The second is that the rules are ecologically adapted to the environment, which means that they "fit to reality." The main purpose of this article is to apply these ideas to learning rules—methods for constructing, selecting, or evaluating competing hypotheses in science—and to the methodology of machine learning, of which connectionist learning is a special case. The bad news is that ecological validity is particularly difficult to implement and difficult to understand in all cases. The good news is that it builds an important bridge from normative psychology and machine learning to recent work in the philosophy of science, which considers predictive accuracy to be a primary goal of science.

Key words: Bayesianism, complexity, decision theory, fast and frugal heuristics, machine learning, philosophy of science, predictive accuracy, simplicity.

1. Introduction

It is often interesting for researchers working in different areas in the science of knowledge to "compare notes." The philosophy of science is one such area, the rationality of human decision making is another, and machine learning and AI constitute a third area of epistemology. I will argue in this article that many of the points that Gigerenzer, Todd, and the ABC Research Group (in press) raise about the bounded, ecological rationality of human behavior also arise in the philosophy of science and in machine learning. The idea is that in every case, fast and frugal heuristics emerge from an underlying complexity in a process of evolution that is anything but fast and frugal. Simplicity is just the tip of a very large iceberg. The underlying complexity is important in all these fields, but especially in machine intelligence and AI.

In everyday situations, human beings often need to act on information quickly, in real time. Gigerenzer et al. (in press) therefore argue that rational decisions are

(1) *Bounded*. The best-known theory of unbounded rationality would be a form of Bayesianism, according to which a rational agent chooses the action that maximizes expected payoff, where the expected payoff is calculated from the probabilities of *all* payoffs for each possible action in every possible environment. The problem is that expected payoffs are monumentally expensive to

calculate. An alternative view (Gigerenzer et al., in press) defines rationality in terms of fast and frugal decision-making heuristics. These simple decision rules are *fast because that they are frugal* in what they take into account.

(2) *Ecological.* Rational decision-making rules are not only cost effective, but also adapted to specific (ecological and social) environments. Rationality is defined by its "fit to reality" and not merely by an adherence to general principles of coherence, such as logical consistency, or the probability calculus.

Boundedness is a pragmatic requirement of real-time decision making. If incoming information requires an immediate response, such as when the sighting of a predator requires an immediate action, then the decision mechanisms must be in place for the decision to be fast. To be fast, they must be frugal in what they take into account. However, fast and frugal decision rules are no good if they do not work. If simple rules do not lead to *appropriate* actions then they are not adapted to their environment, and they are not ecologically valid.

The requirement of ecological validity goes beyond the pragmatic requirements of speed and frugality. But it goes beyond them in a limited way. It does not require that a rule give good payoffs *in every possible environment*. In this view of rational decision rules, "fit to reality" is satisfied by the fit to the *actual* environment in which the decision maker is located. Even so, it is a requirement that is quite foreign in theories of rationality that require only that agents act optimally in accordance with their beliefs. According to such theories, decision makers act rationally even when their beliefs are entirely false or inappropriate. A taxpayer who believes that she will avoid taxes by not filing a tax return is acting rationally on a theory that says that rational action is nothing more than the maximization of *expected* payoff, where "expected" merely refers to the person's beliefs. Such decision making is not rational because it is not ecologically valid, according to the theory of Gigerenzer et al. (in press). As a consequence, an agent who acts on irrational beliefs is not acting rationally. Naturally, this raises another question: How do we characterize the rationality of our beliefs? Perhaps Gigerenzer and colleagues would give the same sort of answer, that rational beliefs are those that are formed on the basis of learning rules that are fast, frugal, and ecologically valid.

The main focus of this article is on fast and frugal learning, where learning is thought of as a decision about what to believe, or how much weight should be given to different beliefs. I use the terms "learning" and "beliefs" in a very general way, which includes classical and operant conditioning in animals, learning in neural networks or machines, and the construction and evaluation of hypotheses in science as special cases (section 2). All of these types of learning involve the extraction of information from observational or sensory data. From this broad perspective, we may also regard the extraction of information from stored data, or from memory, as relevantly similar. This process is more commonly referred to as reasoning or inference. The story of Dennett's robot (Dennett, 1984), about the well-known complexities faced in designing a machine to carry out simple instructions, is meant to motivate this generalization by reminding us that the potential unboundedness

of this process (also known as the frame problem) is a central issue in machine intelligence (section 3).

My main conclusion in this article is that there is a conservation law for complexity that says that simplicity achieved at the higher level is at the expense of complexity at a lower level of implementation. Such a "law" appears to be consistent with the examples I examine in this article. In addition to the example of Dennett's robot, I look at automated scientific discovery (section 4), in which programs such as BACON are fed data in a preprocessed form that it took human astronomers 1,500 years to develop. The no-free-lunch theorems in machine learning (section 5) show that predictions of the future based on the past do not guarantee a "fit to reality" in an a priori sense. Successful prediction must presuppose the applicability of general principles, such as time invariance, the Markov condition, and spatial symmetries, to pare down the space of rival hypotheses (in this sense, the "lunch" provided by predictive learning is not "free"). Learning in connectionist networks (section 3) looks as if it is based on a relatively simple learning algorithm, called backpropagation (although it is not clear that it is either very frugal or very fast). Even if it once looked quite simple and universal, there are now variations on this rule that work better in different contexts. And even if the rule is appropriate to the task, it may require a judicious choice of neural architecture to be in place, and a sequence of learning stages of increasing complexity.

It is clear from these examples that the complexities underlying fast and frugal learning are required in order to satisfy the ecological requirement. "Fit to reality" would be implausible otherwise. There are important lessons here for the Bayesian theory of decision making, which are developed in section 6.

Bayesians could soften the force of the unboundedness objection by supposing that fast and frugal heuristics are set in place by an "offline" comparison of alternative candidates according to their expected payoffs. I argue that this Bayesian response does not automatically succeed in satisfying the ecological requirement. It does not provide an adequate account of the rationality of the initial probability assignments, at least not in the sense that Gigerenzer et al. (in press) require.

If the study of ecological adaptation is to assume a more central role in decision theory, then it seems appropriate that the concept itself should be made more precise. Even for the special case of learning, it is not clear what it means for a learning rule, or a rule for theory evaluation, to "fit to reality." In section 7, I describe how this question arises in the philosophy of science. Traditionally, "fit to reality" would require, roughly speaking, a rule to favor true hypotheses more frequently than alternative rules. However, scientific hypotheses often have dubious theoretical implications, such as "light is a wave," "planets are embedded on crystalline spheres," "space and time are absolute," or "no particle can have a precise velocity and position simultaneously." But what if *none* of these theoretical claims are exactly true? Do we want to deny then that scientific methods are *ecologically* rational? I believe that the relevant concept of ecological validity is

better captured in terms of the concept of predictive accuracy (Forster and Sober, 1994). False theories can be predictively accurate. And evaluation methods that never favor true hypotheses may be more predictively accurate than those that do not (e.g., which is more ecologically adaptive—a stopped clock that is exactly right once every 12 hours, or a clock that is 3 minutes slow?).

In section 7, I also refer to some theorems in mathematical statistics that directly address the connection between fast and frugal evaluation criteria in science and the goal of predictive accuracy. The problem with standard Bayesian approaches to scientific rationality is that they do not address this question at the present time.

A unifying theme in this article is to agree with Gigerenzer et al. (in press) that boundedness and ecological validity must be considered together—and that ecological validity is an important feature of human decision making that is too often overlooked.

2.   Fast and Frugal Learning

Gigerenzer et al. (in press) talk about fast and frugal decision making in all walks of life. As a philosopher of science, I will argue that the same points apply to the evaluation of scientific hypotheses. Scientists judge the success of models on an intuitive basis, often guided by a quick informal assessment of the simplicity of the model and how well a model fits seen data, or by how well the model anticipates new phenomena. Such methods are relatively fast, and they are frugally applied.

For example, scientists appear to use a kind of "don't fix it if it ain't broke" heuristic, according to which a model is not abandoned unless it yields an obvious discrepancy either with data or when the model fails to fit into a network of models in a unified way (a discrepancy is what Kuhn (1970) calls an *anomaly*). If there is no discrepancy, then there is no problem to be solved, and therefore no change. For this reason, Kepler's laws of planetary motion are still widely used in contemporary astronomy even though they predate Newton. Idealizations are not abandoned simply because they are idealizations (Forster, in press a). An advantage of the "don't fix it if it ain't broke" heuristic is that it makes the learning process more bounded than it might otherwise have been. Scientists do not have to complicate their hypotheses just because the world is *really* more complicated than their current hypotheses assume.

The same frugality of learning is an important part of Staddon's (1983) discrepancy theory of learning in animals. Consider an experimental arrangement that was invented by W. K. Estes and B. F. Skinner in 1943 (Staddon 1983, p.137). The animal (usually a rat) is first trained to press a lever at a fairly steady rate. This is known as the *baseline* behavior, for the effect of subsequent conditioning is measured by changes in this behavior. Then, an intermittent stimulus, such as a buzzer, is introduced to the animal, and after the initial "novelty" effect wears off, the animal returns to the baseline behavior. For now the buzzer predicts nothing and

is ignored. The rats are then divided into two groups. The control group is subject to intermittent electric shocks irrespective of whether the buzzer is on or off, while the experimental group is subject to intermittent shocks only when the buzzer is on. In the control group the baseline behavior continues, while in the experimental group the animals stop pressing the lever while the buzzer is on. The buzzer sounds for variable lengths of time at unpredictable intervals, so the occurrence of the shocks can only be predicted from the sound of the buzzer. The suppression of the baseline behavior in the experimental group is, therefore, evidence that the animal is using the buzzer to predict the shocks. Now consider the phenomenon of *blocking* (Staddon 1983, p. 415). After the animal has been conditioned to suppress its lever pressing in response to the buzzer, we add a light stimulus concurrent with the buzzer. The rat will not learn to predict the shocks from the light stimulus despite the fact that the light stimulus is also a perfect predictor of the shocks (to test this claim, remove the buzzer).[1] Therefore, learning in animals is more bounded than the alternative, which would be to learn all salient correlations.

Fast and frugal methods of judging models in science are also ecologically adaptive. Whether idealizations are retained depends on whether they minimize anomalies, and this depends on how well nature "cooperates" in any particular context. Past predictive performance is a useful indicator of future performance only if nature is uniform in at least some respect. But does the converse of this claim hold?

If successful human decision making in science and in everyday situations is simple, fast, and frugal, then doesn't this bode well for the prospects of automating these tasks in a machine? Doesn't this show that machine intelligence is just around the corner? By now the academic public is leery of such claims, and rightfully so. But it is a less trivial, and more important question to understand why it is so difficult to implement fast and frugal heuristics in a machine. The quick answer is that decision rules are just the tip of a huge cognitive infrastructure needed to implement fast and frugal decision making. Witness the complexity of the human brain. It is a hugely complex apparatus whose job it is to prepare for rational action in real time. A decision rule may be written simply in a natural language, but to implement the rule one at least needs to know the language. Thus simple rules may be an essential feature of our intelligence (simple heuristics that make us smart), but before we can act on the rule, we go through decades of learning and education. For example, the recognition heuristic (Goldstein and Gigerenzer, in press) makes use of the simple fact of whether or not we recognize the name of a city, or the name of a company, to judge which cities are bigger or which stocks are the best investments. While such recognition-based decision making is simple at one level, it requires complex recognition machinery underneath and many years of accumulated experience to work.

Likewise, in science, there may be a simple way of evaluating competing models (e.g., other things being equal, choose the model that best fits the phenomena), but there are two major gaps between the statement of the rule and its automated

implementation in a machine. For one, there is the gap between the statement of the heuristic and its meaning in specific contexts. Which other factors must be equal, or, when they are not equal, which other factors need to be traded off against fit? Simplicity? But what is simplicity? How much weight should be give to it? Second, there is always a judicious choice of competing models to which the rule is applied. Scientists do not consider every possible rival hypothesis. Their choices are well educated by past experience and even intuition. It is not clear that any of this can be easily automated.

In any case, my conjecture is that the obstacles to our understanding of how science works are related to the difficulties in implementing fast and frugal heuristics in machines. Originally, I had planned to draw the lessons from the philosophy of science and then apply them to the problem of machine intelligence. In the end, I realized that the points are better known and more clearly explained in the context of machine intelligence than in the philosophy of science. So I have reversed the order of the discussion, first addressing the difficulties faced in AI and then raising the analogous issues in the philosophy of science. The bottom line is the same—that the issues in each case are similar in relevant respects.

3.   Dennett's Robot

The subject of machine intelligence is too vast to deal with in any generality. Instead, I will simply discuss two failures in AI and then draw some general lessons about these. The first example is a fictitious story told by Dennett (1984) about a robot that failed to solve a bomb retrieval problem. The second concerns a scientific discovery program called BACON (Langely et al., 1987), which its authors claim has rediscovered Kepler's harmonic law of planetary motion.

Dennett's story is about a group of engineers who are trying to design an intelligent robot. Their prototype robot, R1, is given the task of retrieving its spare battery from a room with a bomb inside it. The robot gets inside the room and see its battery on a wagon in the room. It decides to retrieve the battery by pulling the wagon out of the room. Just after it exits the room with the wagon, the bomb explodes. The trouble is that the bomb was also on the wagon, so the robot and the battery are blown to smithereens. The robot knew that the bomb was on the wagon, but failed to take account of the implication that if the bomb is on the wagon, and the wagon is pulled out of the room, then the bomb will not remain in the room. Back to the drawing board.

This time the engineers design the new robot-deducer, called R1D1, which deduces the consequences of pulling the wagon out of the room. So, the new robot enters the room to begin deducing the thousands of effects of its proposed course of action. The color of the walls will not change when the wagon is pulled out, the room will not move when the wagon is pulled out, and R1D1 has just finished deducing that the wheels of the wagon will revolve a greater number of times than

there are wheels on the wagon, when the bomb explodes. Back to the drawing board.

Now the engineers realize that the robot should ignore the consequences of a proposed action that are *irrelevant* to the outcome. The new prototype is the robot-relevant-deducer, or R2D1 for short. The robot approaches the room and stops. The engineers yell at the robot to do something, to which it replies that it has computed thousands of irrelevant consequences of its actions, and has duly ignored them, and is in the process of computing many more…when the bomb explodes. As Dennett puts it, R2D1 falls far short of the fabled perspicacity and real-time adroitness of R2D2 from Star Wars.

One of the lessons of Dennett's story concerns the problem of unbounded rationality. Each robot is endowed with the knowledge needed to solve the task. The problem is not one of information gathering. The problem is one of information *extraction* or retrieval. R1D1 faced this problem because it aimed to deduce all possible consequences—relevant and irrelevant—of its action without selective search or a stopping rule. R2D1 addressed the problem of relevance by simply adding in relevance as a utility, which does not work because the procedure is still unbounded.

Fast and frugal heuristics promise to solve this problem by providing a bounded search procedure. However, it is not exactly obvious how it would work in this case. If the robot were intelligent, in the way in which a bomb disposal officer is intelligent, then there would be a fast a frugal heuristic available—namely, the one we can express in English by the rule "retrieve the battery from the room…and leave the bomb in the room." The problem in the case of the robot is not so much the absence of a fast and frugal heuristic, but the absence of the cognitive substratum needed to implement the fast and frugal heuristic.

Of course, one could insist that fast and frugal heuristics could be useful for facing the possible contingencies. If the bomb is on the wagon, then pick up the battery and carry it out. If the bomb is attached to the battery, then…? Well, fast and frugal heuristics are not meant to be infallible. After all, this is a problem that even a human operative might not solve.

The problem with contingencies is that there are too many of them to consider individually. Yet human beings are capable of acting intelligently in situations they have never experienced or even considered. How is this possible? The answer is that we are able to *generalize* from situations we have considered or experienced by virtue of some kind of similarity metric and some form of generalization. This is an underlying cognitive *capacity* that human designers cannot replicate in a machine by explicitly considering a finite number of scenarios, no matter how many. In fact, the trouble is not just with the number of scenarios. The problem is that no scenario we consciously consider is complete in all relevant details. We are unaware of most of the information we take account of in most of our rudimentary decision-making.

It is for this reason that there has been a rising interest in machines that learn, for part of what they appear to learn is the similarity metric needed to generalize

from one case to another. Connectionist networks (Rumelhart et al., 1986) make up one important class of such learning machines. However, it is becoming increasingly clear that, for learning machines to generalize *well*, learning must be well structured. Either the right neural architecture is in place prior to learning (by a designer or by evolution, or by some selection process) or the data are fed to the network in successive stages, or some other organizational mechanism is in place (in machine learning, these are called *biases*).

Fast and frugal heuristics do generalize well (Gigerenzer et al., in press). However, my point is that the *implementation* of such heuristics by human beings is in most cases only possible because of the presence of an immensely complex substratum of cognitive capabilities, which have been set in place by eons of evolution and years of learning. I believe that an analogous point can be made about scientific discovery.

4.   Automated Scientific Discovery

Despite the title of this section, I am going to limit myself to one example, namely, the well-picked-on example of BACON (Langely et al., 1987).[2]   BACON is a computer program designed to discover simple laws such as Kepler's harmonic law of planetary motion. Discovery is a kind of learning, so the recent interest in machine learning encompasses projects of this kind. The question is whether it shows more promise that connectionist learning. Certainly, if a machine can learn something like Kepler's law, then it has achieved an impressive degree of generalizability, for Kepler's law remains at the heart of astronomy even today. The question is whether BACON has modeled the process of discovery in its entirety, or merely a small, and perhaps trivial, part of it. I will argue that the most important part of the discovery process is the development of the underlying conceptual framework. This is the part that BACON is not designed to solve.

Kepler's harmonic law says that the mean radius of planetary orbits around the sun (R) cubed divided by the time of revolution around the sun (T) squared is equal to a constant. That is, this ratio is the same for all planets. This is Kepler's third law of planetary motion, and it is largely independent of the other two. The way that BACON discovers Kepler's law is that it tests various possible ratios, from the simplest to the more complex, and stops when it finds that the ratio is sufficiently constant. It copes with noisy data, and it is an impressive programming achievement. Yet it took 1,500 years to discover Kepler's law from the time of Ptolemy's quite sophisticated geocentric theory of planetary motion. How can it be that a computer program can complete in seconds what took intelligent human beings 1,500 years to achieve?

The answer is that human beings were not handed measurements of R and T on a platter. What we observe when we observe the planets are not the motions of the planets relative to the sun, but their apparent motions relative to the earth. R and T pertain to the motion of the planets relative to the sun. The problem of computing

these quantities from the apparent motions of the planets is an immensely complex one, which was eventually solved by Copernicus. To give you some idea of the problem, suppose there is a series of concentric merry-go-rounds revolving around a common center at different rates. You are on the third one from the middle. It is very dark and only one seat on each merry-go-round is illuminated with a very dim bulb. From the motion of these lights relative to a very distant background, you have to work out the three-dimensional motions of everything you can see. Good luck!

If there were only one light, this problem would be unsolvable in principle because it is impossible to resolve a single resultant motion uniquely into two vector components. Copernicus was only able to solve the problem on the conjecture that there was one vector component (namely your motion relative to the large light at the center) common to all the resultant motions. This was the difficult part of the problem. The discovery of Kepler's law came soon after.

There are two points to this analogy: (1) in scientific discovery you are not told when there is something there to be discovered, and (2) the discovery may require a conceptual innovation before the data can be presented in the form on which a computer program such as BACON can operate. Copernicus's innovations built up the cognitive substratum on which the discovery of Kepler's laws depended. Kepler's laws, in turn, laid the conceptual groundwork for Newton's discovery of universal gravitation, and so on.

What feature does this example have in common with Dennett's robot? Only this: Both examples raise the problem of information *extraction*. In the case of the robot, it already has the information but cannot retrieve what is *relevant* to the problem. In the case of BACON, it is just given what is relevant, and nothing more. A truly intelligent program should be able to extract this information for itself. I am waiting for the program called COPERNICUS.

5. No-Free-Lunch Theorems

So far in this article, I have argued that simple instructions given to a robot, or simple methods of scientific discovery, must be supported by an underlying complexity, whose function is to extract salient information from observational data and from stored data, and whose existence is easily overlooked exactly because its invisibility is a part of its function.

Another way of making the point is this: If learning is viewed *in its entirety*, then there is no simple universal method of learning. The apparent counterexamples (like error backpropagation in connectionist networks, or statistical testing in science) owe their existence to the fact that in different contexts there are different structures in place that allow for the methods to succeed. In machine learning, the slogan is that every machine has a built-in bias, which allows it to succeed in one kind of learning problem at the expense of failing in others. The idea is that there are no free lunches in learning—if learning succeeds, then it is because

of design features built into the machine. The so-called no-free-lunch theorems (Wolpert 1992, 1995, 1996) provide support for this conclusion. But they will also deepen our understanding of bounded rationality and ecological validity.

Here is a very easy example of a no-free-lunch theorem. Consider an imaginary universe that lasts for exactly 2 days, where on each day there exists exactly one object, which is either a circle or a square. There are exactly four possible histories that this world may have: (circle, circle), (circle, square), (square, circle) and (square, square). A predictive hypothesis tells us what to predict on the second day given what is observed on the first day. In this world, there are four predictive hypotheses: *Same* = "same on both days," *Diff* = "different on both days," *Circle* = "circle on second day no matter what," and *Square* = "square on second day no matter what." Given that the four histories have the same prior probability (¼), what is the probability that each predictive hypothesis will make a correct prediction on the second day? If you work it out, you will find that the probability is ½ for all four predictive hypotheses.

Let us now complicate the imaginary universe just a tad. Instead of just one dichotomous variable (the shape variable with two values—circle and square), there is also a color variable with two values (red or green) and a size variable (small or large). That means that on any particular day, an object is in one of 8 possible states. Now suppose that there are two objects instead of one, so that the universe is in 64 possible states on any particular day. Further suppose that there are 10 days in the history of our universe, and we are interested in predicting the state of the universe on the 11th day. There are now 64^{10} possible histories in our universe. This is already an astronomically high number—approximately equal to 1.15×10^{18}. A predictive hypothesis is a function that maps all of these possible histories to one of the 64 possible universe states on the 11th day. There are approximately 10 to the power of 6.38×10^{19} different functions, and therefore the same number of different predictive hypotheses! The complexity of the hypothesis space is mind-boggling even in this extremely simple universe.

The no-free-lunch theorems point out that if we give equal prior probability to each possible universe (a true "ignorance" prior), then every predictive hypothesis has exactly the same prior probability of success. So far, it may seem that the conclusion may not seem surprising—if we are completely ignorant of which possible world we are in, then we are completely ignorant of the future. However, the problem runs deeper. For suppose that we know the history of the universe during the first 10 days. If we add this information, then we are still completely ignorant of the future. Our newfound knowledge falsifies all but 64 predictive hypotheses, but these 64 hypotheses are all equally probable and they all predict a different state on the 11th day. We are still completely ignorant of the future.

The importance of the no-free-lunch theorems is that they show what epistemology is not. One cannot begin from a blank slate (with nothing more than a knowledge of the space of possible universes). One must begin with an a priori prejudice (a "bias" as the machine learning theorists call it) before the knowledge-

seeking process can get off the ground. In that sense, there is no free lunch, which is where the theorems get their name.

Therefore the computational complexity of our toy universe is really only one side of the epistemological coin. For even if the computational problem is solved, there can be no way of differentiating among predictive hypotheses unless we *begin* by discriminating among them. Only if some possible universes have greater prior weights than others can there be any discrimination among predictive hypotheses.

For example, one way of getting the epistemological process off the ground would be to adopt the Markov principle, which says that the prediction of the state of the universe on day $n + 1$ depends only on the state of the universe on day n. If we restrict the class of predictive hypotheses to those that treat the first 9 days as irrelevant to what happens on the 11th day, then one cuts down the space of possible predictive hypotheses. But we gain much more than that if we adopt a further assumption (known as stationarity), which says that the *same* hypothesis that predicts day 2 from the state on day 1, day 3 from the state on day 2, and so on, also predicts the state on day 11 from the state on day 10. With the Markov condition and stationarity, we may now use past data to further winnow down the class of plausible hypotheses. It takes some broad principles, such as the Markov condition and the stationarity assumption, to get the epistemological ball rolling.

This point is so important that I think that the history of science is full of instances in which quantities have been *invented* so that the space of possible hypotheses may be restricted via assumptions like the Markov principle. Examples include the introduction of instantaneous velocities in mechanics, which ensure that an exhaustive description of the state of a system at an instance of time will determine future states, the postulation of genes to explain the statistics of heredity (Arntzenius, 1995), and intervening variables (brain states) to explain learning data (Sober, 1998). Conceptual innovations of these kinds are motivated by their empowerment of the same broad principles, which are in turn motivated by the need to constrain the class of predictive hypotheses.

On the other hand, there are many possible conceptions that we should not introduce. Goodman (1965) invented the concept of *grue*, which I define as follows: Object *x* is grue at time *t* if and only if *x* is green at time *t* and *t* < 2100 AD, or *x* is blue at time *t* and *t* ≥ 2100 AD. Moreover, suppose we define *bleen* as: Object *x* is bleen at time *t* if and only if *x* is blue at time *t* and *t* < 2100, or *x* is green at time *t* and *t* ≥ 2100. Then we may replace the words "green" and "blue" with the words "grue" and "bleen." Let use refer to these as *grolor* terms, as opposed to color terms. Then predictive hypotheses that are restricted by the principle of time invariance (stationarity) will be replaced by grolor versions in the new language. That is, stationarity will imply grolor invariance rather than color invariance. Thus, the standard hypotheses are now eliminated from consideration. The lesson of this (annoying) example is quite real. A general principle such as stationarity only has a precise meaning relative to a fixed vocabulary. If the vocabulary is not restricted in any way, then the principle will not restrict the space of hypotheses.

The argument is leading to the following conclusion. The ecological validity of our rules for selecting amongst competing hypotheses depends on the fact that there are some prior restrictions on the space of hypotheses to which these rules are applied. There is no such thing as the ecological validity of these rules in and of themselves. This conclusion applies to all decision rules.

The no-free-lunch theorems and Goodman's grue problem show that there is no way of *proving* the ecological validity of selection rules from a priori principles. Nevertheless, one should not therefore ignore the question of ecological validity. For example, it is a worthwhile exercise to entertain and test conjectures about the role of evolutionary "learning" in molding our perceptual systems and the vocabulary of ordinary language in restricting the space of hypotheses that we normally entertain (as illustrated by the fact that stationarity would change its meaning if we were to change from color terms to grolor terms). This process is not viciously circular, for we can easily discover that the ecological validity of a particular rule is inconsistent with what our best scientific theories say about the uniformity of nature. For example, a rule that depends on the simple addition of velocities (Galilean transformations) for its validity will not extend to velocities near the speed of light. We know that from Einstein's theory of relativity. But it will be valid within the domain of everyday experience.

6.   Bayesian Decision Making

Bayesian decision theory is often described as the best decision theory we have. Its basic tenet (at least the version I shall consider) is that a rational decision is one that maximizes an agent's expected utility. "Utility" is a quantitative measure of the value of the outcome of an action to the agent—the payoff for the agent, in other words. The goal of the decision is to maximize the payoff, but in the face of uncertainty about what the payoff will be, what should a decision maker do? The answer is that if the agent can assign probabilities to the possible outcomes of an action, and knows the utility of each outcome, then the agent should weigh the utilities by their probabilities. The gives the *expected* utility of each possible action. The decision is now made according to which possible action will lead to the greatest expected utility.

These probabilities of each possible outcome can be computationally expensive to determine because they are generally calculated as the average of the probability of an outcome in an infinite number of possible environments, where the terms in the average are weighted by the probability of each environment. Therefore, Bayesian decision making is also generally computationally expensive.

However, to be charitable to Bayesianism, it should be recognized that Bayesians are free to make a distinction analogous to the distinction between act utilitarianism and rule utilitarianism in moral philosophy (Urmson 1953).[3] *Act utilitarianism* entails the view that a particular action is justified as being right by showing that it promotes the ultimate end (happiness). On the other hand, *rule utilitarianism*

views a particular action as being right by showing that it is in accord with some moral rule, such as "Keep promises," "Do not murder," or "Tell no lies." A moral rule is shown to be correct by showing that the recognition of that rule promotes the ultimate end (Urmson 1953, p. 35). The moral rules are the fast and frugal guides to a moral life.

Rule Bayesianism would be a version of Bayesianism that says to act rationally is to act in accordance with a set of simple rules, and a set of rules is shown to be correct by showing that it maximizes utility better than an alternative set of rules. The justification of the rules is still computationally expensive, but it is done "offline" and therefore allows that decision making in real time is fast and frugal. So, if fast and frugal heuristics could be analyzed in Bayesian terms, then we might understand why they work as well as they do. It is natural for a Bayesian to argue that unbounded rationality can explain the success of bounded rationality. However, it is far from clear that this promise is fulfilled.

The main purpose of this section is to understand why Bayesian justifications of fast and frugal heuristics would not help establish their ecological validity, at least not without further argument. Some Bayesians treat priors as subjective, and some treat them as objective. For subjective Bayesians, the problem in a nutshell is that different choices of prior probabilities lead to a choice of different sets of rules. In fact, with enough ingenuity, a subjective Bayesian can work backward to a prior that yields any set of rules as a utility maximizer. To illustrate the idea, consider someone who values money but who rejects the rule that he should use his automatic deduction to reduce his taxes because of an unfounded and irrational belief that a use would lead to the mysterious disappearance of all of his savings, leaving him with less money. It is easy to say such a belief, and the associated probability distribution, is irrationally held. While that is intuitively correct, the point is that expected utility maximizers are not automatically rational in this intuitive sense.

If the Bayesian method can lead to different decision rules when it adopts different priors, then it is impossible for all of these derived rules to be ecologically valid. How do we decide which rules are ecologically valid and which are not? We cannot trust a rule unless it is accompanied by an argument showing that the priors are capturing some uniformity of the environment. That is where the burden of a Bayesians' proof must lie.

Yet, Bayesians may be puzzled about how to address the question. How does one determine the fit *with reality* of a prior probability? At some stage, Bayesians have to begin with a 'first' prior, which is necessarily independent of any evidence (in the sense that it is not updated in light of any evidence). Yet any such Bayesian prior will distribute its weight over a vast range of *possible* environments, where at most one is the actual environment. So, how can we measure the ecological validity of a Bayesian prior except by the degree to which it gives weight to the actual environment to the exclusion of all other possible environments?

The resolution of this puzzle serves to clarify the meaning of 'ecological validity'. It cannot require of a Bayesian that her prior gives zero weight to every

possible environment except the real one. Rather it requires that the prior narrows down the set of all possible environments to those that share some relevant feature of the actual environment. The prior must *truly* capture some regularity of nature or some aspect of reality. While this is a weaker requirement, the adjective 'truly' is critical. For if the prior is merely engineered in order to reproduce a Bayesian equivalent to fast and frugal heuristics that are *thought* to be ecologically valid, then it is clear that such backwards engineering cannot provide any justification of their validity. Rather, there would have to be some independent justification for thinking that the prior truly captured a relevant regularity of the environment.

Here is one way *not* to solve the problem. Measure the ecological validity of an action by its expected utility, then optimize the ecological validity of our decisions by maximizing expected utility. The flaw in the argument arises from a subtle conflation of utility and expected utility. It is the maximization of the utility that defines ecological validity. Or more exactly, ecological validity must be defined in terms of the gain in utility over many applications of the rule. However, this is an objective kind of average, which has no direct connection with the *subjective* probabilities that define the expected utilities as they are most commonly understood by Bayesians.

Therefore, any Bayesian who claims to address the question of ecological validity must appeal to the objectivity of priors in some sense. These objective priors are usually justified by Laplace's principle of indifference, which says that all possibilities are given equal weight. Yet, as the no-free-lunch theorems show, a rigorous application of this idea leads to another problem—namely, that no hypothesis is favored probabilistically over any other. Objectivity, in the sense of being equally ignorant of all possibilities, does not achieve ecological validity in any world. So, this becomes the second way *not* to solve the problem.

This is not to say that an "ignorance" prior can never be the right prior for a Bayesian to adopt. The correct conclusion is that the so-called "ignorance" priors never truly express a total ignorance over the hypothesis space. Perhaps the easiest demonstration of this fact is to note that a nonlinear transformation of the variables will transform a uniform prior to an equivalent prior that is *not* uniform (Forster, 1995). That is, a uniform prior changes its meaning when different variables are used (just as in the previous section, where the meaning of stationarity changes in different languages). Therefore, the use of uniform priors does not establish ecological validity of Bayesian decision making, at least not by itself.

Independent support for the same conclusion derives from Jaynes's (1971) discussion of Bertrand's (1907) paradox: Imagine a simple experiment of randomly tossing a straw onto a floor on which we have drawn a circle (see Figure 1). If the straw intersects the circle (in two places), what is the chance that the length of the chord formed by the straw is longer than the side of an equilateral triangle inscribed inside the circle? Suppose, without loss of generality, that the straw lands horizontally on the upper half of the circle. Then, the straw intersects the line OA
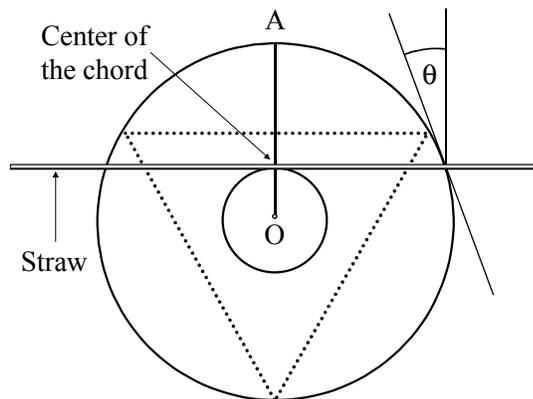
Figure 1. In Bertrand's example (see text), there are three different ways of applying Laplace's principle of indifference to answer the question: What is the chance that the straw intersects the equilateral triangle (as shown)? The first distribution gives equal weights to all intersection points on the line OA, the second distribution gives equal probability the angle of intersection of the straw on the circumference (angle θ), and the third gives equal probability to each possible area of the inner circle. Each leads to a different answer to the original question (1/2, 1/3, and 1/4, respectively).

at some point, which is the center of the chord. The question is: What is the chance that the center of the chord is below the horizontal dotted line in Figure 1?

The purpose of Bertrand's example is to show that Laplace's principle of indifference does not (by itself) solve this problem because there are many ways of defining equally possible situations.   Three of these are assigning a uniform probability distribution to: (a) the distance between the center of the chord and the center of the circle; (b) the angle of intersection of the chord on the circumference (angle θ in Figure 1); or (c) the area of the interior circle touching the center of the chord (the area of the inner circle in Figure 1). These assignments respectively lead to answers of 1/2, 1/3, and 1/4 to Bertrand's question. Not all of these answers are ecologically valid (Jaynes cites empirical data to support answer (a)).   In this particular example, Jaynes (1971) shows that if the probability distribution of events is invariant under translation and rotation (so that the same answer holds when one moves the circle) then (a) is the unique answer. As I read it, Jaynes has shown that the justification for ecological validity of answer (a) does not derive from the uniformity of the prior at all (since all three answers have that in common), but from an invariance with respect to situations that are relevantly similar according to the laws of nature.

Therefore, the correct issue concerning the objectivity of priors is whether the prior restricts the hypothesis space in the right way. The "right way" may depend not only on whether it enforces a general principle but also on the language to which it is applied. This not only goes beyond current Bayesian practice, but it also suggests that the ecological validity of decision rules in different environments will rest on different regularities of nature. This is exactly the issue of specificity versus

generality, which Gigerenzer and Todd (in press, p. 18) claim is at the heart of the success of fast and frugal heuristics:

> The main reason for their success is that [these heuristics] make a trade-off on another dimension: that of generality versus specificity. While coherence criteria are very general—logical consistency, for instance, can be applied to any domain—the correspondence criteria that measure a heuristic's performance against the real world require much more domain-specific solutions.

On the one hand, it is clear that some degree of specificity is required if rules are tailored to a particular context. On the other hand, the general regularities of nature have to be exploited if predictive accuracy is to be achieved. In my cursory discussion of some examples in the previous section, I have argued for the importance of general principles, such as the Markov condition, stationarity, and spatial symmetry. The key is to exploit these generalities and to utilize information pertaining to a particular environment at the same time. These two elements must work in concert, and this is clearly a highly complicated task even if the implementation of the decision rule is very simple at the surface.

How can we know which are the right general principles of nature to apply? Do they *always* apply, in a domain-general way? Otherwise, how can we tell when different particular principles apply? Clearly I do not have general answers to these questions. Yet, these are problems that science grapples with every day. Perhaps it is therefore instructive to ask how science establishes its ecological validity.

7. The Ecological Validity of Science

A theme of this article has been the analogy between how science works and the operation of other information extractors in nature—something that has long been suggested by the common sense analogy between seeing and understanding. The seminal work of Marr (1982) on the mechanisms of human vision is a fine study of how a computational system can exploit general environmental regularities. Our visual system is not a universal seeing machine that would work in any possible world. It is one that exploits the contingencies in our environment by using edges as indicators of object boundaries, texture gradients as indicators of depth, and looming as an indicator of velocity. It draws connections between visual cues and real-world structures based on contingent regularities such as the distance between our eyes, the distortion of our lenses (or lack thereof, as the case may be), rigidity of bodies in motion, or the constancy of colors, shapes, and size. Marr's is a story about the ecological validity of human vision.

Are there are any new lessons to be learned about the concept of ecological validity from the philosophy of science? For one thing, we may learn that the idea of ecological validity is deceptively vague. How is "fit to reality" defined, precisely? For example, in the philosophy of science, there is a dispute between instrumentalists and realists. Instrumentalists view scientific hypotheses as *merely* instruments for the control and prediction of nature. I say "merely" because realists

also agree that theories are instruments for prediction. However, realists think that science aspires to the more lofty goal of truth, which includes not only the truth of prediction, but also the truth of theoretical statements about the reality behind the observable phenomena.

Even within the common core of these schools of thought, there is a disagreement about how to define the "fit to reality" with respect to observational phenomena. On the one hand, the standard approach is to say that a hypothesis is *empirically adequate* if and only if all its observable consequences (past, present, and future) are true (van Fraassen, 1980). On the other hand, Forster and Sober (1994) argue that predictive fit (predictive accuracy) is more interestingly defined in terms of statistical measures of fit, such as the log-likelihood or the sum of squared residuals, applied to hypothetical situations in which predictions are made.

If ecological rationality is meant to be a precise notion, then the variety of ways in which "fit to reality" is defined might be an embarrassment. However, I think that any of these kinds of fit would form a valid explication of the idea. The really interesting question is: Which kind of "fit to reality" is *actually* achieved by decision rules in science? The argument against theoretical truth or empirical adequacy as goals of science is not that we would not want them if we could get them. The argument against them is that they are very rarely achieved. Thus, the reasons why predictive accuracy is a more interesting alternative to empirical adequacy is that (1) in cases where empirical adequacy is achievable the achievement is equally well described in terms of predictive accuracy, and (2) there are many situations in which empirical adequacy is not achieved (it is a black-and-white notion), but at least some *degree* of predictive accuracy is achieved.

The dominant philosophy of science in North America is a Bayesian philosophy of science. A general Bayesian decision-theoretic philosophy of science defines a rational decision to be the prediction that has the highest expected utility, where the utilities would measure any kind of payoff. Of course, Bayesians are not forced to consider all possible utilities. In fact, the most common Bayesian view of science is one in which the utility of accepting a hypothesis is a function of only the truth or falsity of the hypothesis (Earman, 1992). Suppose the payoff for accepting a true hypothesis is 1, and the payoff for accepting a false hypothesis is 0. I will refer to this as the *basic truth utility*, and the corresponding Bayesian acceptance as the *basic Bayesian model*. With the basic truth utility, it is an easy exercise to show that the expected utility of a hypothesis reduces to its probability. More specifically, the expected utility of accepting a hypothesis at any time is the probability of the hypothesis given the total evidence accepted at that time. Therefore, if there is a decision to be made among rival hypotheses, the basic Bayesian rule is to accept the most probable hypothesis. Not only is this a form of act Bayesianism (section 6), but it also defines "fit with reality" in the simplest terms.

When scientists choose between rival hypotheses, the payoffs need not be as simple as I have described them. There are two distinct ways in which Bayesians might complicate their basic model. (1) They could *replace* basic truth payoffs

with another more complex quantity, such as predictive accuracy. (2) They could add utilities, like informativeness, or simplicity, to the basic truth utility. The latter approach is the standard one pursued in the philosophy of science literature (e.g., Levi, 1973, 1984; Maher, 1993). In any such extension of basic Bayesianism, the probability that a hypotheses is true still plays a role in defining expected utilities. Let me refer to the basic Bayesian model and its extensions as *standard Bayesianism*.

How does the issue of bounded rationality bear on standard Bayesian philosophy of science? First, the computation of probabilities can be very expensive. Remember that these probabilities are calculated by Bayes's theorem, which at first seems very simple: The posterior probability of *H* is proportional to the prior probability of *H* times the likelihood of *H*.

However, appearances are sometimes deceptive. When scientists make acceptance decisions, it is often on the level of scientific models, which are highly disjunctive hypotheses. A model might specify the *form* of a function to be fitted to data without specifying the values of adjustable parameters that appear in the equations. To calculate the likelihood of a model *H*, $\Pr(E|H)$, one must average the likelihoods of all the members of *H* corresponding to every possible set of parameter values. Such integrations are computationally expensive in most cases.

Is Bayesianism the only approach paying this price? If the cost arises from the calculation of likelihoods, then it may appear that any approach that uses likelihoods is subject to the same costs. However, this initial impression is incorrect, because other model selection approaches use the maximum likelihood. For example, Akaike's Information Criterion (AIC) is a rule of selecting from amongst competing models (Akaike, 1973, 1974) by trading off the *maximum* log-likelihood of *H* against the number of adjustable parameters. A maximum likelihood is much easier and more frugal to calculate than an average likelihood, and the number of adjustable parameters is easy to determine. Other methods (for a survey, see Forster, in press c) may simply look at how well a model fitted to one data set succeeds in predicting a second set of data (this is a kind of cross-validation method). Again, this quantity is easily computed.

Moreover, the Akaike framework (Forster and Sober, 1994; Forster, in press a) provides a dramatic example of how the issue of ecological validity can be addressed, albeit in a limited way. Akaike (1973, 1974) derives his model selection rule as a way of maximizing predictive accuracy (Forster and Sober, 1994), where predictive accuracy is defined as the expected fit of a model with respect to unseen data of a certain kind. Here the expectation is an objective one defines in terms of the probability distribution that generates the data. It is treated as an unknown, although the theorems he proves do depend on some properties of the distributions. In particular, it is assumed that past and future data are generated by the same distribution, and that repeated estimations of adjustable parameters conform to the normal distribution in the way ensured for asymptotically large sample sizes by the central limit theorem. Under these relatively weak conditions, Akaike proves

that a formula involving fit (maximum likelihood) and simplicity (the paucity of adjustable parameters—although see Forster, 1999, for some needed qualifications) yields an estimate of predictive accuracy. So, Akaike's theorem is a bridge drawn between the decision rule (the AIC rule) and the goal of the decision rule (predictive accuracy). That is, the theorem establishes a kind of "fit to reality," which explains why the rule actually tends to succeed in maximizing predictive accuracy.

While the AIC rule is itself fast and frugal, and therefore simple in a pragmatic sense, it also says that simpler *hypotheses* tend to have greater ecological validity (in the sense of predictive accuracy) than they would otherwise have. The two kinds of simplicity are different. At the level of the decision rule, the simplicity is a pragmatic kind of simplicity, as the terms "fast" and "frugal" imply. At the object level, the rule favors simpler hypotheses. While both kinds of simplicity provide obvious pragmatic advantages, being more economical to use and easier to understand, the Akaike criterion is primarily concerned with ecological validity.

Ockham's razor is a famous principle that practicing scientists have appealed to for centuries as an informal, and imprecise, appeal to simplicity. Newton based his argument for gravitation on a similar principle (Newton, 1687/1934, pp. 398-400; 546-547): "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. Therefore, to the same natural effects we must, as far as possible, assign the same causes." The AIC rule gives a precise numerical tradeoff between fit and simplicity when they are measured in terms of likelihood and paucity of parameters, whereas Newton's criterion is vague and imprecise. Can Akaike's theorem therefore explain the ecological validity of Newton's vague and imprecise appeal to simplicity? In cases such as Newton's theory of gravitation, which improved fit and increased simplicity at the same time, *almost any criterion* will favor Newton's theory over its rivals. Therefore, the justification of Newton's argument for gravitation does not rest on the claim that the AIC rule would have favored Newton's hypothesis. Rather, the explanation is that Akaike's *framework* can provide an argument that the evidence that Newton cites, including the paucity of postulated causes, overwhelmingly supports his theory's greater "fit with reality"—at least for one way of making that claim precise.

The ecological validity of fast and frugal appeals to simplicity as a criterion of theory choice does not always rest on Akaike's theorem. In fact, extensions to Akaike's theorem have already been used to investigate the exceptions to the AIC rule. Failures of the AIC rule can arise not only when the assumptions of Akaike's theorem are systematically violated (Kieseppä, 1997), but also under other conditions. The particular cases are rather difficult to understand, so I can only provide a quick summary: For one, there is the subfamily problem (Forster and Sober, 1994), which points out that AIC can fail to maximize predictive accuracy if models are constructed post hoc so as to maximize fit and simplicity at the same time. Then there is Kukla (1995), who raises a similar problem (also see my reply in Forster, 1995a). And there is the problem of "selection bias" if the set of models is very

large, whereby it will be probable that some models will accidentally receive an unduly high score (Zucchini, 2000). Finally, there is an issue concerning how AIC addresses Goodman's grue problem (De Vito, 1997; Forster, 1999).

What all of these problems suggest to me is that the ecological validity of any fast and frugal decision rule, like AIC, depends on restricting the choice of rival hypotheses to which the rule is applied. Moreover, when the set of rival models is restricted, then the restriction must be "judicious." This is exactly the issue that I raised earlier for Bayesianism, for a "judicious" choice of prior is exactly one that restricts the range of rival hypotheses in the "right" way. The difference is that Akaike's framework addresses this issue directly and provides some answers.

While I have no general proof that similar issues arise for any fast and frugal heuristic, it seems likely to me that their ecological validity will prove to be far from simple. My suspicion is that the right preselection of actions, on which a fast and frugal decision rule operates, is critical to their reliable operation—and, therefore, to their ecological validity.

8.   Main Conclusion

Gigerenzer and Todd (in press) state that their central goal is "to understand human behavior and cognition as it is adapted to specific environments (ecological and social), and to discover the heuristics that guide adaptive behavior" (p. 25). But to *understand* fast and frugal heuristics is to understand why they are ecologically adaptive, which may include not only how they evolved or were learned, but also why they are ecologically well adapted.

Perhaps there is a law of conservation of complexity that says that simplicity achieved at the higher level is at the expense of complexity at a lower level. Such a law appears to be consistent with the examples I have examined in this article. Accordingly, their "apparently paradoxical thesis: *Higher order cognitive mechanisms can often be modeled by simpler algorithms than can lower order mechanisms*" (Gigerenzer & Todd, in press, p. 31) seems to me exactly as one would expect.

The notion of ecological rationality is particularly important because it builds a bridge between the psychological sciences, which are most often unconcerned with truth or predictive accuracy, and philosophers of science, who are primarily interested in science as a knowledge-seeking process, rather than the psychology of practicing scientists (Forster, in press b). I have attempted to show that, contrary to popular belief, many of the same issues arise in different fields such as AI, machine learning, psychology, and the philosophy of science. It is therefore useful to consider them all at once, in the hope that we may jointly contribute to a unified science of epistemology.

Notes

[1] We know that the animals are *capable* of learning to associate the light and the shocks. For if we were to modify the experiment so that the connection between the buzzer and the shocks is randomized when the light is introduced, then a discrepancy will appear, and the animals will succeed in learning to predict the shocks from the light stimulus. In fact, if the light stimulus is paired with the shocks in the new bout of conditioning while the buzzer is *off*, then the rate of learning is accelerated—a phenomenon known as superconditioning.

[2] Another interesting but more involved example of an AI discovery program is TETRAD, developed by Spirtes et al. (1993) to infer causal conclusions from statistical data. It has generated considerable controversy. See Humphreys and Freedman (1996) and replies from Korb and Wallace (1997) and Spirtes et al. (1997). I will not discuss it here, although I believe that some of the broader issues I raise in this article apply to that example.

[3] I owe the point, and the reference, to Branden Fitelson.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle. In Petro, B. N. & Csaki, F. (eds.), *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado, 267-281.

Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* vol. AC-19, 716-23.

Arntzenius, F. (1995), A heuristic for conceptual change. *Philosophy of Science, 62,* 357-369.

Bertrand, J. (1907), *Calcul des Probabilites*. Paris: Hermann et Fils.

Dennett, D. (1984), Cognitive wheels: The frame problem of AI. In Hookway, C. (ed) *Minds, machines and evolution*. Cambridge: Cambridge University Press, 129-151.

De Vito, S. (1997), A gruesome problem for the curve fitting solution. *British Journal for the Philosophy of Science, 48***,** 391- 396.

Earman, J. (1992), *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge: The MIT Press.

Forster, M. R. (1995a), The golfer's dilemma: a reply to Kukla on curve-fitting. *British Journal for the Philosophy of Science, 46***,** 348-360.

Forster, M. R. (1995b), Bayes and bust: The problem of simplicity for a probabilist's approach to confirmation. *British Journal for the Philosophy of Science, 46***,** 399-424.

Forster, M. R. (1999), Model selection in science: The problem of language variance. *British Journal for the Philosophy of Science*, *50,* 83-102.

Forster, M. R. (in press a), The new science of simplicity. In Keuzenkamp, H. A., McAleer, M., & Zellner, A. (eds.) *Simplicity, inference and econometric modelling*. Cambridge: Cambridge University Press.

Forster, M. R. (in press b), The problem of idealization and other hard problems in the philosophy of science. In Nola, R. & Sankey, H. (eds) *After Popper, Kuhn, and Feyerabend: Issues in theories of scientific method*. Australasian Studies in History and Philosophy of Science series, Kluwer.

Forster, M. R. (in press c), Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology*.

Forster, M. R. & Sober, E. (1994), How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions. *British Journal for the Philosophy of Science, 45***,** 1 - 35.

Gigerenzer, G. & Todd, P.M. (in press), Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P.M. Todd, and the ABC Research Group, *Simple heuristics that make us smart*. New York: Oxford University Press.

Gigerenzer, G., Todd, P. M. and the ABC Research Group (in press), *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, D. G. & Gigerenzer, G. (in press), The recognition heuristic: how ignorance makes us smart. In G. Gigerenzer, P.M. Todd, and the ABC Research Group, *Simple heuristics that make us smart*. New York: Oxford University Press.

Goodman, N. (1965), *Fact, fiction and forecast,* Second Edition. Cambridge, Mass.: Harvard University Press.

Humphreys, P. & Freedman, D. (1996), The grand leap. *British Journal for the Philosophy of Science, 47*, 113-123.

Jaynes, E. T. (1971), The well-posed problem. In Godambe, V. P. & Sprott, D. A. (eds.), *Foundations of Statistics*. Toronto: Holt, Rinehart & Winston, 342 - 354.

Kieseppä, I. A. (1997), Akaike information criterion, curve-fitting, and the philosophical problem of simplicity. *British Journal for the Philosophy of Science, 48***, 21-48.

Korb, K. B. & Wallace, C. S. (1997), In search of the philosopher's stone: Remarks on Humphreys and Freedman's critique of causal discovery. *British Journal for the Philosophy of Science, 48*, 543-553.

Kuhn, T. (1970), *The structure of scientific revolution* (2nd ed.). Chicago: University of Chicago Press.

Kukla, A. (1995), Forster and Sober on the curve-fitting problem. *British Journal for the Philosophy of Science, 46*, 248-259.

Langely, P, Simon, H. A., Bradshaw, G. L. & Zytkow, J. M. (1987), *Scientific discovery: computational explorations of the creative process*. Cambridge, Mass.: The MIT Press.

Levi, I. (1973), *Gambling with truth; an essay on induction and the aims of science*. Cambridge, Mass.: The MIT Press .

Levi, I. (1984), *Decisions and revisions*. Cambridge: Cambridge University Press.

Maher, P. (1993), *Betting on theories*. Cambridge: Cambridge University Press.

Marr, D. (1982), *Vision.* New York: W. H. Freeman & Co.

Newton, I. (1687), *Principia Mathematica*. Motte's translation as revised by F. Cajori (1934), Berkeley: University of California Press.

Rumelhart, D. E., McClelland, J., and the PDP Research Group (1986), *Parallel distributed processing, Volumes 1 and 2*. Cambridge, Mass: The MIT Press.

Sober, E. (1998), Black box inference: when should intervening variables be postulated? *British Journal for the Philosophy of Science, 49*, 469 - 498.

Spirtes, P., Glymour, C. & Scheines, R. (1993), *Causation, prediction and search*. New York: Springer.

Spirtes, P., Glymour, C. & Scheines, R. (1997*)*, Reply to Humphreys and Freedman's review of *Causation, prediction, and search*. *British Journal for the Philosophy of Science, 48*, 555-568.

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

van Fraassen, B. (1980), *The scientific image.* Oxford: Oxford University Press.

Urmson, J. O. (1953), The interpretation of the moral philosophy of J. S. Mill. *Philosophical Quarterly*, *3*, 33-39.

Wolpert, D. H. (1992): On the connection between in-sample testing and generalization error. *Complex Systems 6***, 47-94.

Wolpert, D. H. (1995), The relationship between PAC, the statistical physics framework, the Bayesian framework, and the VC framework. In Wolpert, D. H. (ed.) (1995), *The mathematics of generalization*. Reading, MA: Addison-Wesley, 117-214.

Wolpert, D. H. (1996), The lack of a priori distinctions between learning algorithms. *Neural Computation 8*, 1341 - 1390.

Zucchini, W. (in press), An introduction to model selection. *Journal of Mathematical Psychology*.