QED

Information Processing and Bounded Rationality: A Survey

Barton L. Lipman

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-1993

Discussion Paper #872

Information Processing and Bounded
Rationality: A Survey

by

Barton L. Lipman
Queen's University

January 1993

Information Processing and Bounded Rationality: A Survey

by

Barton L. Lipman

Department of Economics
Queen's University
Kingston, Ontario K7L 3N6

email: lipmanb@qucdn.queensu.ca

Queen's University Discussion Paper #872

First Draft

January 1993

ABSTRACT

This paper surveys some recent attempts to formulate a plausible and tractable model of bounded rationality. I focus in particular on models which view bounded rationality as stemming from limited information processing. I discuss computability, partitional models (such as automata, perceptrons, and optimal networks), nonpartitional models, and axiomatic approaches.

I. Introduction.

In response to criticisms of the assumption of perfect rationality (see, for example, Simon [1955, 1976], Selten [1978], Binmore [1987, 1988]), much attention has turned to finding a tractable and plausible model of *bounded rationality*. Herbert Simon, the originator of the phrase, defines bounded rationality as "rational choice that takes into account the cognitive limitations of the decision–maker — limitations of both knowledge and computational capacity." (Simon [1987].) A useful way to view bounded rationality is to think of an agent as an information processor. "Inputs" (information) flow to the agent, he processes them in some fashion, and "outputs" (decisions) come out. Bounded rationality refers to choice which is imperfect in the sense that the output is often not the "correct" one, but which is sensible in that it can be understood as an attempt by the agent to do reasonably well. Put differently, the procedure used is a reasonable compromise between accuracy of the output and the difficulties involved in processing.

Despite the fact that many economists view bounded rationality as a more realistic and more appropriate assumption than perfect rationality at least for many situations, there are very few papers which explore the implications of bounded rationality. The reason for this is very simple: there is no clear agreement on how one models this phenomenon. Instead, the focus of the literature has been on exploring various modeling approaches. In this survey, I will review some of the approaches that have been suggested in recent years. I should warn the reader that, given the state of the literature, this survey will be quite unlike the others in this volume. While I will mention some theorems, the focus will be on ideas and approaches, not results or established facts.

The various approaches I discuss have one theme in common: all treat bounded rationality as limited information processing. However, they vary greatly in the way these limitations are modeled. In Section II, after presenting some notation and terminology, I discuss with the most fundamen-

tal restriction on information processing: *computability*. It seems quite reasonable to believe that real agents cannot use functions that cannot be computed. Surprisingly, restricting agents to such functions can have very strong implications in economic models (Spear [1989], Anderlini [1991], Anderlini and Felli [1992]).

In Sections III and IV, I look at models which put more severe restrictions on information processing. Section III focuses on *partitional* models of information processing. The common feature of these is that information processing can be summarized in terms of a partition of the set of "external states" or "inputs." A large number of apparently quite different models fit into this framework. In this section, I first discuss some models which assume some form of cost function for partitions. I primarily discuss the automata literature (Rubinstein [1986], Abreu and Rubinstein [1988], Kalai and Stanford [1988]), but also briefly discuss a few related models. Secondly, I discuss models where the costs of partitions are, at least to some extent, derived from a more detailed model of the way the partition is constructed. In this context, I discuss threshold models (Rubinstein [1992], Cho and Matsui [1992a, 1992b]) and optimal processing networks (Mount and Reiter [1990], Radner and Van Zandt [1992]).

In Section IV, I discuss *nonpartitional* models — that is, models in which information processing does not generate a partition (Bacharach [1985], Geanakoplos [1989, 1992]). This approach seems to have the advantage of greater generality in that partitions are special cases of nonpartitional structures. On the other hand, it seems more difficult to endogenize information processing in this framework. While all the models in Section III derive the nature of information processing from some primitives at least to some extent, so far, there has been very little analogous work with nonpartitional structures.

In Section V, I discuss three recent papers which can be thought of as attempts to provide an axiomatic derivation of certain forms of information

processing. In particular, I discuss Morris' [1992b] work on nonpartional information structures, Lipman's [1992] derivation of impossible possible worlds, and Gilboa and Schmeidler's [1992b] notion of case–based reasoning. Some concluding remarks are offered in Section VI.

Remark 1. Unfortunately, I have had to omit many topics related to these issues. In this remark, I will briefly comment on a few of the most serious omissions. First, I have left out the literature on intransitive indifference (see Suppes, *et al* [1989] for a survey or Sileo [1992] for an interesting recent paper), despite the fact that this seems to make a very nice model of imperfect perceptual ability. Second, I have omitted the work on non-additive probabilities (see Gilboa and Schmeidler [1992a] for a particularly intriguing introduction to the subject), even though many (such as Shafer [1976], Binmore [1991], and Walley [1991]) have described this and related approaches as more realistic models of humans. Third, I have omitted a variety of work on relaxations of expected utility, even though some of these papers seem to be motivated by bounded rationality considerations (particularly Rubinstein [1988] and Segal [1990]). Fourth, I have not included any discussion of the "decision–theoretic" approach to game theory (Bernheim [1984], Pearce [1984], Aumann [1987], Aumann and Brandenburger [1991]), despite its focus on the effects of knowledge regarding the rationality of others. Finally, I only briefly mention a few applications of the theory. This is not to suggest that applications are unimportant — to the contrary, I argue in the conclusion that they are crucial. However, to date, there have not been very many applications of these models to specific problems in economics and game theory.

II. Computability.

Each of the models I discuss can be roughly described in the following way. The agent observes some external events, which I shall often refer to as *inputs*. The way the agent processes these inputs determines his "state of mind." His state of mind determines the way he views the available actions

and thus ultimately determines his behavior.

More formally, let Ω denote the set of *external states*. The agent has a prior probability distribution, q, on Ω. In most of the models I discuss, the agent will directly observe the state ω, so these will also be the inputs. Let A be the set of *actions* the agent can choose among. So that the choice problem is nontrivial, I will always assume that A has at least two elements. The agent chooses a *behavior rule*, which is a function $f : \Omega \to A$.

Remark 2. Normally, one thinks of the external state as a complete specification of all relevant factors. When external states are assumed to be observed directly, it is perhaps better to view an external state as specifying relevant factors only in as much detail as an agent who perfectly processes information could determine.

The most basic requirement we could place on information processing is to restrict the agent to using a *computable* or *recursive* behavior rule. There are several equivalent ways to define this set of functions. Loosely speaking, it is the set of functions f for which an algorithm exists such that for each input ω, the algorithm eventually (that is, in finitely many steps) produces output $f(\omega)$. This requirement is only restrictive when Ω is infinite — any function with a finite domain is computable. Typically, one assumes that Ω is countable.

It is easy to show that there are only countably many computable functions. To see this intuitively, imagine you are writing a computer program. There are only finitely many different commands you can use at any given line of the program. Furthermore, the program must be of finite length. Hence there are countably many different programs you could write. Therefore, there are at most countably many functions you can compute. On the other hand, if Ω is infinite, the set of all functions from Ω to A is uncountable. In this sense, "most" functions are not computable.

For a more intuitive sense of why some functions are not computable, consider the so–called *Halting problem*. We say that a computer program *halts* on a given input if it eventually ceases computation and gives an answer. The Halting problem is the problem of determining whether a given computer program eventually halts on a given input. Obviously, if a given program does halt on a given input, we can learn this fact simply by running the program. However, if it does not halt, it is not obvious that we can determine this fact. Certainly, if we can recognize an infinite loop in the program, we can determine that it will not halt. But clearly there is no simple procedure for finding such a loop. In fact, one can show that the Halting problem is *undecidable* — there is no computable function which determines whether a given program halts on a given input for every program and every input.

Of course, it is difficult to imagine a real agent using a function which is not computable and so it seems only reasonable to restrict agents to computable functions. A number of authors have shown some very surprising consequences of this restriction in certain contexts. For example, Anderlini [1991] considers shows that in common interest games with communication, a restriction to computable strategies and, loosely speaking, computably trembling–hand perfect equilibria implies cooperation, even in a one–shot game. Unfortunately, to explain this and other papers in detail would require a lengthy digression on the mathematics of computable functions. I encourage the interested reader to consult Cutland [1980] for an especially readable introduction to the subject and Anderlini, Spear [1989], or Anderlini and Felli [1992] for some of the more interesting applications.

III. Partitional Models.

In this section and the next, I turn to models which analyze information processing in more detail. In the first class of models I discuss, information processing is represented by a *partition* Π of Ω. A partition Π of a set, say Z, is a collection of subsets of Z with the property that every $z \in Z$ is

in exactly one of these subsets. The elements of the partition Π are often referred to as *events*. A partition Π is said to be *finer* than a partition Π' if for every event $\pi' \in \Pi'$, there are events $\pi_1, \ldots, \pi_k \in \Pi$ such that $\cup_i \pi_i = \pi'$. This is also described by saying that Π' is *coarser* than Π. Intuitively, if Π is finer than Π', then learning which event of Π contains a given z conveys more information than only learning which event of Π' contains z.

Partitions are intimately related to *equivalence relations*. An equivalence relation, say R, on a set Z is a binary relation which is reflexive (that is, zRz always holds), symmetric (zRz' implies $z'Rz$), and transitive (zRz' and $z'Rz''$ implies zRz''). The simplest example of an equivalence relation is the standard notion of "equal to," or $=$. A familiar example to economists is the usual indifference relation. Given any equivalence relation, we can partition Z into *equivalence classes* as follows. For any given $z \in Z$, let

$$R(z) = \{z' \in Z \mid zRz'\}.$$

Thus $R(z)$ is the set of points equivalent under R to z. These sets are called the equivalence classes of R. It is not hard to show that for any z and z', either $R(z) = R(z')$ or else $R(z) \cap R(z') = \emptyset$. Furthermore, since $z \in R(z)$, every point in Z is contained in an $R(z)$ set for some z. Hence the equivalence classes of R form a partition of Z, called the partition *induced* by R. Similarly, given any partition Π, we can define an equivalence relation induced by the partition by saying that z and z' are equivalent if they are contained in the same event in Π.

To return to information processing, the idea is that the agent only processes his information to the point of determining which event of Π contains it — after processing, this is *all* he knows. Thus Π describes how thoroughly the agent processes his information. If Π has only one event (which would have to be the entire set Ω), then the agent is not processing his input at all — he effectively ignores it. By contrast, a partition that has a different event for each different ω involves complete processing. That is, the agent processes the input so thoroughly that he recognizes every

possible distinction between external states. More generally, the finer is his partition, the more processing of his information he is carrying out.

This modeling approach may seem odd. After all, the agent does see the input, so why is he treated as only able to know which event of some partition contains it? To understand the approach, it is important to keep in mind that there is a distinction between "seeing" the input and recognizing all the implications it has. For example, suppose the inputs are real numbers and that given input ω, the agent wishes to compute the value of some very complicated function g at the point ω. For concreteness, suppose the input ω received by the agent is 3. Certainly, the agent knows that the input is 3 — it takes virtually no cognitive effort at all for the agent to recognize the input in this sense! However, the agent may not immediately recognize all the implications of this input. That is, he may not immediately recognize what number $g(3)$ is. In this sense, there is a relevant attribute of the input which the agent does *not* recognize. To be more concrete still, suppose that the processing the agent does leads him to certain beliefs regarding the value of $g(3)$. Suppose, further, that he would have identical beliefs regarding the value of $g(3.1)$ if the input had been 3.1 instead. Then we can treat these two inputs as equivalent from the agent's point of view and hence would include them in the same event of the partition.

The agent's payoff depends on his action, the state, and information processing costs. His optimization can be described in two steps. First, given a partition Π and an event $\pi \in \Pi$, the agent chooses his action a to maximize

$$E_\omega[u(a, \omega) \mid \omega \in \pi].$$

The notation "$\mid \omega \in \pi$" means that given the event π, the agent updates his probability distribution q over Ω using Bayes' Rule to concentrate the distribution on π. He then chooses the action which maximizes his conditional expected utility. Let $f(\pi)$ denote an a which solves this maximization. Abusing terminology slightly, I will call such a function a *behavior rule*.

7

Let

$$V(\Pi) = \sum_{\pi \in \Pi} \mathrm{E}_\omega[u(f(\pi), \omega) \mid \omega \in \pi]q(\pi).$$

Thus V gives the *ex ante* expected payoff associated with the information partition Π.

An alternative way to describe the first step, which will sometimes prove convenient, is to say that given a partition Π, the agent chooses the optimal behavior rule f from the set of behavior rules which are *measurable* with respect to Π. A function $f : \Omega \to A$ is measurable with respect to Π if for every event $\pi \in \Pi$ and every ω, $\omega' \in \pi$, we have $f(\omega) = f(\omega')$. Letting $\mathcal{F}(\Pi)$ denote the set of functions measurable with respect to Π, then, we can redescribe the first step by saying that the agent chooses some $f \in \mathcal{F}(\Pi)$ to maximize $E_\omega u(f(\omega), \omega)$. Again, let $V(\Pi)$ denote the agent's expected utility at this optimum. It is important to note that the equivalence of these two approaches does not always hold with nonpartitional information structures.

In the second step, the agent chooses Π to maximize $V(\Pi) - c(\Pi)$, where c gives the expected information processing costs. In some of the models, the agent is constrained to choose Π from some set $\mathcal{P}$ of feasible partitions.

Remark 3. People are often troubled by the assumption that the agent chooses the partition optimally. Certainly, it seems odd to model bounded rationality by assuming a particular form of optimal choice! Even worse, the problem of finding the optimal partition seems like a more complex problem that simply choosing the optimal action given some ω in the first place. There are at least two replies to this criticism. First, it seems more reasonable to believe that the knowledge agents have is general knowledge rather than specific knowledge. Put differently, they know how to solve problems, even if they do not know the solution to every problem. By analogy, there is a sense in which deriving the quadratic formula is harder than finding the roots to a given quadratic equation. However, *ex ante*,

one does not know which quadratic equation one will need to solve. Hence one finds it more useful to know the quadratic formula than to try learning the roots of various quadratic equations! The choice of Π and f models the agent's choice of how to solve problems and, by this argument, it seems more reasonable to believe that this choice is made well than that the choice given any particular ω is made well. An alternative response is that if we are uncomfortable with the assumption that the agent chooses Π and f optimally, we must be saying that we have not completely specified the model. That is, our model must be viewed as a model of the world *as the agent perceives it*. If the model is truly a complete description of the agent's perception, then how can he fail to choose what he perceives to be best for him? (This is basically the argument given in Lipman [1991].) A different form of this view will be discussed briefly in Section V.

The rest of this section is divided in two parts. First, I discuss some models which make certain assumptions about the cost functions and derive their implications. The main body of research along these lines has been the automata literature, but I also very briefly discuss a variety of other applications of this idea. Second, I discuss two models which analyze the construction of the partitions in more detail.

A. Exogenous Cost Functions.

1. Automata.

In response to a suggestion by Aumann [1981], Neyman [1985] and Rubinstein [1986] initiated the study of repeated games in which players were restricted to using *finite state automata* (more precisely, Moore machines) to implement their strategies. Both papers treated limitations on rationality as limitations on the number of states of the automata, Neyman by imposing constraints on the number of states, Rubinstein by assuming that states are costly. These papers led to a large body of research, much

of which is summarized in Kalai [1991].[1]

An automaton is a stylized description of a simple computing device, which has been studied in the computer science literature since the mid–1950's. (See Hopcroft and Ullman [1979] for an introduction to the subject.) Briefly, an automaton consists of a set of *internal states*, one of which is designated the *initial state*, a *transition function* which specifies how the automata changes states in response to the opponents' actions, and an *output function* which gives an action as a function of the state. (I will give some examples below.) Rather than focus on automata themselves, I will follow an equivalent approach first noted by Kalai and Stanford [1988], which makes clear how automata fit into the framework above.

Suppose we have an infinitely repeated game. Let H denote the set of possible histories of play. A strategy in the game is a function σ which specifies an action as a function of the history of the game. After any history, say h, the remaining game is still an infinitely repeated game. Hence a strategy for the overall game, σ, in effect, specifies a continuation strategy for the game which follows the history h. This is what Kalai and Stanford call the *induced strategy $\sigma|h$*. More precisely, $\sigma|h$ is the strategy which, on a history h', specifies the action that σ would prescribe on the history h followed by h'. We can then say that two histories, h and h', are equivalent under σ if they lead to the same induced strategy — that is, if $\sigma|h = \sigma|h'$. It is easy to show that this is an equivalence relation so that it generates a partition of H which will be denoted by $\mathcal{H}(\sigma)$. Clearly, if we know which event of this partition a history lies in, we know enough about the history to determine the strategy it induces. Kalai and Stanford show that the number of internal states of the smallest automata which plays a given strategy is equal to the number of sets in this partition.

---

[1] Some relevant papers written since Kalai's survey are Neme and Quintas [1992], Piccione [1992], and Piccione and Rubinstein [1992].

10

To understand this notion more intuitively, it is useful to consider a few examples. Suppose we are analyzing a repeated Prisoners' Dilemma game. The actions are "cooperate" and "defect." The strategy "always cooperate" has exactly one induced strategy: namely, itself. On any history at all, the continuation strategy is identical to the original strategy. Hence the partition has just one event equal to the set of all histories. The automaton this corresponds to has just one state, an output function which specifies cooperation in this one state, and a trivial transition function which stipulates remaining in this one state in response to any action by the opponent.

A slightly more complex example is "tit–for–tat." This strategy begins by cooperating and, on every subsequent history, does whatever the opponent did on the previous play. This strategy has two induced strategies. On the "empty history" — that is, at the beginning of the game — or on any history where the opponent cooperated in the previous period, the induced strategy is tit–for–tat. On any history where the opponent defected in the previous period, it is the variant of tit–for–tat which begins by defecting and then does whatever the opponent did on the previous move. Hence the partition has two events: those histories where the opponent cooperated on the last move and those histories where the opponent defected on the last move. The automaton this corresponds to has two states, one which specifies cooperation as its output and one which specifies defection. The transition function specifies moving to the cooperative state in response to cooperation and the defecting state in response to defection.

For a last example, consider a k–period trigger strategy. This strategy cooperates until the opponent defects. At this point, it enters a "punishment mode," defecting for the next k periods regardless of what the opponent does. Any choices by the opponent are ignored during this period. After the k defections, the strategy returns to cooperation until the next defection by the opponent. It is not hard to see that this strategy has $k + 1$ induced strategies. On the empty history, histories where the opponent has never defected, or any history where the punishment is com-

plete, the induced strategy is the k–period trigger strategy. Then there are k induced strategies corresponding to the k periods of punishment. In any one of them, there is a certain number of periods in which the player intends to defect and moves by the opponent are ignored. Since the number of remaining periods like this differs over the punishment cycle, each such induced strategy is different. Hence the partition this strategy induces has $k + 1$ events. Again, the corresponding automaton has $k + 1$ states, one with cooperation as its output and the others with defection. The transition function specifies staying in the cooperative state in response to cooperation. In response to defection while in the cooperative state, it moves to the first of the k punishment states. From there, it moves through the k punishment states in sequence in response to any output, finally moving back to the cooperative state.

To relate this to the previous framework, the set of histories plays the role of the set of external states, so $H = \Omega$. (Of course, in this case, the probability distribution over H is determined endogenously by the strategy choices of the players in the game.) The set of "actions," A, is the set of strategies for the infinitely repeated game. Hence any strategy σ can be described as a function f from $\mathcal{H}(\sigma)$ into A where $f(h) = \sigma|h$. Thus we can divide the choice of a strategy σ into the choice of a partition of the set of histories, Π, and a function from Π to the set of strategies.[2] The cost function c in most of the literature is a function only of the number of events of the partition and is an increasing function of this. Lipman and Srivastava [1990] and Banks and Sundaram [1990] analyze the implications of alternative cost functions.

---

[2] It is important to note that not every partition of H can be generated by some strategy. For example, there is no strategy that induces the partition which puts every history except a single one–period history into one event and this single history into another. Similarly, not every function from such a partition to A will form a legitimate strategy.

2. Related Models.

Rosenthal [1991a, 1991b] suggests that agents adopt simple rules of thumb which are approximately optimal for a wide variety of circumstances, even though they may never be exactly optimal. (This idea has a long history in economics. See, for example, Baumol and Quandt [1964].) To model this, he assumes that there is a set of possible games, Γ, that the agent might play. The input the agent recevies is simply the game to be played, so $\Omega = \Gamma$. For simplicity, suppose that there is the same set of feasible strategies in each of these games. This set of strategies will be the set of actions A. A rule of thumb is a function $r : \Gamma \rightarrow A$ specifying what action to use in which game. Rosenthal assumes that there are costs to these rules, where the costs increase with the number of different actions which might be used. Putting it differently, say that two games, G and G', are equivalent given r if $r(G) = r(G')$. Let Π be the partition this equivalence relation induces. Now we see that the agent's choice of r corresponds directly to a choice of a partition Π and a f measurable with respect to Π. Rosenthal's cost function is simply an increasing function of the number of events of the partition.

Dow [1991] considers a model of search with limited memory. Here the input is a pair of prices, one observed in the first period and one in the second, where the agent knows he will not be able to remember the first one exactly. An action is a choice of which price to purchase the good at. Dow's formulation of limited memory is that the agent has a partition of the set of possible prices and, in the second period, only remembers which event of the partition the first period price fell into. Hence he works directly with a partition Π. He also assumes that costs are proportional to the number of events of the partition.

A last example of this type of model treats computation as equivalent to costly information acquisition. It has often been suggested (see, e.g., Conlisk [1988] or Lipman [1991]) that computation is the extraction of

13

information from facts one already knows and hence is analogous to acquisition of information. This is modeled by treating computation is simply a choice of a partition Π, such that the agent must choose f to be measurable with respect to Π. In this literature, it is much less common to assume that the cost of a partition is simply an increasing function of the number of events in the partition.

B. Modeling the Construction of the Partition.

1. Threshold Models.

In this class of models, there are no costs associated with information processing, so $c(\Pi) = 0$ for all $\Pi \in \mathcal{P}$. However, the processing of information is constrained to take a particular form. Rather than describe the set of feasible partitions directly, it is described indirectly by requiring that the partition be constructed according to the following two–step procedure. First, the agent "translates" ω into a real number using what I will call a *processing function* μ. Second, the agent observes whether or not $\mu(\omega)$ is above a certain "threshold." The important point is that the agent is only able to divide Ω into two sets — those for which $\mu(\omega)$ is above the threshold and those for which it is not.

The threshold form may seem arbitrary at first glance. On the other hand, it does provide a simple way of restricting the construction of partitions. Intuitively, the agent is only able to determine which of two "positions" ω gives greater support to. Naturally, one may be interested in generalizing the idea to multiple thresholds or some other intermediate level of information processing.

One example of the use of threshold models in economics is Rubinstein's [1992] application of Minsky and Papert's [1988] perceptrons.[3] Here,

---

[3] The only other examples I am aware of are Cho and Matsui [1992a, 1992b].

it is assumed that $\Omega \subseteq \mathbf{R}^k$ for some $k \geq 2$. The processing is constrained in the following way. The agent has some number of functions, ϕ, called *perceptrons*, each of which analyze the input ω. A perceptron of order n only conditions on n components of the vector ω — that is, it is completely independent of what appears in the other $k - n$ components. Then the processing function is $\mu(\omega) = \sum_i \phi_i(\omega)$ where the ϕ_i's are the agent's perceptrons. Notice that the agent's sophistication can be restricted either by restricting the number or order of his perceptrons.[4]

To get a sense of how these restrictions affect the set of feasible partitions, consider the following example drawn from the Rubinstein paper. Suppose there are four states in Ω: (b, b), (b, d), (d, b), and (d, d) where b and d are real numbers, $b \neq d$. Obviously, all perceptrons are either of order 1 or order 2. Because the conditioning has to have a threshold form, no feasible partition has more than two events. However, it is easy to see that it is possible to generate any such partition with a single perceptron of order 2.

Suppose, though, that one is restricted to order 1 perceptrons. Clearly, there is no need to have more than two perceptrons since there is no need to condition twice on the same component. Let ϕ_1 and ϕ_2 be two order 1 perceptrons, where ϕ_i conditions only on the ith component of ω. To get a sense of some of the possible partitions, let $\phi_1(b) = 1$, $\phi_1(d) = 2$, $\phi_2(b) = -2$, and $\phi_2(d) = -4$. Then $\mu((b, d)) = -3$, $\mu((d, d)) = -2$, $\mu((b, b)) = -1$, and $\mu((d, b)) = 0$. Clearly, by choosing the threshold appropriately, one can obtain a partition with (b, d) by itself and the other three in a single event (with a threshold of -2.5), a partition with (b, d) and (d, d) in one event and the other two states in the other (threshold of -1.5), or a partition with (d, b) in one event and the other three in the other (threshold of $-.5$). More generally, it is easy to see that with an appropriate choice

---

[4] Minsky and Papert [1988] discuss a variety of other ways to require perceptrons to depend only on limited amounts of information about the input.

of ϕ_1 and ϕ_2, one can obtain any two event partition where one event is a singleton. (To do so, simply make ϕ_i on the ith component of this singleton very negative and make ϕ_i equal zero otherwise.) As to two–event partitions where each event contains two states, we showed above how to obtain $\{\{(b,d),(d,d)\},\{(b,b),(d,b)\}\}$. It is not hard to find perceptrons which generate the partition $\{\{(b,d),(b,b)\},\{(d,d),(d,b)\}\}$.

This leaves only the partition $\{\{(b,d),(d,b)\},\{(b,b),(d,d)\}\}$. This partition *cannot* be generated with order 1 perceptrons. To see this, suppose that it is generated by with perceptrons ϕ_1 and ϕ_2 and threshold α. Then we must have[5]

$$\phi_1(b) + \phi_2(d) > \alpha$$
$$\phi_1(d) + \phi_2(b) > \alpha$$

and

$$\phi_1(b) + \phi_2(b) < \alpha$$
$$\phi_1(d) + \phi_2(d) < \alpha.$$

Clearly, though, these two pairs of inequalities are contradictory. Summing the first two yields the opposite inequality obtained from summing the second two.

2. Processing Networks.

As with the threshold models, the work on processing networks provides a more detailed analysis of how information is processed. Because I understand it better (or, at least, believe I do!), I will focus on the approach taken by Radner [1989] and further pursued in Van Zandt [1990] and Radner and Van Zandt [1992].[6] The interested reader should also consult the related work of Mount and Reiter [1990].

---

[5] Of course, reversing all four inequalities also generates this partition, but this possibility leads to the same contradiction.

[6] A very clear introduction to this work is contained in Radner [1992].

In these models, information flows into a network of *processors*. Each processor has an "in–box," a register where it keeps a running total, and the ability to send its total to certain, prespecified processors. In each period, a processor can remove an item from its in–box, add it to the current register,[7] and send the value of the register to another processor. If it sends the value of its register to another processor, this information goes into the receiving processor's in–box and the sending processor's register is set to zero. Also in each period, information from outside the network flows into the in–boxes of certain prespecified processors. The choice variables in a network are: (1) the number of processors, (2) which processors receive how much incoming information, (3) when processors send their register to other processors, and (4) which processors they send the information to.

To understand the model, consider the following example drawn from Radner [1992]. Suppose that information only comes in to the network in the first period and that it consists of a vector of 40 numbers. The goal of the network is to compute the sum of these 40 numbers. Put differently, the set of inputs, Ω, is $\mathbf{R}^{40}$. The partition we wish to generate is the one is induced by the equivalence relation which treats vectors as equivalent if their components add to the same number. We wish to generate this partition at the lowest possible cost where the cost is increasing in both the number of processors required and the number of periods it takes to compute the sum.

One network which generates this partition is the following: Each of eight processors receives five of the incoming numbers. Over the course of the first five periods, the processors add the five numbers together. At the end of period 5, the eight processors send the totals to four more processors, each of which receives information from two processors. In periods 6 and 7, these processors add the incoming numbers together. At the end of period 7, they forward the totals to two other processors, each of which receives

---

[7] Addition can be replaced by any associative operation.

information from two processors. They total their incoming numbers in periods 8 and 9, forwarding the totals to a single processor at the end of period 9. Finally, in periods 10 and 11, the last processor computes the overall sum. This network requires 15 processors and takes 11 periods to compute the sum.

Clearly, though, there is much redundancy in this network since the "higher up" processors are idle while waiting for the lower processors to add. In fact, the 40 numbers can be summed by a network with only 8 processors in 8 periods. To do this, each of the eight processors receives five numbers. As above, the processors spend periods 1 through 5 adding their numbers. At this point, four of the processors send their totals to the other four, each processor receiving one number. This is added to the processor's previous total in period 6. At the end of this period, two of the four processors send their totals to the other two. These numbers are added to previous totals in period 7, after which one processor sends its total to the other. Finally, the overall total is computed in period 8. Radner [1989] shows that this network is efficient in the sense that one cannot generate this partition with both fewer processors and less delay.

Since there is a tradeoff between processors and periods, this is not necessarily the minimum cost network. For example, the sum could also be computed by a network with one less processor in one more period. Depending on the relative cost of processors versus delay, either (or neither) network could be optimal.

More generally, it may be optimal to choose a network which only provides a coarser partition. The partition described above is the finest partition that a network of processors can generate. Once the costs of processors and delay are specified,[8] one can derive a cost function for any

---

[8] In the general framework described at the beginning of this section, the costs of the partition are given by some function $c(\Pi)$, completely independently of the "benefits"

feasible partition. Then, given a probability distribution over Ω, a set of actions A, and a payoff function u, one can compute the optimal partition, the optimal network for generating it, and the optimal behavior rule measurable with respect to that partition. When new information arrives in every period as in Van Zandt [1990] or Radner and Van Zandt [1992], the problem is vastly more complex, but the idea is the same.

I should emphasize that Radner primarily discusses these networks as models of firms, rather than of individual agents.[9] In this context, the structure of optimal networks is itself of great interest.

IV. Nonpartitional Models.

Traditionally, information is treated as a partition (see Blackwell [1951], Aumann [1976], etc.). Changing the approach of the previous section slightly, suppose that there is a set of inputs or signals S and a function $\xi : \Omega \rightarrow S$ which tells which input the agent observes as a function of the external state. In the traditional analysis, we simply say that ω and ω' are equivalent if $\xi(\omega) = \xi(\omega')$ and assume that information is described by the partition of Ω induced by this equivalence relation. The models discussed in Section III can all be thought of as assuming that $\Omega = S$ and $\xi(\omega) = \omega$. They relax the standard assumption that the agent fully processes his information. If he did, his partition would be the partition into singletons. Instead, he typically chooses some coarser partition because of constraints on or costs of information processing. However, these models do maintain the traditional assumption that processed information takes the form of a partition.

---

of the partition, $V(\Pi)$. When delay matters, it may be more natural to model costs through the use of a discount factor. In this approach, benefits and costs cannot be separated quite so simply, but the basic principle is unchanged.

[9] See Turnbull [1992] for a related approach to modeling firms as networks of "processing machines."

However, if an agent does not fully process the information a signal reveals, he may not have his information in the form of a partition. Bacharach [1985] was the first in the economics literature to point out this possibility. As he noted, there is a very large, very old literature in philosophy on *modal logic* which addresses this subject. (See Hughes and Cresswell [1968] or Chellas [1990] for an introduction.) Geanakoplos [1989] was the first to define and analyze a decision theory for nonpartitional information.

To relate this approach to that used in Section III, I will continue to let Ω denote the set of external states as well as the set of inputs. Instead of representing information processing by a partition, it will be represented by what Geanakoplos [1989] calls a *possibility correspondence*. This is a function $P : \Omega \to 2^{\Omega}$, where 2^{Ω} is the power set of Ω — that is, the set of all subsets of Ω. Intuitively, $P(\omega)$ is the set of ω' that the agent thinks are possible after processing input ω. Here P is *not* a choice variable. Instead, it is assumed to satisfy certain properties, some of which I discuss below.

To understand the difference between this approach and the partitional models, it is important to recognize that there is still one aspect of nonpartitional models that is quite partitional — and unavoidably so. In both partitional and nonpartitional models, we can partition Ω by treating as equivalent inputs that the agent responds identically to. In the nonpartitional case, we would do this by saying that ω and ω' are equivalent if $P(\omega) = P(\omega')$ and constructing the partition this relation induces. The key difference between the partitional and nonpartitional models is that the former treat the agent as basing his decisions on this partition, while the latter drop this assumption. In other words, in the partitional models, the agent understands how he is responding to inputs and behaves accordingly. In nonpartitional models, he has only an imperfect understanding of his own information processing.

Some properties of the possibility correspondence that might be assumed are the following. The first assumption, often called *nondelusion*, is

that $\omega \in P(\omega)$. In other words, in state ω, the agent does not mistakenly believe that the state could not possibly be ω. For obvious reasons, this assumption is almost always imposed.

A second property is what Geanakoplos calls *knowing that you know* and is also called *positive introspection*. It says that $\omega' \in P(\omega)$ implies $P(\omega') \subseteq P(\omega)$. To understand this property intuitively and to see why it gets its name, suppose it does not hold. Suppose ω is the true state and suppose that $\omega' \in P(\omega)$ but $P(\omega') \not\subseteq P(\omega)$. For simplicity, let $P = P(\omega)$ and let β denote the statement that the true state is in P. Since ω is the true state, the set of states the agent considers possible is P. Hence the agent knows that β is true. However, in one of the states he thinks of as possible, namely ω', he would not know that the state is in P since $P(\omega') \not\subseteq P$. Hence he considers a state possible in which he does not know that β is true. In this sense, he is not sure that he knows that β is true. That is, he knows that β is true but does not know that he knows this fact.

The final property is called *knowing that you don't know* or *negative introspection*. It says that if $\omega' \in P(\omega)$, then $P(\omega) \subseteq P(\omega')$. To understand this property intuitively, suppose it does not hold. Again, suppose ω is the true state and that $\omega' \in P(\omega)$. Let $P' = P(\omega')$ and let β denote the statement that the state is in P'. Suppose $P(\omega) \not\subseteq P'$. Then the agent does not know whether β is true since some of the states he considers possible are not in P'. However, in one of the states he considers possible, namely ω', he would know that the true state is in P' and hence would know that β is true. Hence he does not know that β is true but considers it possible that he does know that β is true. In this sense, he does not know that he does not know that β is true.

It is easy to see that if all three properties hold, then the $P(\omega)$ sets form a partition of Ω. Hence partitional information is a special case. As noted earlier, nonpartitional models allow the possibility that the agent does not understand his own information processing and the properties above should

be interpreted in this light. Negative introspection is generally seen as a less realistic assumption than positive introspection. As Geanokoplos [1989] notes, it is the fact that people don't typically notice nonoccurrence that enabled Sherlock Holmes to surprise Dr. Watson by pointing out the importance of the dog that *didn't* bark in the night.

The decision theory Geanakoplos proposes assumes that if the agent's set of possibilities is P, then he chooses his action to maximize $E_\omega[u(a, \omega) \mid \omega \in P]$. As before, this expectation is computed by updating some prior on ω using Bayes' Rule to concentrate the distribution on $\omega \in P$.[10] In this case, the optimization *cannot* be described as choosing an optimal f which is measurable in the usual sense. As a result, the agent's view of his actions *ex ante* and after receiving his information may be quite different.

To see this point clearly, consider the following example, drawn from Geanakoplos [1992]. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and assume that the agent's prior, q, has $q(\omega_1) = q(\omega_3) = 2/7$ and $q(\omega_2) = 3/7$. Let $A = \{a_1, a_2\}$ and let u be given by the following table:

	ω_1	ω_2	ω_3
a_1	-1	1	-1
a_2	0	0	0

Let $P(\omega_1) = \{\omega_1, \omega_2\}$, $P(\omega_2) = \{\omega_2\}$, and $P(\omega_3) = \{\omega_2, \omega_3\}$. (This possibility correspondence satisfies nondelusion and positive introspection, but not negative introspection.) Note that given either $P(\omega_1)$ or $P(\omega_3)$, the updated probability on state ω_2 is $3/5$. Hence a_1 is optimal given either possibility set. Obviously, a_1 is optimal when the possibility set is $P(\omega_2)$. Hence the optimal f would appear to be $f(\omega) = a_1$ for all ω. However,

---

[10] One could take a more general approach, allowing any probability distribution with P as its support. See Morris [1992a] or Morris and Shin [1992].

notice that the *ex ante* expected payoff to this f is $-1/7$, while the *ex ante* expected payoff to always choosing a_2 is 0.

There are two odd aspects to this example. First, the agent's optimal plan of action *ex ante* and his chosen behavior in response to information differ. This kind of behavior is referred to as *dynamic inconsistency*.[11] Second, it is particularly striking that this conflict is so severe that, *ex ante*, the agent would prefer not to get information at all! Geanakoplos [1989] gives necessary and sufficient conditions for this problem to not arise.[12] However, he does not resolve the dynamic inconsistency problem.

There have been a few applications of nonpartitional information structures. Geanakoplos [1989, 1992], Samet [1990], and Rubinstein and Wolinsky [1990] consider under what conditions certain standard results with partitional information go through with nonpartitional structures (particularly, Aumann's [1976] theorem on agreeing to disagree and Milgrom and Stokey's [1982] no–trade theorem). Also, Brandenburger, Dekel, and Geanakoplos [1992] analyze correlated equilibria with these information structures, generalizing the approach of Aumann [1987]. Finally, Morris [1992a] provides a nonpartitional generalization of Blackwell's [1951] theorem on comparing the value of information partitions.

Aside from Morris' [1992a] work comparing the value of information structures, there has been very little work on deriving optimal nonpartitional information processing.[13] An example due to Robinson [1992] sug-

---

[11] The dynamic inconsistency problem often arises outside the context of standard expected utility. In fact, there are a variety of results which show that dynamic consistency essentially requires expected utility. See, for example, Epstein and Le Breton [1992] or Machina [1989].

[12] See Morris and Shin [1992] for a generalization and interpretation of this result.

[13] Shin [1988] does derive nonpartitional information processing, but not using any op-

gests that, in fact, agents who can choose between partitional and nonpartitional information will often prefer the latter. To see the intuition, simply note that much of what an agent learns will be irrelevant. If it costs anything to remember what is irrelevant, the agent will prefer to forget. To be more concrete, Robinson's example has the following structure. Suppose there are two states, ω_1 and ω_2, and two actions, a_1 and a_2. Suppose that a_i is optimal in state ω_i, $i = 1, 2$. Suppose that the agent's prior has the property that in the absence of any information, he will choose action a_1. Clearly, then, the information partition $\{\{\omega_1\}, \{\omega_2\}\}$ is no more useful to the agent than the nonpartitional possibility correspondence $P(\omega_1) = \Omega$ and $P(\omega_2) = \{\omega_2\}$. Assuming that it is even slightly more costly for the agent to process input ω_1 enough to recognize it perfectly, he will always prefer the nonpartitional structure to the partitional one. Note also that there would seem to be no conflict at all between the agent's *ex ante* plans and his actual response to information in this example. In general, though, the dynamic consistency problem is likely to greatly complicate the analysis of optimal nonpartitional information processing.

There is one conceptual problem which may pose a serious obstacle to endogenizing the amount of information processing in a nonpartitional model. Recall that these models are based on the assumption that the agent only imperfectly understands his information processing. But if the agent chooses the properties of his information processing, how can he not understand how he processes? In at least some contexts, this problem does not seem to arise. Certainly, in Robinson's example, there seems to be no difficulty with interpretting $P(\omega_1) = \Omega$ as saying that the agent processes the information ω_1 correctly (*i.e.*, recognizes that the state is ω_1), but then forgets this information. It hardly seems odd to suppose that the agent correctly anticipates the possibility that he will forget some fact.

---

timality considerations. He shows that if one interprets knowledge as provability in propositional logic, then the implied possibility correspondence is nonpartitional.

24

V. Axiomatic Approaches to Information Processing.

So far, I have not addressed the question of how any of the approaches above might be derived axiomatically. In the absence of an axiomatic treatment, we lack a systematic way to evaluate the assumptions being made in the different approaches. With an axiomatic derivation, we can, in principle, answer the question "what does it mean to assume that an agent behaves this way?" In this section, I describe three recent papers which analyze information processing from an axiomatic perspective.

Remark 4. A different advantage of an axiomatic approach is that it seems to avoid the criticism discussed in Remark 3 that bounded rationality is being modeled by assuming rationality. Roughly, the axiomatic approach begins with a description of the agent and then translates this into a model of information processing. Clearly, it then makes no sense to ask whether the agent can carry out this information processing accurately. If the processing is simply a representation of what the agent is doing, the question boils down to asking whether an agent is able to do whatever it is that he does!

The three papers discussed in this section have one fundamental similarity. Traditionally, axiomatic decision theory takes a single preference relation and derives some kind of utility representation for it. Along the way, a notion of conditional preferences — that is, preferences conditional on information — are derived, generally based on assumptions about information processing. All three papers reverse this process: they begin with a set of conditional preferences, one for each piece of information the agent might receive. Then a representation is derived which includes a representation of information processing.

A. Nonpartitional Information.

Morris [1992b] provides an axiomatic approach to nonpartitional infor-

mation structures. The idea is to identify what states the agent considers possible by looking at his preferences over state–contingent choices. Loosely speaking, if he does not care what he gets in state ω, then it is as if he knew that the true state is not ω.

More formally, following Savage [1954], let X be a set of *consequences*. A consequence is intended to be a complete description of the payoff–relevant results of a given choice. An *act* is a function from Ω to X. In other words, an act or action is defined in terms of the relationship between external states and consequences it induces. The set of actions A for this model is the set of all possible acts. If $a \in A$, then, $a(\omega)$ is the consequence that action a yields in state ω. For each $\omega \in \Omega$, the agent has a preference ordering $\succ_\omega$ over A. The interpretation of this is that $\succ_\omega$ is the way the agent ranks the acts given whatever information the agent derives from processing the input ω.

Savage defines a $\succ_\omega$–*null state* as a state ω' such that for every pair of acts a_1 and a_2 with $a_1(\omega'') = a_2(\omega'')$ for all $\omega'' \neq \omega'$, we have $a_1 \sim_\omega a_2$. ($\sim_\omega$ is the indifference relation associated with $\succ_\omega$.) That is, ω' is $\succ_\omega$–null if the agent does not care what happens to him in that state — only the consequences for other states matter. Morris then defines a possibility correspondence by letting $P(\omega)$ denote the set of ω' such that ω' is not $\succ_\omega$–null.

With this framework, Morris relates conditions on preferences to the conditions on the possibility correspondence discussed in Section IV. For example, Morris defines a property he calls *static coherence*. To understand this property, suppose that the agent is able to choose one act from a finite set $D \subseteq A$. For each input ω, we could ask which act in D is optimal given the preferences $\succ_\omega$. Intuitively, this would tell us how the agent will choose in response to information. Let a_ω^D denote the agent's optimal choice in response to input ω. Define a new act $\tilde{a}^D$ by setting $\tilde{a}^D(\omega) = a_\omega^D(\omega)$. Intuitively, this is the act that the agent is effectively choosing by

the way he responds to information. Oversimplifying slightly, the agent's preferences are said to satisfy static coherence if for every D and every ω, the agent weakly prefers $\tilde{a}^D$ to every act in D. Morris shows that static coherence (together with another condition) implies positive introspection and a weakening of nondelusion.

He also discusses some other ways to define the possibility correspondence from preferences and relationships among these different approaches. Finally, he discusses changes in the agent's information by considering preferences indexed by time as well as by ω. He uses this approach to address the dynamic consistency issue. Morris does not provide axioms generating the particular decision theory proposed by Geanakoplos [1989] or the weaker version proposed by Morris [1992a].

B. Impossible Possible Worlds.

In Lipman [1992], I make a distinction between the external state and the input. So I have a separate set of inputs or signals, S. To define the "correct" inference, I assume that there is a function $\xi : \Omega \to 2^S$ where $\xi(\omega)$ is the set of signals that are "true" at ω. For any input s, let $\Omega(s)$ denote the set of ω such that $s \in \xi(\omega)$. In other words, $\Omega(s) = \xi^{-1}(s)$. For simplicity, suppose that $\Omega(s) \neq \emptyset$ for all $s \in S$. If $\Omega(s) = \Omega(s')$ — that is, if s and s' contain the same information about Ω — then I will say that s and s' are *logically equivalent*. Like Morris, I assume that there is a set of consequences X and let A denote the set of acts. For each $s \in S$, there is a preference ordering $\succ_s$ over A which gives the agent's preferences in response to input s.

The representation I derive is what I call an *extended expected utility representation*. The "expected utility" part refers to the fact that I derive a utility function over consequences and a probability distribution over states such that one act is preferred to another iff it yields higher expected utility. The "extended" part refers to the fact that I do this not with the state set

Ω but with an enlarged state set Ω^*. The way the state set is enlarged is itself a part of the representation. The "states" in Ω^* which are not in Ω are referred to as *impossible possible worlds*. They are impossible in the sense that only the original states are really possible; however, they are possible in the sense that the agent does not recognize that they are impossible.

In particular, this kind of representation allows the possibility that s and s' are logically equivalent and yet $\succ_s \neq \succ_{s'}$ — in words, the agent does not recognize logical equivalence. Normally, we treat two logically equivalent pieces of information as equivalent to the agent, even though this is clearly not consistent with reality. Generally, we learn something when we are shown that two mathematical statements are equivalent, so we ourselves do not immediately recognize logical equivalence!

Roughly, the way the impossible possible worlds are constructed is the following. Say that s and s' are *preference equivalent* if $\succ_s = \succ_{s'}$ and are *strongly equivalent* if they are both logically equivalent and preference equivalent. When s and s' are strongly equivalent, I require either both or neither to be true in each of the new states. That is, I maintain their logical equivalence in the new states. If they are logically equivalent but not preference equivalent, then I must add in new states in which one input is true and the other is false. The trick is to add these new states in such a way as to maintain the right relationships.

The impossible possible worlds can be thought of as a representation of the "logic" in the agent's mind. To see the idea, let $\Omega^*(s)$ denote the set of $\omega \in \Omega^*$ such that s is true. Then when $\Omega^*(s) \subseteq \Omega^*(s')$, the agent effectively infers that s' is true when he learns that s is true. That is, he deduces s' from s. Thus, for example, an agent who satisfies $\Omega^*(s) \subseteq \Omega^*(s')$ whenever $\Omega(s) \subseteq \Omega(s')$ carries out all appropriate logical deductions. In a similar manner, one can define a variety of notions regarding what kind of deductions the agent carries out. Since the impossible possible worlds are derived from the agent's preferences, this means that the agent's logic is

derived from his preferences.

To understand the relationship to Morris' work, consider the following simple example. Suppose there are three states, ω_1, ω_2, and ω_3. Consider the possibility correspondence $P(\omega_1) = P(\omega_2) = \{\omega_1, \omega_2\}$ and $P(\omega_3) = \{\omega_2, \omega_3\}$. This nonpartitional information structure would be derived by Morris if ω_3 is the only $\succ_{\omega_1}$–null state and the only $\succ_{\omega_2}$–null state, while ω_1 is the only $\succ_{\omega_3}$–null state. In my model, this situation is represented differently. Let s be an input which is true only in states ω_1 and ω_2 and let s' be an input which is true only in state ω_3. To model the agent above, I would introduce a new state, say ω^*, which is identical to ω_2 except that s' is true there. Then when the agent receives input s, he infers that the true state is in $\{\omega_1, \omega_2\}$. When he receives the input s', he infers that the true state is in $\{\omega^*, \omega_3\}$. More generally, the nonpartitional structure is replaced by a partitional structure where some of the worlds are impossible possible worlds which "mimic" certain of the normal states.

C. Case–Based Reasoning.

Gilboa and Schmeidler [1992b] provide a completely different approach to axiomatically deriving a form of information processing. They derive what they refer to as *case–based reasoning*, named after an approach used in artificial intelligence. Loosely speaking, the idea in artificial intelligence, as put forth by Riesbeck and Schank [1989], is that people reason by analogy to past experience. They remember "similar" situations in the past and assume that what worked in a similar situation is likely to work again.

More formally, a *case* is a tuple (p, a, r) where p is a *problem* (an element of some set P), a is an action (an element of A), and r is a *result* (a real number). Let C denote the set of possible cases. A *memory* is a finite set of cases, each of which involves a different problem. In the terminology used here, one can think of the memory together with the current problem as serving the role of the input or external state. For this reason, I will

use ω to denote a particular specification of the memory and the current problem.

As in Morris [1992b] and Lipman [1992], the key is to analyze how preferences vary with the input. Oversimplifying a bit, for each input ω, there is a preference relation $\succ_\omega$ over A. They give some axioms on the preferences such that they can be represented in the following manner. There is a function $\psi : P \times P \to [0, 1]$ where $\psi(p, p')$ is interpreted as describing the similarity between problems p and p'. Given an input ω which specifies current problem p, the agent prefers action a_1 to action a_2 iff

$$\sum_{p'} \psi(p, p') a_1(p') > \sum_{p'} \psi(p, p') a_2(p')$$

where the sum is taken over problems p' in the memory. The function $a_i(p')$ is the result of choosing action a_i in problem p' according to the memory if in fact this is the action that was chosen. Otherwise, $a_i(p') = 0$. Intuitively, then, the agent evaluates each action by looking through his memory to times when the action in question was used and computing a weighted sum of the results obtained where the weights are the similarity of the current problem to these past problems.

To get a sense of how this representation works, consider the following simple example. Suppose that $P = \{1, 2, \ldots\}$ and $A = \{a_1, a_2, \ldots, a_n\}$. Suppose that $\psi(p, p') = 1$ for all p and p' — that is, all problems are equally similar to one another. When the agent faces the first problem, his memory is the empty set since he has not faced any problems in the past. Gilboa and Schmeidler treat the sum over an empty set as equal to 0. Hence we have

$$\sum_{p'} \psi(1, p') a_i(p') = 0$$

for all i, so the agent is indifferent between all actions. For simplicity, suppose that, when indifferent, he adopts the tie–breaking rule of choosing the lowest numbered action. So he would choose a_1. If the payoff to a_1 is strictly positive, then when the agent faces problem 2, he will choose a_1

since

$$\sum_{p'} \psi(2,p')a_1(p') = \psi(1,2)a_1(1) = a_1(1) > 0,$$

while the "expected payoff" to any other action is zero. He will continue choosing a_1 until the payoff accumulated so far falls below zero. Note that it is the *cumulative* payoff, not the *average* payoff that is relevant with this representation.

Once the accumulated payoff to a_1 falls below zero, the agent switches to action a_2, continuing with it as long as its accumulated payoff exceeds zero, and so on. Once he has chosen every action at least once, then in each period he will choose whatever action has the highest cumulative payoff to date.

Though I include this paper in the section on axiomatic approaches, Gilboa and Schmeidler seem to downplay the axioms. They appear to view the axioms they use as more of a technical device and to be more interested in exploring this particular representation. Certainly, this form of information processing is quite different from others in the literature. Also, as Gilboa and Schmeidler note and as the example suggests, it has the distinct advantage of having a natural dynamic component to it. That is, the behavior of the agent over time is inherently a crucial part of the model. For this reason, it may have especially interesting applications in dynamic models.

VI. Concluding Remarks.

It seems almost pointless to conclude by suggesting areas for future research — there is so much left to be done that almost everything is a possibility for future research! Nevertheless, I will comment briefly on some directions that seem especially promising.

First, I will not even attempt to claim to be unbiased when I say that axiomatic approaches seem to be a very important direction to pursue. As

argued above, it is only with an axiomatic approach that we can understand our assumptions on the way the agent processes information.

On the other hand, the axiomatic approach is unlikely to find new ideas on how to model information processing. It certainly seems much more plausible that this approach is only useful after interesting models are found and then as a way of gaining understanding about the models. I believe that the development and exploration of other approaches to modeling the costs or constraints on information processing is still very important. The most promising direction for developing other models would seem to be following the lead of the threshold models and network models in considering more details of how information is processed. Certainly, it is unlikely that a general specification of a set of feasible partitions and a cost function on partitions will lead to very concrete results. While these two specific formulations may not prove to be the most useful in the end, the general idea of filling in details this way seems very promising. Perhaps some blending of nonpartitional approaches with this kind of more detailed look at information processing will be fruitful.

Finally, I believe it is very important to try applying these models to real economic problems. So far, there has been only a small amount of work in this direction. I believe that we will only really understand these models when we see what they say in the context of real economic problems, rather than simple and artificial examples.

REFERENCES

Abreu, D., and A. Rubinstein, "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, November 1988, 1259–1282.

Anderlini, Luca, "Communication, Computability, and Common Interest Games," St John's College, Cambridge working paper, 1991.

Anderlini, L., and L. Felli, "Incomplete Written Contracts," London School of Economics working paper, 1992.

Aumann, R., "Agreeing to Disagree," *Annals of Statistics*, 4, 1976, 1236–1239.

Aumann, R., "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morganstern*, Zurich: Bibliographisches Institut, 1981.

Aumann, R., "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, January 1987, 1–18.

Aumann, R., and A. Brandenburger, "Epistemic Conditions for Nash Equilibrium," Harvard Business School working paper, 1991.

Bacharach, M., "Some Extensions to a Claim of Aumann in an Axiomatic Model of Knowledge," *Journal of Economic Theory*, 37, 1985, 167–190.

Banks, J., and R. Sundaram, "Repeated Games, Finite Automata, and Complexity," *Games and Economic Behavior*, 2, June 1990, 97–117.

Baumol, W., and R. Quandt, "Rules of Thumb and Optimally Imperfect Decisions," *American Economic Review*, 54, 1964, 23–46.

Bernheim, B. D., "Rationalizable Strategic Behavior," *Econometrica*, 52, November 1984, 1007–1028.

Binmore, K., "Modeling Rational Players: Parts I and II," *Economics and Philosophy*, 3, 1987; 4, 1988.

Binmore, K., "DeBayesing Game Theory," University of Michigan working paper, 1991.

Blackwell, D., "The Comparison of Experiments," in *Proceedings, Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, 93–102.

Brandenburger, A., E. Dekel, and J. Geanakoplos, "Correlated Equilibrium with Generalized Information Structures," *Games and Economic*

Behavior, 4, April 1992, 182–201.

Chellas, B., *Modal Logic: An Introduction*, Cambridge: Cambridge University Press, 1980.

Cho, I.-K., and A. Matsui, "Induction and Bounded Rationality in Repeated Games," University of Pennsylvania working paper, 1992a.

Cho, I.-K., and A. Matsui, "Learning and the Ramsey Policy," University of Pennsylvania working paper, 1992b.

Conlisk, J., "Optimization Cost," *Journal of Economic Behavior and Organization*, 9, 1988, 213–228.

Cutland, N., *Computability: An Introduction to Recursive Function Theory*, Cambridge: Cambridge University Press, 1980.

Dow, J., "Search Decisions with Limited Memory," *Review of Economic Studies*, 58, January 1991, 1–14.

Epstein, L., and M. Le Breton, "Dynamically Consistent Beliefs Must be Bayesian," University of Toronto working paper, 1992.

Geanakoplos, J., "Game Theory without Partitions, and Applications to Speculation and Consensus," Yale University working paper, 1989.

Geanakoplos, J., "Common Knowledge," in Y. Moses, ed., *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fourth Conference*, San Mateo: Morgan Kaufmann Publishers, 1992, 254–315.

Gilboa, I., and D. Schmeidler, "Additive Representations of Non–Additive Measures and the Choquet Integral," Northwestern University working paper, 1992a.

Gilboa, I., and D. Schmeidler, "Case–Based Decision Theory," Northwestern University working paper, 1992b.

Hopcroft, J., and J. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Reading: Addison–Wesley, 1979.

Hughes, G. E., and M. J. Cresswell, *An Introduction to Modal Logic*, London: Methuen and Co., 1968.

Kalai, E., and W. Stanford, "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, March 1988, 397–410.

Kalai, Ehud, "Artificial Decisions and Strategic Complexity in Repeated Games," in *Essays in Game Theory*, New York: Academic Press, 1991.

Lipman, B., and S. Srivastava, "Informational Requirements and Strategic

Complexity in Repeated Games," *Games and Economic Behavior*, 2, September 1990, 273–290.

Lipman, B., "How to Decide How to Decide How to ... : Modeling Limited Rationality," *Econometrica*, 59, July 1991, 1105–1125.

Lipman, B., "Decision Theory with Impossible Possible Worlds," Queen's University working paper, preliminary draft, 1992.

Machina, M., "Dynamic Consistency and Non–Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Literature*, 27, 1989, 1622–1668.

Milgrom, P., and N. Stokey, "Information, Trade, and Common Knowledge," *Journal of Economic Theory*, 26, 1982, 17–27.

Minsky, M., and S. Papert, *Perceptrons*, Cambridge: MIT Press, 1988.

Morris, S., "Revising Beliefs and Knowledge: A Generalization of Blackwell's Theorem," University of Pennsylvania working paper, 1992a.

Morris, S., "Revising Knowledge: A Decision Theoretic Approach," University of Pennsylvania working paper, 1992b.

Morris, S., and H. Shin, "Noisy Bayes Updating and the Value of Information," University of Pennsylvania working paper, 1992.

Mount, K., and S. Reiter, "A Model of Computing with Human Agents," Northwestern University working paper, 1990.

Neme, A., and L. Quintas, "Equilibrium of Repeated Games with Cost of Implementation," *Journal of Economic Theory*, 58, October 1992, 105–109.

Neyman, A., "Bounded Rationality Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma Game," *Economic Letters*, 19, 1985, 227–229.

Pearce, D., "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, November 1984, 1029–1050.

Piccione, M., "Finite Automata Equilibria with Discounting," *Journal of Economic Theory*, 56, February 1992, 180–193.

Piccione, M., and A. Rubinstein, "Finite Automata Play a Repeated Extensive Game," University of British Columbia working paper, 1992.

Radner, R., "The Organization of Decentralized Information Processing," AT&T Bell Laboratories working paper, 1989.

Radner, R., "Hierarchy: The Economics of Managing," *Journal of Economic Literature*, 30, September 1992, 1382–1415.

Radner, R., and T. Van Zandt, "Information Processing and Returns to Scale," *Annales d'Economie et de Statistique*, 25/26, June 1992, 265–298.

Riesbeck, C. K., and R. C. Schank, *Inside Case Based Reasoning*, Hillsdale: Lawrence Erlbaum Associates, 1989.

Robinson, L., "The Rationality of Non–Rational Information," Columbia University working paper, 1992.

Rosenthal, R., "Rules of Thumb in Games," Boston University working paper, 1991a.

Rosenthal, R., "Bargaining Rules of Thumb," Boston University working paper, 1991b.

Rubinstein, A., "Finite Automata Play the Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 39, 1986, 83–96.

Rubinstein, A., "Similarity and Decision–making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)," *Journal of Economic Theory*, 46, October 1988, 145–153.

Rubinstein, A., "On Price Recognition and Computational Complexity in A Monopolistic Model," University of Tel Aviv working paper, 1992.

Rubinstein, A., and A. Wolinsky, "On the Logic of 'Agreeing to Disagree' Type Results," *Journal of Economic Theory*, 51, June 1990, 184–193.

Samet, D., "Ignoring Ignorance and Agreeing to Disagree," *Journal of Economic Theory*, 52, October 1990, 190–207.

Savage, L. J., *The Foundations of Statistics*, New York: Wiley, 1954.

Segal, U., "Two–Stage Lotteries without the Reduction Axiom," *Econometrica*, 58, March 1990, 349–377.

Selten, R., "The Chain–Store Paradox," *Theory and Decision*, 9, 1978.

Shafer, G., *A Mathematical Theory of Evidence*, Princeton: Princeton University Press, 1976.

Shin, H., "Logical Structure of Common Knowledge," Nuffield College, Oxford, working paper, 1988.

Sileo, P., "Intransitivity of Indifference, Strong Monotonicity, and the Endowment Effect," Carnegie Mellon University working paper, 1992.

Simon, H., "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 1955, pp. 99–118.

Simon, H., "From Substantive to Procedural Rationality," in S. J. Latsis, ed., *Method and Appraisal in Economics*, Cambridge University Press, 1976, pp. 129–148.

Simon, H., "Bounded Rationality," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave*, New York: W. W. Norton, 1987, 15–18.

Spear, S., "Learning Rational Expectations Under Computability Constraints," *Econometrica*, 57, July 1989, 889–910.

Suppes, P., D. Krantz, R. D. Luce, and A. Tversky, *Foundations of Measurement, Volume II*, San Diego: Academic Press, 1989.

Turnbull, S., "The Firm as a Team of Automata," Ohio State University working paper, 1992.

Van Zandt, T., "Efficient Parallel Addition," AT&T Bell Laboratories working paper, 1990.

Walley, Peter, *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall, 1991.