Inductive Reasoning and Bounded Rationality

(The El Farol Problem)

by

W. Brian Arthur

Stanford University and Santa Fe Institute

Published in *Amer. Econ. Review (Papers and Proceedings)*, **84**, 406, 1994.

Given at the American Economic Association Annual Meetings, 1994

Session: Complexity in Economic Theory, chaired by Paul Krugman.

Citibank Professor, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501.

Inductive Reasoning and Bounded Rationality

By W. BRIAN ARTHUR *

The type of rationality we assume in economics—perfect, logical, deductive rationality—is extremely useful in generating solutions to theoretical problems. But it demands much of human behavior—much more in fact than it can usually deliver. If we were to imagine the vast collection of decision problems economic agents might conceivably deal with as a sea or an ocean, with the easier problems on top and more complicated ones at increasing depth, then deductive rationality would describe human behavior accurately only within a few feet of the surface. For example, the game Tic-Tac-Toe is simple, and we can readily find a perfectly rational, minimax solution to it. But we do not find rational "solutions" at the depth of Checkers; and certainly not at the still modest depths of Chess and Go.

There are two reasons for perfect or deductive rationality to break down under complication. The obvious one is that beyond a certain complicatedness, our logical apparatus ceases to cope—our rationality is bounded. The other is that in interactive situations of complication, agents can not rely upon the other agents they are dealing with to behave under perfect rationality, and so they are forced to guess their behavior. This lands them in a world of subjective beliefs, and subjective beliefs about subjective beliefs. Objective, well-defined, shared assumptions then cease to apply. In turn, rational, deductive reasoning—deriving a conclusion by perfect logical processes from well-defined premises—itself cannot apply. The problem becomes ill-defined.

As economists, of course, we are well aware of this. The question is not whether perfect rationality works, but rather what to put in its place. How do we model bounded rationality in economics? Many ideas have been suggested in the small but growing literature on bounded rationality; but there is not yet much convergence among them. In the behavioral sciences this is not the case. Modern psychologists are in reasonable agreement that in situations that are complicated or ill-defined, humans use characteristic and predictable methods of reasoning.

These methods are not deductive, but *inductive*.

In this paper I will argue that as economists we need to pay great attention to inductive reasoning; that it makes excellent sense as an intellectual process; and that it is not hard to model. In the main part of this paper, I will present a decision problem—the "bar problem"—in which inductive reasoning is assumed and modeled, and its implications are examined. The system that emerges under inductive reasoning will have connections both with evolution and complexity.

I. Thinking Inductively

How *do* humans reason in situations that are complicated or ill-defined? Modern psychology tells us that as humans we are only moderately good at deductive logic, and we make only moderate use of it. But we *are* superb at seeing or recognizing or matching patterns—behaviors that confer obvious evolutionary benefits. In problems of complication then, we look for patterns; and we simplify the problem by using these to construct temporary internal models or hypotheses or *schemata* to work with.[1] We carry out localized deductions based on our current hypotheses and act on them. And, as feedback from the environment comes in, we may strengthen or weaken our beliefs in our current hypotheses, discarding some when they cease to perform, and replacing them as needed with new ones. In other words, where we cannot fully reason or lack full definition of the problem, we use simple models to fill the gaps in our understanding. Such behavior is *inductive.*

We can see inductive behavior at work in Chess playing. Players typically study the current configuration of the board, and recall their opponent's play in past games, to discern patterns (De Groot, 1965). They use these to form hypotheses or internal models about each others' intended strategies, maybe even holding several in their minds at one time: "He's using a Caro-Kann defense." "This looks a bit like the 1936 Botvinnik-Vidmar game." "He is trying to build up his mid-board pawn formation." They make local deductions based on these—analyzing the possible implications of moves several moves deep. And as play unfolds they hold onto hypotheses or mental models that prove plausible, or toss them aside if not, generating new ones to put in their

place. In other words, they use a sequence of pattern recognition, hypothesis formation, deduction using currently-held hypotheses, and replacement of hypotheses as needed.

This type of behavior may not be familiar in economics. But we can recognize its advantages. It enables us to deal with complication: we construct plausible, simpler models that we *can* cope with. It enables us to deal with ill-definedness: where we have insufficient definition, our working models fill the gap. It is not antithetical to "reason," or to science for that matter. In fact, it is the way science itself operates and progresses.

Modeling Induction. If humans indeed reason in this way, how can we model this? In a typical problem that plays out over time, we might set up a collection of agents, probably heterogeneous, and assume they can form mental models, or hypotheses, or subjective beliefs. These beliefs might come in the form of simple mathematical expressions that can be used to describe or predict some variable or action; or of complicated expectational models of the type common in economics; or of statistical hypotheses; or of condition/prediction rules ("If situation Q is observed/predict outcome or action D"). These will normally be subjective, that is, they will differ among the agents. An agent may hold one in mind at a time, or several simultaneously.

Each agent will normally keep track of the performance of a private collection of such belief-models. When it comes time to make choices, he acts upon his currently most credible (or possibly most profitable) one. The others he keeps at the back of his mind, so to speak. Alternatively, he may act upon a combination of several. (However, humans tend to hold in mind many hypotheses and act on the most plausible one (Feldman, 1962).) Once actions are taken the aggregative picture is updated, and agents update the track record of all their hypotheses.

This is a system in which learning takes place. Agents "learn" which of their hypotheses work, and from time to time they may discard poorly performing hypotheses and generate new "ideas" to put in their place. Agents linger with their currently most believable hypothesis or belief model, but drop it when it no longer functions well, in favor of a better one. This causes a built-in hysteresis. A belief model is clung to not because it is "correct"—there is no way to

know this—but rather because it has worked in the past, and must cumulate a record of failure before it is worth discarding. In general, there may be a constant, slow turnover of hypotheses acted upon. We could speak of this as a system of *temporarily fulfilled expectations*—beliefs or models or hypotheses that are temporarily fulfilled (though not perfectly), that give way to different beliefs or hypotheses when they cease to be fulfilled.

If the reader finds this system unfamiliar, he or she might think of it as generalizing the standard economic learning framework which typically has agents sharing one expectational model with unknown parameters, acting upon their currently most plausible values. Here, by contrast, agents differ, and each uses several subjective models instead of a continuum of one commonly held one. This is a richer world, and we might ask whether, in a particular context, it converges to some standard equilibrium of beliefs; or whether it remains open-ended, always discovering new hypotheses, new ideas.

It is also a world that is evolutionary—or more accurately co-evolutionary. Just as species, to survive and reproduce, must prove themselves by competing and being adapted within an environment created by other species, in this world hypotheses, to be accurate and therefore acted upon, must prove themselves by competing and being adapted within an environment created by other agents' hypotheses. The set of ideas or hypotheses that are acted upon at any stage therefore coevolves. [2]

A key question remains. Where do the hypotheses or mental models come from? How are they generated? Behaviorally, this is a deep question in psychology, having to do with cognition, object representation, and pattern recognition. I will not go into it here. But there are some simple and practical options for modeling. Sometimes we might endow our agents with *focal* models—patterns or hypotheses that are obvious, simple and easily dealt with mentally. We might generate a "bank" of these and distribute them among the agents. Other times, given a suitable model-space, we might allow the genetic algorithm or some similar intelligent search device to generate ever "smarter" models. Whatever option is taken, it is important to be clear that the framework described above is independent of the specific hypotheses or beliefs used,

just as the consumer theory framework is independent of particular products chosen among. Of course, to use the framework in a particular problem, some system of generating beliefs must be adopted.

III. The Bar Problem

Consider now a problem I will construct to illustrate inductive reasoning and how it might be modeled. *N* people decide independently each week whether to go to a bar that offers entertainment on a certain night. For concreteness, let us set *N* at 100. Space is limited, and the evening is enjoyable if things are not too crowded—specifically, if fewer than 60% of the possible 100 are present. There is no way to tell the numbers coming for sure in advance, therefore a person or agent: *goes*—deems it worth going—if he expects fewer than 60 to show up, or *stays home* if he expects more than 60 to go. (There is no need that utility differ much above and below 60.) Choices are unaffected by previous visits; there is no collusion or prior communication among the agents; and the only information available is the numbers who came in past weeks. (The problem was inspired by the bar El Farol in Santa Fe which offers Irish music on Thursday nights; but the reader may recognize it as applying to noontime lunch-room crowding, and to other coordination problems with limits to desired coordination.) Of interest is the dynamics of the numbers attending from week to week.

Notice two interesting features of this problem. First, if there were an obvious model that all agents could use to forecast attendance and base their decisions on, then a deductive solution would be possible. But this is not the case here. Given the numbers attending in the recent past, a large number of expectational models might be reasonable and defensible. Thus, not knowing which model other agents might choose, a reference agent cannot choose his in a well-defined way. There is no deductively rational solution—no "correct" expectational model. From the agents' viewpoint, the problem is ill-defined and they are propelled into a world of induction. Second, and diabolically, any commonalty of expectations gets broken up: If all believe *few* will go, *all* will go. But this would invalidate that belief. Similarly, if all believe *most* will go, *nobody* will go, invalidating that belief. Expectations will be forced to differ.

At this stage, I invite the reader to pause and ponder how attendance might behave dynamically over time. Will it converge, and if so to what? Will it become chaotic? How might predictions be arrived at?

A Dynamic Model. To answer this, let us construct a model along the lines of the framework sketched above. Assume the 100 agents can individually each form several predictors or hypotheses, in the form of functions that map the past *d* weeks' attendance figures into next week's. For example, recent attendance might be:

… 44 78 56 15 23 67 84 34 45 76 40 56 22 35

And particular hypotheses or predictors might be: *predict next week's number to be*

– the same as last week's [35]

– a mirror image around 50 of last week's [65]

– 67 [67]

– a (rounded) average of the last four weeks [49]

– the trend in last 8 weeks, bounded by 0, 100 [29]

– the same as 2 weeks ago (2-period cycle detector) [22]

– the same as 5 weeks ago (5-period cycle detector) [76]

– etc. …

Assume each agent possesses and keeps track of a individualized set of *k* such focal predictors. He decides to go or stay according to the currently most accurate predictor in his set. (I will call this his *active* predictor). Once decisions are made, each agent learns the new attendance figure, and updates the accuracies of his monitored predictors.

Notice that in this bar problem, the set of hypotheses currently most credible and acted upon by the agents—the set of active hypotheses—determines the attendance. But the attendance history determines the set of active hypotheses. To use John Holland's term, we can think of these active hypotheses as forming an *ecology*. Of interest is how this ecology evolves over time.

Computer Experiments. For most sets of hypotheses, analytically this appears to be a difficult question. So in what follows I will proceed by computer experiments. In the experiments, to generate hypotheses, I first create an "alphabet soup" of predictors, in the form of several dozen focal predictors replicated many times. I then randomly ladle out k (6 or 12 or 23, say) of these to each of 100 agents. Each agent then possesses k predictors or hypotheses or "ideas" he can draw upon. We need not worry that useless predictors will muddy behavior. If predictors do not "work" they will not be used; if they do work they will come to the fore. Given starting conditions and the fixed set of predictors available to each agent, the future accuracies of all predictors are predetermined. The dynamics in this case are deterministic.

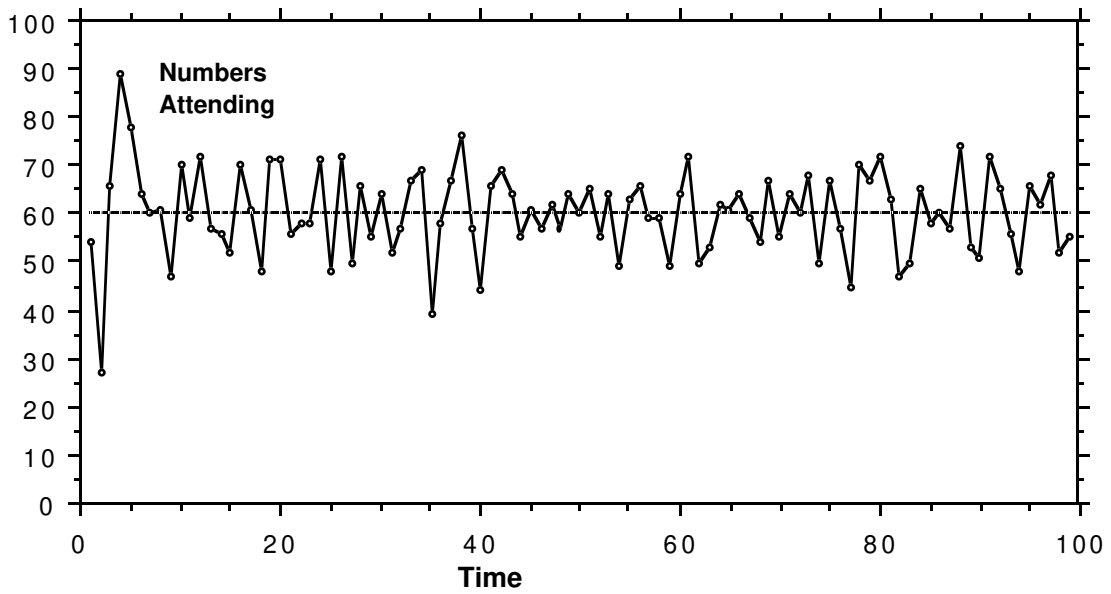

FIGURE 1. BAR ATTENDANCE IN THE FIRST 100 WEEKS.

The results of the experiments are interesting (Fig.1). Where cycle-detector predictors are present, cycles are quickly "arbitraged" away so there are no persistent cycles. (If several people expect many to go because many went three weeks ago, they will stay home.) More interestingly, mean attendance converges always to 60. In fact, the predictors self-organize into an equilibrium pattern or "ecology" in which of the active predictors, those most accurate and therefore acted

upon, on average 40% are forecasting above 60, 60% below 60. This emergent ecology is almost organic in nature. For, while the population of active predictors splits into this 60/40 average ratio, it keeps changing in membership forever. This is something like a forest whose contours do not change, but whose individual trees do. These results appear throughout the experiments, robust to changes in types of predictors created and in numbers assigned.

How do the predictors self-organize so that 60 emerges as average attendance and forecasts split into a 60/40 ratio? One explanation might be that 60 is a natural "attractor" in this bar problem; in fact if we view it as a pure game of predicting, a mixed strategy of forecasting above 60 with probability 0.4 and below it with probability 0.6 is Nash. But still this does not explain how the agents approximate any such outcome, given their realistic, subjective reasoning. To get some understanding of how this happens, suppose 70% of their predictors forecasted above 60 for a longish time. Then on average only 30 people would show up. But this would validate predictors that forecasted close to 30, restoring the "ecological" balance among predictions, so to speak. Eventually the 40%–60% combination would assert itself. (Making this argument mathematically exact appears to be non-trivial.) It is important to be clear that we do not need any 40-60 forecasting balance in the predictors that are set up. Many could have a tendency to predict high, but aggregate behavior calls the equilibrium predicting ratio to the fore. Of course, the result would fail if all predictors could only predict below 60—then all 100 agents would always show up. Predictors need to "cover" the available prediction space to some modest degree. The reader might ponder what would happen if all agents shared the same set of predictors.

It might be objected that I lumbered the agents in these experiments with fixed sets of clunky predictive models. If they could form more open-ended, intelligent predictions, different behavior might emerge. We could certainly test this using a more sophisticated procedure, say genetic programming (Koza, 1992). This continually generates new hypotheses—new predictive expressions—that adapt "intelligently" and often become more complicated as time progresses. But I would be surprised if this changes the above results in any qualitative way.

III. Conclusion

The inductive-reasoning system I have described above consists of a multitude of "elements" in the form of belief-models or hypotheses that adapt to the aggregate environment they jointly create. Thus it qualifies as an *adaptive complex* system. After some initial learning time, the hypotheses or mental models in use are mutually co-adapted. Thus we can think of a *consistent* set of mental models as a set of hypotheses that work well with each other under some criterion—that have a high degree of mutual adaptedness. Sometimes there is a unique such set, it corresponds to a standard rational expectations equilibrium, and beliefs gravitate into it. More often there is a high, possibly very high, multiplicity of such sets. In this case we might expect inductive reasoning systems in the economy—whether in stock-market speculating, in negotiating, in poker games, in oligopoly pricing, in positioning products in the market—to cycle through or temporarily lock into psychological patterns that may be non-recurrent, path-dependent, and increasingly complicated. The possibilities are rich.

Economists have long been uneasy with the assumption of perfect, deductive rationality in decision contexts that are complicated and potentially ill-defined. The level at which humans can apply perfect rationality is surprisingly modest. Yet it has not been clear how to deal with imperfect or bounded rationality. From the reasoning given above, I believe that as humans in these contexts we use *inductive* reasoning: we induce a variety of working hypotheses, act upon the most credible, and replace hypotheses with new ones if they cease to work. Such reasoning can be modeled in a variety of ways. Usually this leads to a rich psychological world in which agents' ideas or mental models compete for survival against other agents' ideas or mental models—a world that is both evolutionary and complex.

Footnotes

* Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501, and Stanford University. I thank particularly John Holland whose work inspired many of the ideas here. I also thank Kenneth Arrow, David Lane, David Rumelhart, Roger Shepard, Glen Swindle, and colleagues at Santa Fe and Stanford for discussions. A lengthier version is given in Arthur (1992). For parallel work on bounded rationality and induction, but applied to macroeconomics, see Sargent (1994).

[1] For accounts in psychological literature, see Bower and Hilgard (1981), Holland *et al.* (1986), Rumelhart (1980), and Schank and Abelson (1977). Not all decision problems of course work this way. Most of our mundane actions like walking or driving are subconsciously directed, and for these pattern-cognition maps directly in action. Here connectionist models work better.

[2] A similar statement holds for strategies in evolutionary game theory; but there, instead of a large number of private, subjective expectational models, a small number of strategies compete.

REFERENCES

Arthur, W. Brian, "On Learning and Adaptation in the Economy," Santa Fe Institute Paper 92-07-038, 1992.

Bower, Gordon H. and Hilgard, Ernest R., *Theories of Learning*, Englewood Cliffs: Prentice Hall, 1981.

De Groot Adriann, *Thought and Choice in Chess*, in the series *Psychological Studies,* 4, Paris: Mouton & Co., 1965.

Feldman, Julian "Computer Simulation of Cognitive Processes," in Harold Borko (ed.), *Computer Applications in the Behavioral Sciences*, Prentice Hall, 1962.

Holland, John H., Keith J. Holyoak, Richard E. Nisbett and Paul R. Thagard, *Induction.* Cambridge, Mass: MIT Press, 1986.

Koza, John. *Genetic Programming.* Cambridge, Mass: MIT Press, 1992.

Rumelhart, David, "Schemata: the Building Blocks of Cognition," in R. Spiro, B. Bruce, and W. Brewer (eds.), *Theoretical Issues in Reading Comprehension.* Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Sargent, Thomas, J. *Bounded Rationality in Macroeconomics.* Oxford University Press, 1994.

Schank R. and R.P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.* Hillsdale, N.J.: Lawrence Erlbaum, 1977.