Boundedly *versus* Procedurally Rational Expectations

Scott Moss
Centre for Policy Modeling
Manchester Metropolitan University
Manchester, M1 3GH, UK.
s.moss@mmu.ac.uk

and

Esther-Mirjam Sent
Department of Economics
University of Notre Dame
Notre Dame, IN 46556, USA.
sent.2@nd.edu

I don't see the point of not assuming that agents know the correct model of the economy. This is not economics.[1]

Some economists who have relied on the rational expectations hypothesis are now seeking to demonstrate that rational expectations equilibria can emerge in models with agents who are artificially intelligent. They typically model agents' intelligence through the use of genetic algorithms. However, these algorithms misrepresent current understanding of human cognition as well as well-known and long-standing evidence from business history and the history of technology. This paper implements a well-validated representation of human cognition in SDML, a logic-based programming language that is optimised for representations of interactions among agents. Within that software environment, a model of a transition economy is developed with three production sectors and a household sector. The numerical outputs from that model are broadly in accord with the statistical evidence from the Russian economy. The model itself is developed explicitly to incorporate qualitatively specified characteristics of entrepreneurial behaviour in that economy. Unlike conventional economic models, transactions are negotiated and effected explicitly — there are no unspecified or under-specified "markets".

0

[1] The whole, wonderfully revealing, report by a *Journal of Economic Dynamics and Control* reviewer on [13].

1 Boundedly rational expectations

Adaptive expectations models contain an asymmetry in the sense that in those models econometricians, who are presumed to be fully knowledgeable, forecast better than agents, who are presumed to rely on mechanistic backward-looking extrapolative rules. One explanation for the rise of rational expectations economics is that it eliminates this asymmetry by placing econometricians and agents on an equal footing (see [30] for further explanations). Hence, rational and consistent expectations in macroeconometric models are often justified on the grounds that it would be wrong to assume that econometric modelers were smarter than the agents who make the decisions that generate the time-series data used to specify and estimate the model. If those agents are really as smart as the econometricians, and if the econometricians' model of the economy is correct, then the agents will also have specified the correct model of the economy and it will be the same as that of the econometricians.

One consequence of this line of reasoning is that, if all agents know the correct (econometricians') model of the economy, they therefore have the same model. By using this model to form their expectations, they must all have the same expectations and there is nothing analytically to distinguish one agent from another. The point can be put in two ways. Either there is a single representative agent or agents are homogeneous. This consequence is not only logical, but also desirable from the perspective of rational expectations economics. The reason is that many rational expectations models seek to derive macro results from general equilibrium models. Yet, the Sonnenschein-Mantel-Debreu (see [9], [15], and [34]) result illustrates the precarious nature of this endeavour. It shows that under standard assumptions on the individual consumers, such as strict convexity and monotonicity of preferences, so that each agent is characterised by textbook indifference curves and a positive bundle of endowments of all goods, we can derive an excess demand curve for each individual. Summing over all individuals, of whom it is assumed that there are only a finite number, gives the excess demand curve for society as a whole. Under certain not-very-restrictive conditions, three properties will carry over from the individual's excess demand curve to the aggregate demand curve: continuity, a value of total excess demand equal to zero at all prices, and excess demand being homogeneous of degree zero. However, these three properties turn out to be the only properties that carry over from the individual to the aggregate demand function. In particular, the Weak Axiom of Revealed Preference (WARP) may not be satisfied at the aggregate level. Yet, if we are to obtain uniqueness and stability of equilibria, some such restrictions must be imposed. Now, if the behaviour of the economy could be represented as that of a representative agent or a number of identical agents, the situation might be saved, since textbook individual excess demand functions do have unique and stable equilibria (see [12] and [30] for further discussions of this).

One rational expectations economist who has openly worried about such consequences of the rational expectations assumption is Thomas Sargent. For him, one undesirable consequence is the no-trade theorem for rational expectations models, which indicates that even if homogenous agents have different information, they will still not be willing to trade with each other. Consider a purely speculative market, meaning that the aggregate monetary gain is zero and that insurance plays no role. When it is common knowledge that traders are risk averse, rational, have the same prior, and that the market clears, then it is also common knowledge that a trader's expectation of monetary gain, based on given information, must be positive in order for that trader to be willing to trade. In other words, if one agent has information that induces willingness to trade at the current asset price, then other rational agents would be unwilling to trade with that agent, because they realise that that agent must have superior information. The equilibrium market price fully reveals everybody's private information at zero trades for all traders. According to Sargent ([28], p. 113): “The remarkable no-trade outcome works the rational expectations hypothesis very hard.”

Sargent further discovered that rational expectations models are unable to place agents and econometricians on an equal footing. When implemented numerically or econometrically, rational expectations models impute more knowledge to the agents within the model (who use the equilibrium probability distributions in evaluating their Euler equations) than is possessed by an econometrician (who faces estimation and inference problems that the agents in the model have somehow solved). The reason is that close scrutiny of the justification of error terms reveals that identification and estimation of the models require the econometrician to know less than the agents. Whereas agents' decision rules are exact (non-stochastic) functions of the information they possess, the econometrician must resort to some device to convert the exact equations delivered by economic theory into inexact (stochastic) equations susceptible to econometric analysis Sargent ([28], p. 21) therefore acknowledges: “The idea of rational expectations is ... sometimes said to embody the idea that economists and the agents they are modeling should be placed on an equal footing.... These ways of explaining things are suggestive, but misleading.”

In an attempt to solve these problems, Sargent has joined the ranks of economists who are using algorithms from artificial intelligence and artificial life programs to represent heterogeneous agents. Sargent ([28], pp. 2-3) justifies this move as follows:

> I interpret a proposal to build models with boundedly `rational agents' as a call to retreat from the second piece of rational expectations (mutual consistence of expectations) by expelling rational agents from our model environments and replacing them with `artificially intelligent' agents who behave like econometricians. These `econometricians' theorize, estimate and adapt in attempting to learn about probability distributions which, under rational expectations, they already know.

In particular, Sargent seeks to demonstrate that it is not necessary to impose mutual consistency of expectations on all agents in order to get simulation results that approximate the rational expectations outcomes. He suggests that, in general, rational expectations equilibria can emerge from computational models in which “heterogeneous” agents develop models endogenously. Therefore, Sargent remains unwilling to relinquish rational expectations completely. Instead, he tries to reinforce rational expectations by focusing on convergence to this equilibrium ([16], p. 140; [28], p. 133). Furthermore, he seeks to solve some of the problems associated with rational expectations such as multiple equilibria and no-trade theorems ([28], pp. 25, 133-134). Moreover, he desires to deal with some discrepancies in rational expectations ([28], p. 27). Finally, he hopes algorithms from artificial intelligence and artificial life could assist in the computation of equilibria ([28], pp. 106, 152).

Though inspired by the algorithms from artificial intelligence and artificial life developed by researchers associated with the Santa Fe institute, Sargent adopted a restricted interpretation of these algorithms as a result of his reluctance to relinquish expectations. Rather than using algorithms to think about populations as endorsed by many Santa Fe scientists, Sargent saw them as models of the neurons of an individual’s brain ([28], p. 76). Rather than relinquishing the notion of an equilibrium as advocated by many Santa Fe researchers, he focused on convergence to equilibrium ([28], p. 153; [17], p. 372).

There is nothing in the assumption of heterogeneous agents or artificial intelligence or artificial life to make convergence to equilibrium an obvious or, even more emphatically, a necessary outcome. Palmer, Arthur, Holland, LeBaron, and Tayler [24], for example, did not find convergence to rational expectations in their first model of an artificial stock market. They conjectured that in on-off situations that are never going to happen again or in complicated situations where the agents have to do a lot of computing, establishing convergence to rational expectations requires loading almost impossible conditions onto these agents. Whereas prices in Sargent’s experiments fluctuate around the fundamental value, Palmer et al. observe speculative bubbles and crashes. Whereas markets in Sargent’s simulations settle down to a stationary state, Palmer et al. find that their markets do not do this.

In a later paper [3], however, the same authors found that essentially the same model will converge to rational expectations equilibrium provided that learning takes place at a rate which is sufficiently slow. Faster rates of learning generate a market regime in which psychological behaviour emerges, there are significant deviations from the rational expectations benchmark, and statistical "signatures" of real financial markets are observed.

Darley and Kaufman [8] find two similar regimes in dynamic non-cooperative games in which agents learn from more or less local interaction with their neighbours. States resembling rational expectations equilibria (with shared agent perceptions) arise when prediction is easy (because each agent looks at the behaviour of a small number of neighbours so there is not much feedback) but not when prediction is difficult (because each agent notices the behaviour of a lot of other agents and, so there is a lot of feedback). When the model does not generate rational expectations equilibria marketed by mutually consistent models or perceptions, then the output is meta-stable in the sense that it is marked by periods of stasis interspersed with periods of turbulence.

Slow learning and local learning seemingly can support convergence of mutually consistent expectations. Both entail limited inputs of information to the learning process. This result is consistent with the findings of Moss and his colleagues in a number of simulation studies ([19], for example) that local learning in a stable and relatively simple environment leads to agents to hold very similar or identical models of the environment and appropriate actions for achieving their aspirations. However, environments can sometimes become too complicated for the agents' conventional models to provide guides to action and then some exceptional means must be found to limit and focus the information presented to decision-makers.

We have now seen that the use of a particular range of artificial intelligence algorithms, those associated with Holland classifiers, can be used to represent agents who learn to specify mutually consistent mental models of their environments but who need not do so. In order to choose among these models, we require some criteria. The criteria of Sargent and his colleagues appears to be convergence to rational expectations equilibrium. We infer this criterion from Sargent's own methods. In several published works, Sargent ([17]; [28]) tweaks the standard classifiers to establish optimisation and convergence. For example, he imposes the same classifier systems on every agent of a given type. Hence, despite the claim that artificial intelligence algorithms allow the incorporation of heterogeneous agents, Sargent's models retain a high degree of homogeneity. Moreover, instead of adopting classifiers to model markets, he simulates single agents by the interaction of two autonomous classifiers. This leaves him with the awkward corollary that his agents make virtual bids (to themselves?) and transactions (with themselves?) to gain control of their own actions. Finally, rather than using genetic algorithms to provide the environmental flexibility that other artificial intelligence techniques lack, Sargent diminishes their role. In fact, Marimon, McGrattan, and Sargent ([17], p. 358n) acknowledge: "More standard Holland algorithms were first tried without much success, prompting us to produce the modified algorithm."

We conclude that the representations of cognition used by Sargent are arbitrary in the sense that there is no reason independent of his own models to devise and use such representations. They are certainly motivated by a wish to give his models an appearance of plausibility but there is no effort to demonstrate either that these representations are descriptively accurate or that the models in which they are incorporated yield outputs which are empirically verified. The only criterion seems to be that the model outputs should conform to the predictions of rational-expectations theory which, Sargent admits, rests on representations of cognition that are implausible in the extreme.

These analyses are sometimes seen as part of a bounded-rationality research program. However, Herbert Simon, the inventor of the notion of bounded rationality, would certainly reject the Sargent view. Sent [29] describes the difference between the Sargent and Simon approaches to representations of cognition as one that turns on the descriptive accuracy of the representation of cognition as behaviour. Whereas Sargent justifies the use of genetic algorithms to represent cognition on their effective parallelism in computation, Simon argues that cognition depends on serial symbol processing. In effect, Sargent appeals to current views of the physiological basis of all mental activity while Simon appeals to experimental evidence about the epiphenomena of decision-making and learning as observed by experimental psychologists. Furthermore, whereas Sargent sees bounded rationality as a means for strengthening mainstream economics, Simon seeks to develop an alternative to mainstream economics through bounded rationality.

Though Simon was the originator of the bounded rationality concept, he was not in complete control of it. Recognising this, Simon wrote in a letter to Sargent on 11 July 1995: “I could complain and say: ‘I invented it and have a right to decide how it should be defined,’ but as I failed to apply for trademark rights, I guess I have no standing in court.” However, we would like to present some arguments to the jury in the remained of this paper. Our defence is based on the criticism that Sargent-type results are generated by computational models with unknown analytical properties. Moreover, the learning procedures assumed for agents are arbitrary in the sense that there is no reason to believe that actual decision makers learn in a manner that is described by those assumptions. Building on this criticism, our defence further presents an alternative approach.

The difference between the approaches outlined in this introduction and our proposal is important for economists if either (a) the two approaches imply different theoretical and modeling structures yielding different relationships between actions such as policy measures, on the one hand, and the consequences of those actions, on the other, or (b) one approach more usefully supports policy analyses than does the other. In the remainder of this paper, we consider only the second of these criteria.

2 Procedurally rational expectations

Although the phrase "bounded rationality" was originated by Herbert Simon, it has been taken over by more conventional economists and redefined to cohere with mainstream economic theory. For Simon, bounded rationality involves limited information-processing and computational capacities. For the more conventional economist Williamson [36], in contrast, bounded rationality implies limited access to information. The difference is that, for Simon, bounded rationality entails the availability of more information than can be taken into account by decision makers while for economists such as Williamson bounded rationality entails the paucity of information which is therefore a constraint in optimisation procedures. Whereas Williamson is reluctant to accept the notion of satisficing, mainly because he thinks it would denote irrational behaviour, Simon himself considers satisficing to be a direct implication of bounded rationality.

The effect of redefinitions of bounded rationality such as Williamson’s is to leave in tact the underlying precept of mainstream economic theory that agent behaviour can be represented by some constrained optimisation algorithm. The compulsion of economists to adhere to constrained optimisation as the defining characteristic of human behaviour is also manifest in, for example, Sargent's specification of artificially intelligent agents. The genetic algorithms and classifier systems he employs to represent behaviour are different but remain optimising algorithms that arguably misspecify the nature of human cognition.

2.1 Genetic algorithms and classifier systems

Classifier systems are parallel, message-passing, rule-based systems that model their environments by activating appropriate clusters of rules. This structure allows them to model complicated, changing environments, to interpret the internal states of agents in the theory so that the agents seem to progressively "model" their world, to make agents able to build up behavioural repertoires that include chains of actions that are initiated long before the agent obtains the reward, and to make agents able to develop the capacity to plan future actions on the basis of their expectations of what the consequences of those actions will be. In particular, classifier systems have two especially desirable efficiency properties. First, they do not impose heavy memory requirements on the system. Second, much of the information processing can be carried out in parallel.

In classifier systems, the agent is modeled as a collection of basic cognitive units, called classifiers. In contrast to standard economic theory, there are no consistency requirements on the classifiers of which an agent is comprised. In fact, there are some advantages to the notion of an agent as a bundle of possibly inconsistent behavioural propensities. First, requiring consistency imposes great computational costs on the system, as it entails a lot of internal structure and frequent consistency checking amongst different structural components. Second, since the world is always more complicated that our personal experience, maintaining consistency in an agent's behavioural or conceptual system almost necessarily requires a reduction in the agent's range of possible action, in particular in response to novel situations. Finally, evidence seems to suggest that we humans do in fact maintain overlapping and inconsistent conceptual systems and associated behavioural propensities.

An agent that maintains inconsistent behavioural propensities has to have some mechanism that determined on which of these propensities it would actually act. This is where competition enters the stage. There might be more than one winner of the competition at any given time, and as a result a cluster of rules could react to external situations. The competition mechanism in classifier systems depends on a number that is associated with each of the agent's classifiers, its strength, which registers the "memory" of how well the classifier has served in the past in the agent's quest for reward. A classifier's strength is modified over time by one of the system's learning algorithms.

A classifier system adapts or learns through the application of two well-defined machine learning algorithms. The first one, the bucket brigade algorithm, changes classifier strengths by identifying actions that lead to rewards — not just those that produce reward directly, but also those that "set the stage." It changes the strength associated with each classifier with experience in two ways. First, any classifiers whose action is implemented pass some of their strength to their immediate predecessors. Second, the strength of classifiers whose action is implemented when the agent receives an externally specified reward is increased as a function of the reward received. The system can be started off with a set of totally random classifiers. And then, as the environment reinforces certain behaviours and as the bucket brigade does its work, the classifiers organise themselves into coherent sequences that produce at least a semblance of the desired behaviour.

Even if the bucket-brigade credit assignment algorithm works perfectly, it could only rank the rules already present. By itself, it can only lead the system into highly optimised mediocrity. Two mechanisms are required to carry out the operation of replacing old classifiers with new ones. The first determines when replacements take place. It has to recognise situations in which the agent "needs" new classifiers. The second type of mechanism constructs new classifiers, which would probably improve the prospect for the agent to obtain a reward. This is a job for the genetic algorithm, which explores the immense space of possible new classifiers. Genetic algorithms carry out a subtle search for tested, above-average "building blocks" and build new classifiers by combining parts of existing high-strength classifiers. The idea is that useful classifiers work because they are composed of good building blocks, either in the features of the world that trigger them or in the actions they recommend. Trying out new combinations of these building blocks is more likely to produce useful new classifiers than is any kind of random search through the space of possible classifiers.

The genetic algorithm solves the problem of how to set the dial of rationality. The needle can be put at zero initially and then the genetic algorithm decides through mutation and recombination how far up the dial it goes. By adding the genetic algorithm as a third layer on top of the bucket brigade and the basic rule-based system, an adaptive agent not only learns from experience but can also be spontaneous and creative.

Classifier systems have been applied to economics in several ways. Instead of assuming that agents are perfectly rational, they can be modelled with classifier systems and learn from experience like real economic agents. Instead of modeling the economy as a Walrasian general equilibrium, societies of classifier systems can organise a set of interacting economic agents into an economy.

The argument that these implementations of genetic algorithms and classifier systems misspecify human cognition starts from the implication from cognitive science that cognition takes the form of exploitation of what we know and a highly directed exploration of our environment that is focused by our knowledge. This inference follows from the distinction in cognitive science between procedural and declarative knowledge (see, e.g., Anderson [2]).

2.2 Procedural and declarative knowledge

Procedural knowledge is knowledge about how to do something and this knowledge is held by individuals in a way that does not allow it to be communicated directly to other individuals. Declarative knowledge is knowledge of what is true and can be communicated directly to other individuals. For example, an Englishman may have both procedural and declarative knowledge about the game of cricket. He can explain the rules of the game and describe or show a novice how to stand at the wicket or where to stand if he is to play off-stump or what to do if he is the wicket-keeper or the necessity of keeping the bowling arm straight at the elbow. All of this knowledge is declarative. To hit the ball successfully and place it where the batsman wants the ball to land or to spin-bowl so that the ball hits the ground and bounces so as to hit the wicket without coming into the range of the bat require abilities that can only be attained by practice. However well a person might know the rules and be able to describe the practices of cricket, that person will not be able actually to play cricket without acquiring substantial procedural knowledge.

Independently, this same distinction has been made by historians of business, the organisation and technological change who demonstrate its relevance to these areas of economic activity. For example, Edith Penrose [25] calls the two types of knowledge objective and subjective in her seminal analysis of the direction of the growth of the firm. Yet, her differentiation between the two is couched in the same terms as Anderson's [2] discussion of the difference between procedural and declarative knowledge. Similar distinctions — though not quite so explicit as in Penrose — are found in Chandler's (e.g. [6]) work on the development of organisational structures and Rosenberg's ([26], [27]) discussions of the determinants of the direction of technical change.

Since we cannot know everything, a reasonable assumption is that the declarative knowledge we do have comes from the activities in which we engage. How we use this declarative knowledge follows from our experience and, to the extent that experience is necessary to use declarative knowledge effectively, its use is governed by procedural knowledge. In other words, we start from what we know and develop new ideas and perceptions only by extending our experience.

Genetic algorithms and classifier systems, by contrast, search the whole of the environmental information-space randomly and, if well constructed, evenly at the outset and concentrate increasingly on the parts of the information space that yield the best results. In the language of the field, genetic algorithms explore the search space and then exploit the subspaces described by classifiers that yield the greatest fitnesses. For cognitive scientists, human cognition takes the form of exploitation of what we know and a highly directed exploration that is focused by our procedural knowledge.

The difference has significant implications for economics. Either agents are global optimisers in which case genetic algorithms and classifiers can be used to represent that optimisation in conditions of constrained information-processing and computational capacities or they can at best exploit their existing procedural and declarative knowledge in the hope of gaining some local (though possibly large) improvement in their circumstances. Hence, the assumption of global optimisation can evidently support the construction of models with no concern for the procedures by means of which agents actually go about collecting declarative knowledge and then developing their procedural knowledge.

2.3 Implications for representations of cognition

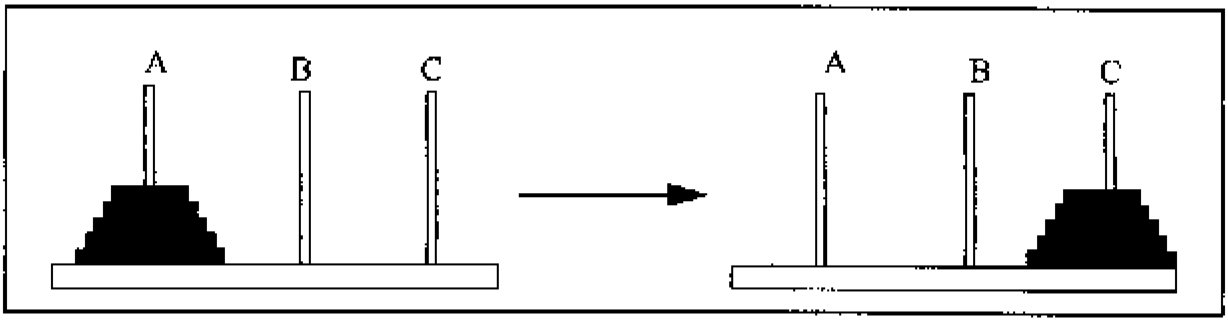
Much of the work in the development of computational cognitive science stems from Simon's work and, in particular, Newell and Simon [23]. The later work in cognitive science, focused as we shall see on computer software architectures to represent cognitive processes, informs the approach taken here.

Indeed, Simon himself is critical of genetic algorithms and classifier systems [33] to represent cognitive processes.

In his distinction between substantive and procedural rationality, Simon [32] stresses the importance of the procedural aspects of cognitive behaviour that are left out by genetic algorithms and classifiers. Procedural rationality concerns the choice or development of procedures for making decisions when the decision maker has effectively limited capacities to process information and calculate appropriate outcomes. Certainly, procedural rationality entails satisficing. Our concern here is to find a representation of satisficing that uses artificial intelligence and supports models of decision making in a macroeconomic environment.

The particular representation reported in this paper is drawn from several cognitive theories that have been implemented as computer software architectures designed to replicate data from psychological experiments. These architectures are Soar [14] and ACT-R [2]. Both of these architectures are based on the concept of a problem-space architecture that itself is a tree structure of goals and subgoals. The original specification of this goal and subgoal structure was developed by Newell and Simon [23] as a planning algorithm. The sort of situation in which it might be used in the Newell-Simon version was planning a trip from an office at MIT to an office at Berkeley. If the goal were to make that journey, a subgoal would be to fly from the nearest airport to MIT to the nearest airport to Berkeley. The subgoal of making that flight would be to get from the MIT office to the airport, which would be undertaken by (say) car or taxi. To take the car would entail the subgoal of getting from the MIT office to the car by walking.

A classic problem on which to test artificial-intelligence algorithms is the Tower of Hanoi problem. This involves moving a set of discs of graded size from one peg to another, using a third peg as an intermediate step. The five-disc Tower of Hanoi problem is illustrated below. The discs can be moved one at a time and it is not permitted to place a larger on a smaller disc. The problem-space architecture for this problem, as specified by Simon [31] is to specify a subgoal of getting the top four discs onto peg B so that the largest disk and be placed on peg C and then to execute the next subgoal of moving the four discs on peg B to peg C. That move entails a subgoal of moving the top three discs to peg A so that the remaining disc cam be placed on the largest disc that is already on peg C. There is then a similar subgoal to get the three-disc tower onto peg A, and so on. Anderson [2] developed a program in ACT-R to learn to solve the Tower of Hanoi problem and compared the results of that program with



the results of experiments with human subjects. He found that the students did indeed learn to use a goal stack in the same way as ACT-R. The actual movements of the discs and the setting of goals and subgoals were accomplished in ACT-R by production (if-then) rules. Figure 1

Figure 1. The Tower of Hanoi Problem

Three points about the ACT-R representation of cognition are relevant here. The first is that the results obtained from ACT-R programs can be compared with the results of psychological experiments to verify the accuracy of a program as a representation of cognitive behaviour in particular circumstances. Secondly, ACT-R is an encoding of an underlying theory of cognition. Thirdly, the representation of the problem-space architecture as rules for moving up and down the goal tree and the rules for performing tasks to achieve each goal can, in principle, be obtained by the standard knowledge-elicitation techniques used for building expert systems.

Taking the first two of these points together, we have a means of encoding procedural knowledge about how agents learn that is informed and justified by a particular theoretical structure and discipline that is independent of the domain of application in economics or the management sciences. Discussions or arguments about the appropriateness of that particular encoding are not likely to be influenced by the results desired for economics models. The third point allows us to develop independent evidence to support a particular encoding of agents' procedural and declarative knowledge.

As mentioned before, we believe that the difference between models such as Sargent's and ours is important if one approach more usefully supports policy analyses than does the other. Therefore, we will report a pilot model of a transition economy in the remainder of this paper. This model will be used to investigate the characteristics of procedures for learning and decision making that are validated in relation to cognitive science and verified in relation to economic time-series data. These procedures take for granted bounded rationality in the sense of Simon. These limitations preclude the assumption of optimising behaviour. Encoding the process of goal formation, learning and declarative knowledge about the environment in a manner that corresponds to encodings in the cognitive sciences, we are able to determine whether procedures for forming expectations and perceptions about the environment are rational in the sense that action based on those perceptions is increasingly likely or, in any case, not less likely to further the attainment of agents' goals.

3 An emerging-market model

The model reported here was developed to capture certain stylized descriptions of the commercial environment faced by newly- privatised and state-owned enterprises in the Russian Federation. One such fact of particular importance is the sharp increase in inter-enterprise arrears that was not being repaid during the first half of 1992 — an episode known as "the arrears crisis". As shown by Alfandari [1], the volume of arrears peaked in the summer of 1992 at some 23 per cent of Russian GNP. The volume of arrears soon subsided to an average of about 5% for the period after 1992. Some experts find this figure to be not out of line with what is a normal amount of overdue trade credit by international standards [1]. A straightforward statistical comparison, however, can be misleading because enterprise arrears in Russia represent a different type of economic relation. In many cases the credit is forced, bears negative real interest and has no fixed repayment period or agreed repayment schedule. To a degree, these peculiar features were conditioned by the shortage of working capital, a drastic fall in demand and other economic consequences of a government's attempt to put an end to the regime of soft budget constraints by lifting price control and removing state subsidies to producers.

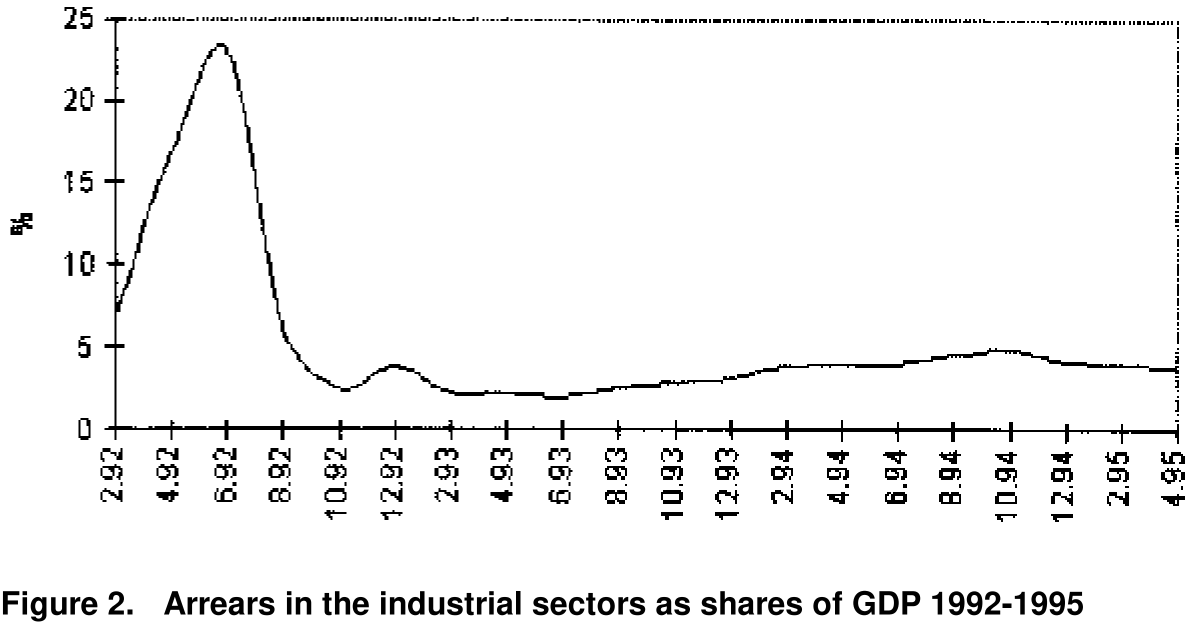


Figure 2. Arrears in the industrial sectors as shares of GDP 1992-1995

Source: Finansy promyshlennosti Rossii, The Monthly Bulletin of the Working Centre on Economic Reform of the Government of the Russian Federation, no., June 1995, p. 2.

Moss and Kuznetsova [21] argue that the evolution of the arrears crisis provides a clear example of how enterprises cope with situations characterised by significant uncertainty. In the Russian case, enterprises were forced to adopt a survival strategy giving priority to existing, recognised constraints. There was no possibility to maximise anything in the framework of those constraints [4]. Indeed, the scale of the accumulated debt and its persistence suggest that debt reduction was not a high priority. For one thing, both the liquid assets of enterprises and their debts have been growing simultaneously [5]. The debt became an element of enterprises' survival strategy and was instrumental in prolonging the existence of a business environment to which they were accustomed, i.e., one governed by soft budget-constraints. Because the accumulated bad debt grew out of proportion on the national scale and became commonplace in all industrial sectors, this important business indicator ceased to be seen as a symptom of poor management efficiency. By the autumn of 1992, 95% of enterprises had bad debts enough to be proclaimed bankrupt on legal grounds [37]. Debt performance had become separated from the business performance of a firm.

The Russian arrears crisis provides us with sufficiently clear stylized facts that we can use to assess whether the outputs from simulations conform to those facts or not. The stylized facts we want to capture, in addition to the arrears increases, are a high average and widely fluctuating rate of price inflation and volatile but trendless outputs. The point is to capture these stylized facts using a credible and validated specification of the processes agents use to develop their own models of their environments.

3.1 The modelling language

The model reported here was implemented in SDML, a strictly declarative modeling language that corresponds to strongly grounded autoepistemic logic (SGAL) (cf. [11], [19]). This means that any model that runs under SDML is formally sound and consistent with respect to the axioms and rules of inference of SGAL. Consequently, models written in SDML can entail qualitative as well as numerical relationships without loss of formal clarity and rigour. This is an important issue that is taken up in some detail in Section 4.1 below.

SDML has a number of object-oriented features that make it particularly useful for modelling cognition along the lines described in Section 2. The particular object-oriented features supporting the model reported here are the type (or class) hierarchy and the container hierarchy.

The type hierarchy is similar to the class hierarchy of, say, C++ but whereas C++ has simple inheritance (each class inherits the methods and instances of one superclass), SDML has multiple inheritance. The basic inheritance class is reproduced as Figure 3 from Moss *et. al.* [19].

The user adds further subtypes, in particular, subtypes of Object, Agent and SDML's predefined Agent subtypes. The type Agent is distinguished from Object in that it has rulebases associated with it. The number of such rulebases varies with the time levels defined by the user. Time levels are discussed in several contexts below.

Agent is the principal type of interest here. Models are specified in terms of instances of agents but these will not normally be instances of the type Agent but of a user-defined subtype of Agent or one of its predefined subtypes (or of more than one of these). Clause definitions and rules are specified in types and are inherited from them by their instances. In this way, the rules for a number of identical agents can be defined in a shared type. Similarly, agents who are not identical may nevertheless share certain rules by means of a common supertype.

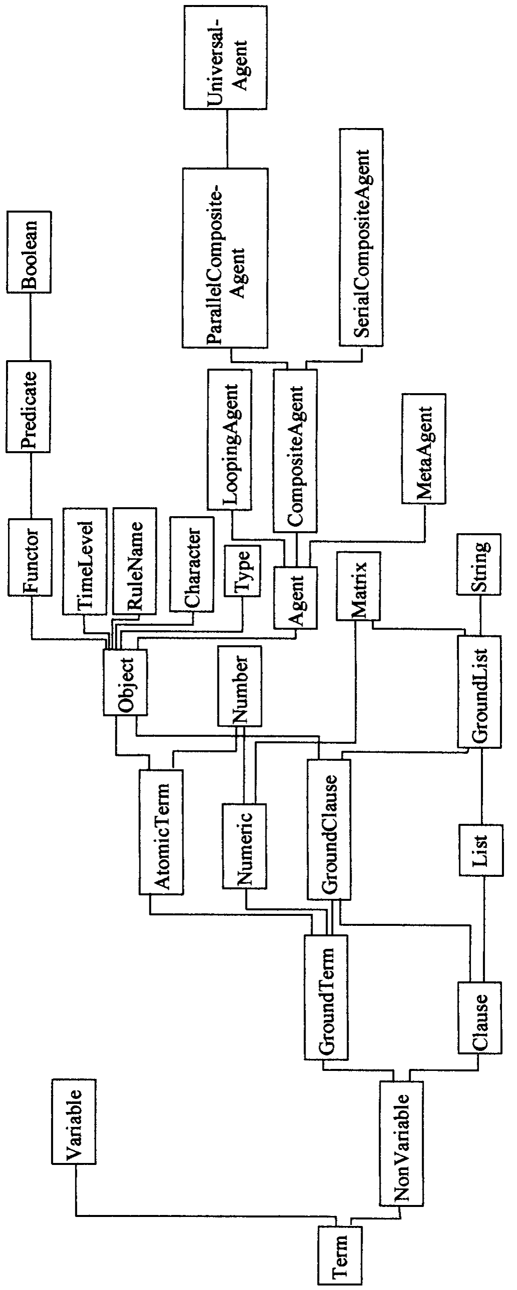


Figure 3. The Basic SDML Type Hierarchy

Abstract supertypes such as ParallelCompositeAgent or LoopingAgent add particular functionality to agents. The instances of every subtype of CompositeAgent can contain other agents (its subagents). Instances of subtypes of ParallelCompositeAgent contain subagents the rulebases of whom fire in parallel. Instances inheriting from type SerialCompositeAgent fire their rulebases in a previously specified order. LoopingAgents loop over time periods and there can be an arbitrary number of such time levels.

Finally, any agent can contain an instance of a subtype of type MetaAgent. Meta-agents can assert statements to, and retrieve statements from, their containers' rulebases in the same way that all agents assert to and retrieve from databases. The main difference is that meta-agents can only write rules to and read rules from their containers' rulebases whereas they and all other agents can read and write statements conforming to any previously defined clause definition when these statements are held on databases. Meta-agents are used to devise the sort of agent routines discussed by Nelson and Winter [22] but to do so as a result of some representation of cognitive activity. These routines represent procedures that are the best the agents and their meta agents have so far found in their attempt to meet their current aspirations.

We will use the type and container hierarchies to put together a model in which enterprises learn about each other in the course of finding enterprise routines that support their goals and aspirations.

3.2 Model structure

The type hierarchy defined for the Russian transition model is reproduced in Figure 4 and the container hierarchy in Figure 5 and Figure 6. The type symbols in boldface in Figure 4 are user defined.

All of the common features of agents that engage in transactions are implemented in the TradingAgent type. In the main, type TradingAgent defines clauses that are used to effect transactions. These clauses include "orderPlacedWith <ProductGroup> <TradingAgent> <Number>" and "orderReceivedFrom <ProductGroup> <TradingAgent> <Number>". The symbol <Type> indicates an instance of that type. If a household (say household-18) decides to purchase 12 units of corn from a farm (say farm-3), it would write "orderReceivedFrom corn household-18 12" to the database of farm-3. At the same time, it would write to its own database "orderPlacedWith corn farm-312". Similarly, if farm-3 decided to accept the order, it would write to the database of household-18 "purchasedFrom corn farm-3 1.305 12" where the price is 1.395 and the quantity 12 units of corn. In order to remember the sale itself, farm-3 would write to its own database "saleTo corn household-181.305 12". In this way the transaction would have been proposed and agreed by all of the different types of agent using the same language. However, households and enterprises behave differently in a number of ways in the model so that those clause definitions and rules that govern the behaviour of each of these subtypes of trading agents separately are implemented in their respective types.

In order to understand the nature of the subtypes of CompositeAgent and of LoopingAgent, it is necessary also to consider the container hierarchy of the model. The outermost container is debtArrearsModel, an instance of type ArrearsInflationModel. The type ArrearsInflationModel is a subtype of LoopingAgent from which it inherits all of the rules and clause definitions required to enable the model to loop over the time levels defined for instances of ArrearsInflationModel in the computational model. Type ArrearsInflationModel is also a subtype of type SerialCompositeAgent from which it inherits the rules and clause definitions to support subagents who fire their rules in a sequence determined by the agent. In this case, the agent environment fires its rules before the agent economy. The agent environment fires its rules once at each date. It is a simple agent the function of which is to introduce representations of environmental changes such as natural disasters into the simulations.

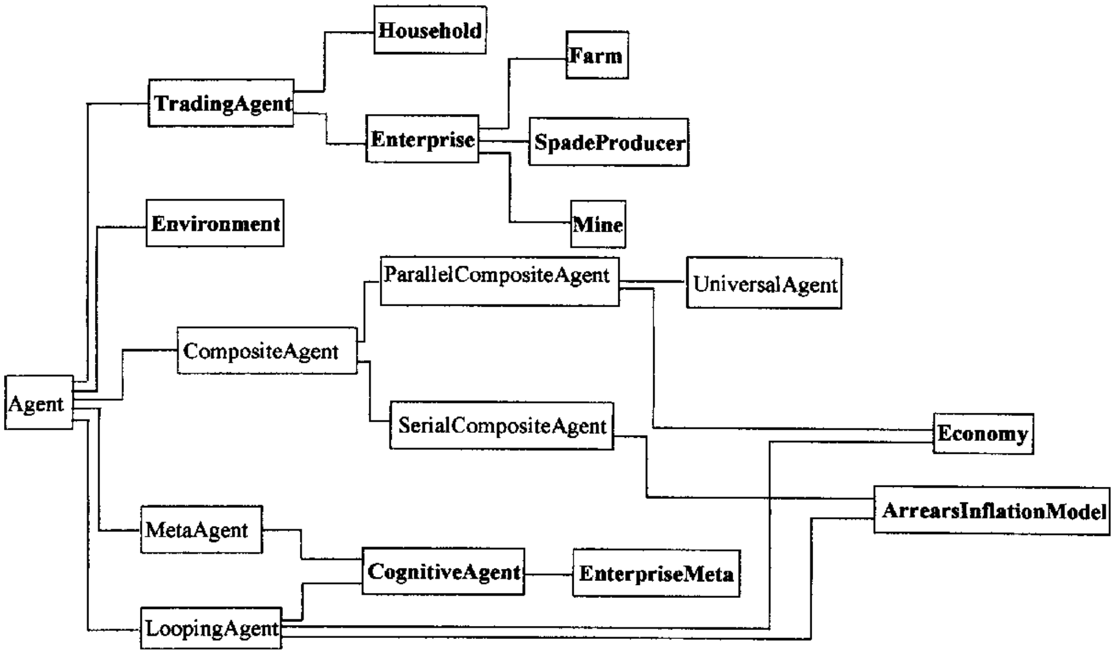


Figure 4. Hierarchy of agent types

The agent economy is an instance of type Economy that itself is a subtype of ParallelCompositeAgent and of LoopingAgent. As an instance of ParallelCompositeAgent, it contains agents who fire their rules effectively in parallel. This is an important feature in a model that represents transactions as the outcomes of communication among agents and where communication is represented by the assertion of clauses to the databases of other agents or some other database that is common to both. Since the agents are active at the same time, we have to represent the fact that a communication can be received only after it is sent. This means, in terms of any computational model, that an agent can retrieve any message written by another agent only at a time period subsequent to the period in which it was written. So, in order to effect a transaction, agent-1 will assert an order for goods to the database of agent-2 at time t; agent-2 will retrieve that assertion at time t+1 and assert its acceptance of the offer to the database of agent-1. This acceptance will be retrieved by agent-1 and time t+2 and the agreement to transact has been concluded.

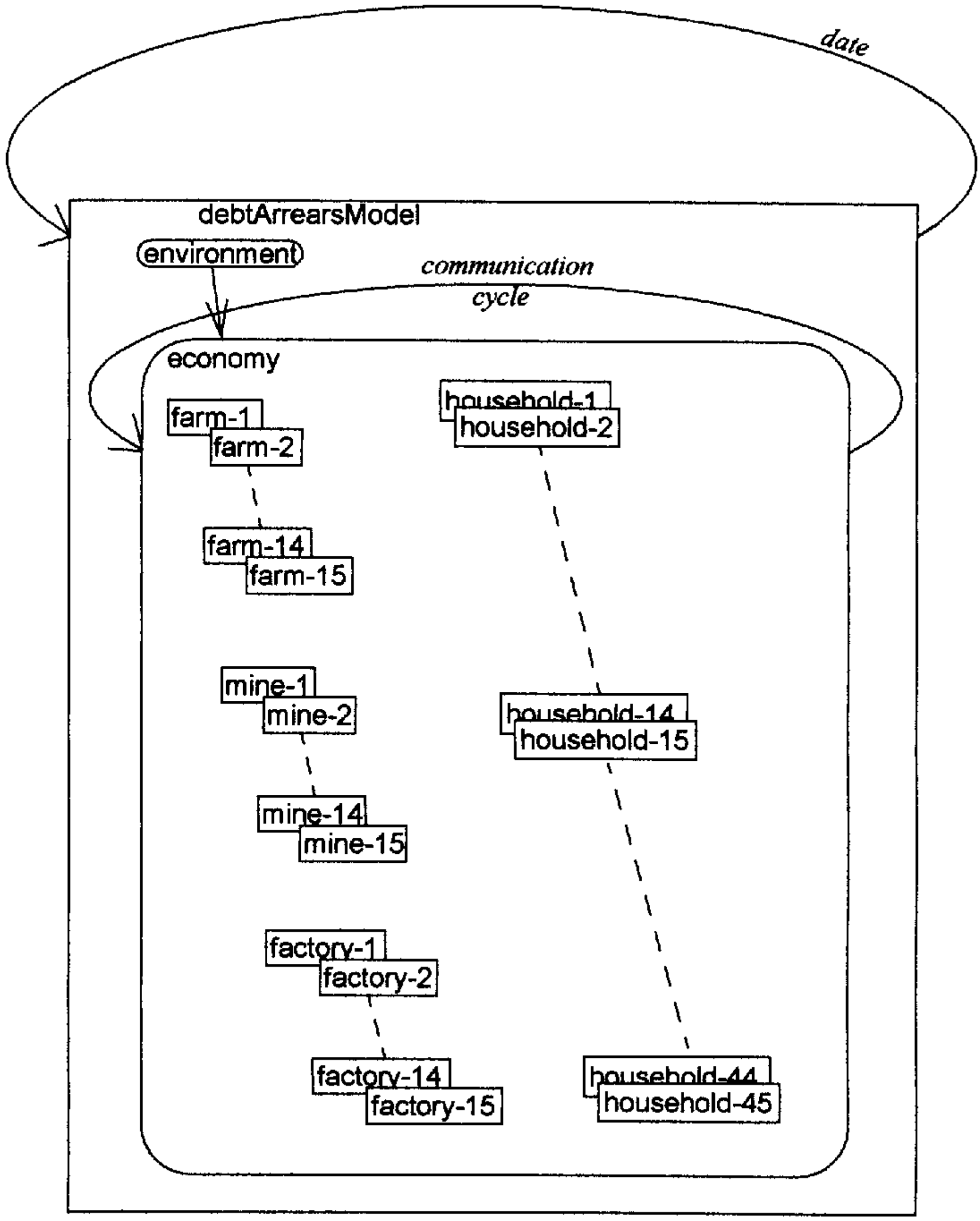


Figure 5. The macro container structure

In this model, transactions take place each date but the negotiations required to effect them we now see will take more than a single time period. For this reason, several communicationCycles are allowed each date. The clauses shared by instances of TradingAgent include, for example, purchasedFrom and soldTo, each of which has arguments to identify the product, the price, the amount and the trading partner.

Instances of type Enterprise represent cognitive agents. In this model, cognition takes the form of the building of mental models that are used to create rules for guiding the decisions of the enterprise. These decision rules are what Nelson and Winter [22] call "routines".

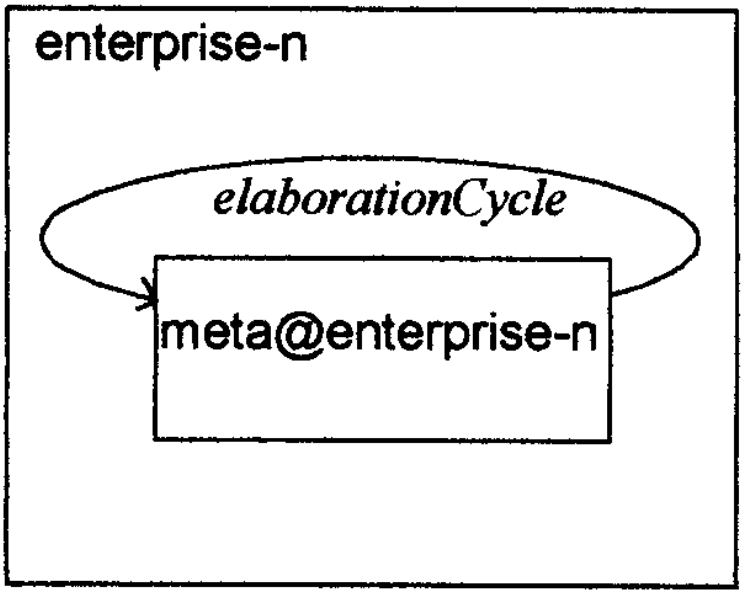


Figure 6. The enterprise container structure

The building and assessment of the mental models are undertaken by subagents of the enterprises. These subagents are instances of type EnterpriseMeta that is a subtype of MetaAgent. Meta-agents can use the rulebases of their containers and other subagents of their containers as databases. In this model, each instance inheriting from type Enterprise contains a meta-agent of type EnterpriseMeta. All instances of EnterpriseMeta build mental models in a procedure derived from computational cognitive science. This involves identifying goals and subgoals and the tasks needed to achieve those goals over a number of time periods called elaboration cycles. One of the tasks to be completed in this process is the writing of decision rules to the rulebase of the containing enterprise. For this reason, type EnterpriseMeta is itself a subtype of LoopingAgent as well as MetaAgent and it loops over time level elaborationCycle. The articulation of instance of type Enterprise, expanded from Figure 5, is given in Figure 6.

3.3 The model setup

In order to capture the main, essential elements of the position of Russian enterprises, we assume the decision variables of the enterprises to be planned output, output price, the wage rate, the offer of employment, orders placed with suppliers and payments to creditors. In addition, the enterprises choose the suppliers with whom they place their orders and the customers whose orders they will fill in whole or in part. Each enterprise notes at each date whether its suppliers have filled the orders placed with them and whether its customers have paid for the goods previously supplied to them. These notes take the form of endorsements attached to the enterprises records of its customers and suppliers. Orders are allocated among known suppliers in proportion to their records of reliability. Sales are allocated first to orders from known customers with the best records of payment. In effect, each enterprise builds up models of the enterprises with which it trades.

These endorsements are also used by enterprises to formulate views about which other firms are the most successful. It is natural to assume that those suppliers who are best at supplying orders and those customers who pay most quickly are also the strongest enterprises. On this basis, enterprises take into account any observable information they have about these trading partners and assume that their behaviour is highly functional. In the model reported here, the only information that one agent can observe about an enterprise is the output prices it sets, its employment of labour and information arising from the transactions in which they engage (supplies, orders and payments). Thus, if one enterprise observes its best trading partners lowering (or raising) prices, it will assume that to lower (or raise) prices increases the values of goal variables and, so, will conjecture a model to that effect.

The goals of the enterprises are sales volume and cash. There is no attempt at optimisation of these values but, rather, the agents seek strategies that will increase the value of one or the other of these goals. In the present setup, neither is given pride of place. In the event that changing the value of one decision variable is expected to increase the value of one goal value and diminish the other, then the action that is considered most likely to have the anticipated outcome will dominate the decision. If the agent has more confidence that the goal value diminution will occur then she will change the decision variable value to reduce or prevent the diminution. If she has more confidence that the other goal value will be increased, then she will change the value of the decision variable in the appropriate direction.

The pre-defined intermediate variables observed by the enterprises are their own purchases, their own sales, their current stocks of real goods (inputs and unsold outputs), their current financial asset holdings (only cash, so far), and the wage bill (the product of the wage rate paid and employment by the enterprise).

The wage bill is the only intermediate variable to be calculated from other variables observed by the enterprise. It gets this special treatment because we assume that it is paid in the same period as the employment it covers. This assumption itself seems appropriate because of the relatively insignificant incidence of delayed wage payments and also because, in inflationary conditions, the impact of delayed wage payments is much the same as offering a lower wage rate. A rather more elaborate setup would be required directly to capture the effects of wage arrears. If such a model were thought likely to be useful, it would be a straightforward extension of the model reported here.

In general terms, the setup reported here was devised only to capture a coarse-grained account of the development of Russian enterprises and to demonstrate how our modelling techniques perform on problems relating to the emergence of new markets and market institutions.

3.4 Results

Our first simulation setup included a specification of the input-output relations and the inclusion as model variables of inter-enterprise debts as well as payments and the usual economic variables of employment, prices, wage rates, production decisions, actual outputs and sales. The Russian experience is of steady, even rising, employment, rapidly rising prices, growing debt and declining output and sales.

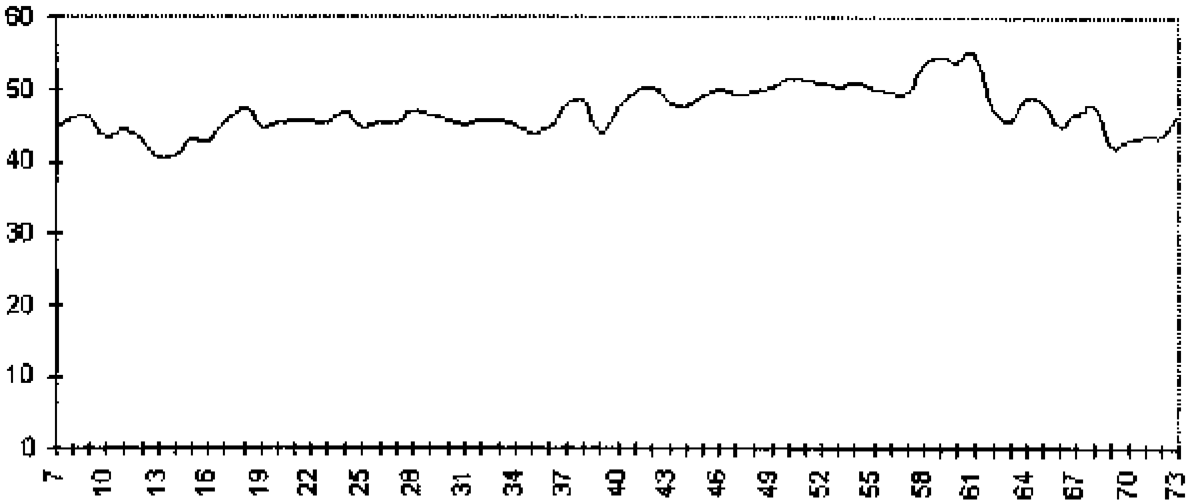


Figure 7. Total employment (simulated)

We have run the setup for 73 periods. What we get is not too dissimilar to what we observe. Employment, for example, does not show any dramatic changes at all although, reflecting variations in production activities over the whole of the simulation, it has not been stable. The time path of employment is given in Figure 7.

In production we observed a collapse after an initial surge (probably due to initial simulation conditions) but starting from date 7 production trends varied from sector to sector. The production of corn remained remarkably stable because corn has at least one relatively stable source of demand in the form of households. The production of both spades and iron shows considerable oscillation but it develops accordingly with the picture of unit sales that demonstrates a certain cyclical pattern. Series for production is shown in Figure 8 and for unit sales in 0

A key result in the development of the simulation setup was the price series. The series, reproduced as Figure 0, now reflects the stylized facts of the Russian experience. However, earlier models generated trendless, though moderately volatile, price series. The achievement of the price series of 0 was obtained by specifying credible information sources.

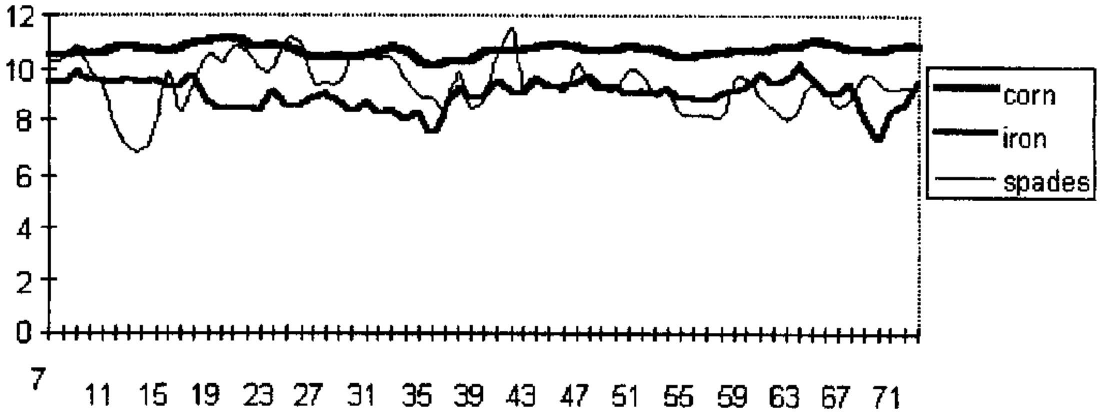


Figure 8. Sectoral production (simulated)

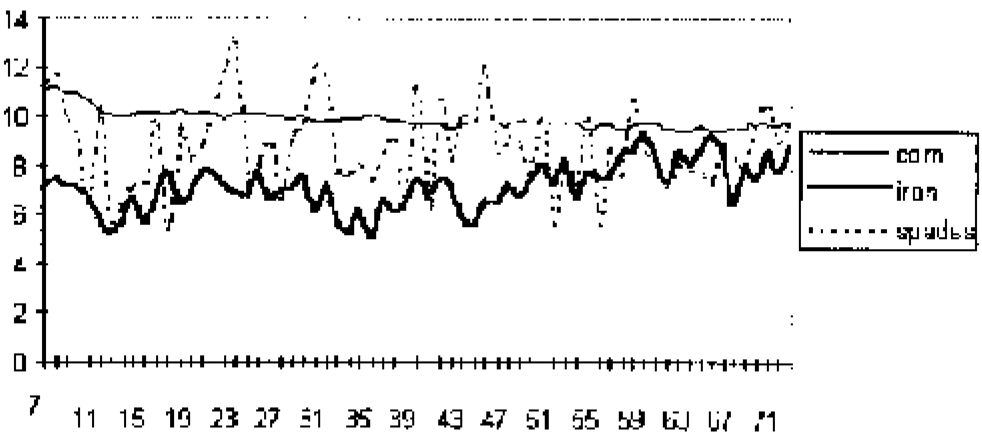


Figure 9. Unit sales (simulated)

3.5 Procedural rationality as modelling

The goals of the enterprises are sales volume and cash. The numerical controls are price, wage rate, employment, input demands and debt payments. In addition, each enterprise has to determine the allocation of its orders for inputs and the allocation of its debt repayments.

Agent cognition is represented by a process of model building that has more in common with the cognitive sciences than with genetic algorithms and classifiers as used by Sargent and others in the economics community. The modelling strategy of each enterprise in the simulation model is to formulate its own models relating selected control variables to at least one goal variable. Initially, following Moss [20]), control variables and goal variables were combined at random and the control variable was either increased or decreased with equal probability. There was a standard generate-and-test procedure so that enterprises kept those of their models that yielded improved goal variable values and abandoned those that did not. Agents kept track of how good their models were by a process of endorsement first suggested by Cohen [7] and implemented by Moss [20]). In general terms, models were endorsed as successful at a date when they were used to formulate an action undertaken in the same period that a goal variable value was improved. They were endorsed as being unsuccessful when the goal value was not improved. Models that were usually successful or usually unsuccessful were endorsed as such. This would happen when four of the five most recent applications of a model earned the successful or unsuccessful endorsements, respectively. Collections of endorsements mapped into numerical values with positive endorsements increasing overall endorsement values and negative endorsements reducing overall endorsement values. The probability of applying a model is proportional to its overall endorsement value. Models with negative overall endorsement values are abandoned.

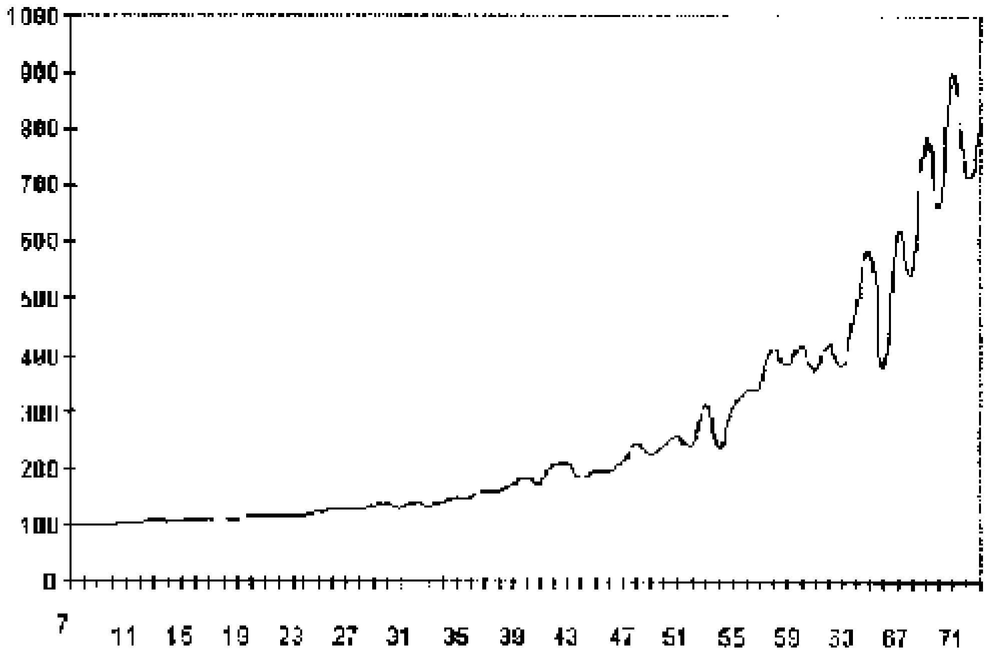


Figure 10. Paasche price index (simulated)

Enterprises also endorsed other agents as being reliable or unreliable suppliers, reliable or unreliable debtors and so on. The resulting overall endorsement values attaching to these other agents were used to determine suppliers and, when orders exceeded stocks of outputs, the order in which to allocate outputs to customers. In this way, stable trading relationships were established among agents in the model.

An important feature of this model was the absence of any *deus ex machina*. There was no auctioneer, no bulletin board on which to post prices, supplies or demands for all agents to see. All information was communicated directly by one agent to another. Learning and expectations about the behaviour of other agents and the system as a whole resulted from an explicit cognitive mechanism that gains its credibility from disciplines outside economics. Procedural rationality entails the selection of information when more is available than an agent can process. It also involves identifying relations from that data without relying on it being correct in some fundamental sense. If the relationships suggest actions that improve on an agent's present position, then they constitute a good model. Otherwise, the model needs to be abandoned or revised. An important means of understanding for humans is the forming of analogies. The analogy in the simulation environment reported here took the form of applying behaviour by agents who appear to be successful to one's own behaviour. If, for example, we know that the farmer to whom we sell our spades pays her bills immediately and is raising the price of corn, then she must be doing well enough to pay her bills so we will try raising the price of our spades in order to do well ourselves.

What did not work (i.e., did not conform to the stylized facts) in the simulation experiments that led to this model was random generation and testing. Classifiers rely on a generate-and-test algorithm, too, but on a much larger scale since a population of models would be generated, mutated, crossed-over and tested at each date rather than just the one model selected by the agent in our simulations. We know that such models can yield outputs that converge towards rational-expectations equilibria. This issue of the relationship between simulation outputs and rational expectations equilibrium is clearly of some importance to economists. It is, therefore, addressed in the following section.

4 Model validation and verification

The validation of a computer program is the process of applying formal methods to ensure that a program design will achieve what is expected of it in appropriate conditions and, in particular, will not get into a confused or illegal state. If a program runs without error in any computer programming language, then that program is consistent and sound relative to that language. That is, the program does not generate or entail mutually contradictory statements and it does not generate statements that the language does not support. Consequently, program validation entails ascertaining that the program is consistent and sound relative to a formal statement of the properties of the programming language.

In this section, we argue that validation should be an important issue in the specification of economic models in general and economic cognition in particular. Two aspects of validation are considered: validation with respect to logical formalisms and validation with respect to cognitive theories.

4.1 Logical validation

The virtue of validating the consistency and soundness of a model relative to a logical formalism is that it removes ambiguity from the specified relationships comprising the model. The particular formalisms that economists rely on are mathematical systems that are well-suited to optimising functions subject to well-specified constraints. Such mathematical bases are inappropriate for the model reported in Section 3 because of its strong reliance on qualitatively defined variables such as endorsements and because there is no element of optimisation in the model. Nonetheless, the model is consistent and sound relative to at least one logical formalism (and, though unproved, probably many such formalisms).

The logical formalism under which the transition model is consistent and sound is a fragment of strongly grounded autoepistemic logic (FOSGAL) [13]. The proof of the consistency and soundness of the model relative to FOSGAL runs as follows:

If a programming language corresponds to a logical formalism, then any program viewed as a set of statements or sentences that runs in that language will necessarily be sound and consistent relative to that logical formalism. One such language, implemented precisely to capture this feature of programs, is SDML which corresponds to FOSGAL. This particular logical formalism of SDML has emerged as one that supports the kind of multi-agent, strictly declarative modelling favoured by the SDML user community. It is by no means the only possible or appropriate logical formalism for modelling organisations. Indeed, the choice of logical formalisms for different classes of problems is already a fruitful field of enquiry in the artificial intelligence literature. A natural extension of the work described in this paper is the explication of the properties of appropriate logics to underpin a language of discourse for the management and economic sciences.

In SDML, and therefore in the model reported here, each agent is defined on a rulebase and database for each period of time. Every rule in the rulebases and every clause asserted to the databases is sound and consistent relative to strongly grounded autoepistemic logic. If any were not, then the model would not run and inconsistency error would be reported.

FOSGAL emerged as a good logical basis for modelling agent behaviour and interaction because of its encoding of negative knowledge. Such encodings are well- recognised to be difficult simply because it is not possible in practice to store all the facts that are not true. For this reason, SDML follows the conventional practice of storing only positive knowledge and dealing with negation by allowing rules that have `not inferred' operators in their antecedents.

SDML was designed principally as a forward-chaining language because drawing inferences from a set of beliefs and facts and then remembering those inferences by asserting them to a database seemed a more natural representation of agent reasoning than backward chaining, in which the implication is stated and then the antecedents evaluated to see if the assertion can be justified. In order to perform forward chaining efficiently, SDML will fire rules but keep track of any assumptions it had to make on the way. This helps to minimise any back-tracking to try alternative assumptions in the more commonly occurring situations. Thus SDML sometimes needs to make inferences from its own lack of inference of certain facts, just as in SGAL one can infer from one's own lack of belief.

4.2 Theoretical validation

The position we have now reached is that a model written in SDML is consistent and sound relative to a particular formal logic that is well-suited to the development of computational models involving interaction among agents that takes the form of direct communication of information. Genetic algorithm and classifier representations of agents are doubtless consistent and sound relative to the mathematics underlying these representations. Hence, we have two, alternative representations of cognition that have equal claim to logical rigour. How then do we choose between them?

It seems likely that the appeal of genetic algorithms to economists such as Sargent is that they are optimising algorithms. However, not all problems are well conditioned to be optimised using genetic algorithms in general and Holland classifiers in particular. In fact, Sargent [17], [28] imposed many restrictions on standard genetic algorithms. Rather than using the standard one-point crossover reproduction mechanism, he included a two-point method of recombination. Rather than adopting classifiers to model markets, he simulated single agents by the interaction of two autonomous classifier systems. Rather than using genetic algorithms to provide the environmental flexibility that traditional artificial intelligence techniques lacked, he diminished their role. Hence, he moved away from standard interpretations of classifier systems in his desire to salvage methodological individualism and neoclassical equilibrium. First, standard genetic algorithm methods differed from methodological individualism in that they simultaneously involved a parallel search involving hundreds or thousands of points in the search space. Second, convergence to a globally suboptimal result was a major concern with genetic algorithms.

We are not aware of any independent argument that Holland classifiers are accurate representations of human cognition. In fact, as discussed before, Simon [33] opposes this interpretation. Hence, their main virtue seems to be that, under some circumstances at least, computational models representing agents as Holland-type classifiers yield outputs that converge over simulated time towards rational expectations equilibria.

To find that a result can be achieved in a variety of different ways and, in particular, using a variety of distinct approaches is a standard means in the natural and mathematical sciences of building confidence in the result. The fact that genetic algorithms can be used to generate rational expectations equilibria in computational models thus enhances confidence in the importance of the rational expectations hypothesis. If the attempt to generate the same result with different techniques sometimes leads to the intended result and sometimes does not, then the failure can itself lead to a deepening of our understanding of the phenomena under investigation. One possibility is that a number of different approaches to the investigation yield a similar set of conditions under which the original result can be expected. In this way, we begin to identify the conditions of application of, in this case, the rational expectations hypothesis.

One such development looks to be emerging from the papers by Arthur *et. al*. [3], [24] and by Darley and Kaufman [8] as cited earlier. These papers follow Sargent in relying on classifiers to represent key aspects of cognition. A complementary approach in the investigation of the conditions in which we might expect rational expectations equilibria to arise is suggested by computational cognitive science. Both Soar and ACT-R offer points of departure for this approach. Although we do not yet have complete results, we can exhibit the key features of and some early results from an implementation of key ideas from cognitive science into the transition model reported above.

The core concept drawn from both Soar and ACT-R (and originally from Newell and Simon [23]) is the problem-space architecture (PSA). The PSA rests on the experimentally verified idea that, in undertaking complex decisions, we engage in a process of sub-goaling. What is involved here is readily seen from the PSA devised for a new version of the model reported in Section 3.

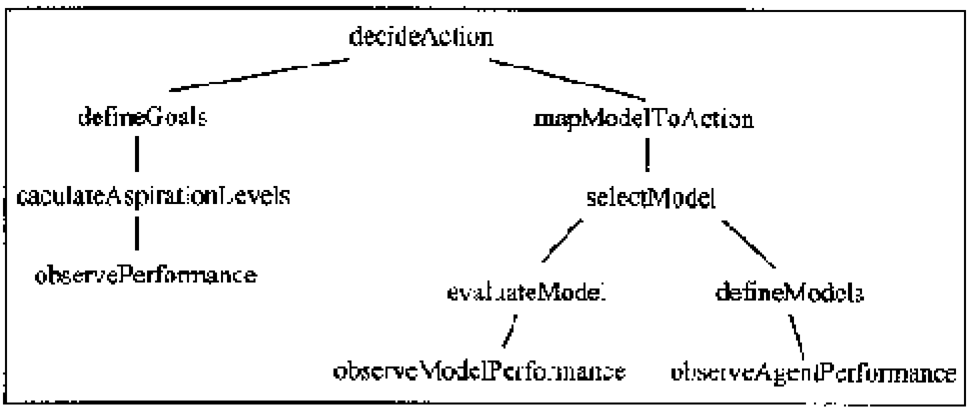


Figure 11. Transition enterprise problem-space architecture

The label of each problem space in Figure 11 indicates a task to be completed. The top-level task (decideAction) that must be completed at each date is to decide what actions to take. Examples of the actions to be decided upon are price levels, wage rate, planned output, input demands and how much of current cash resources to pay out. Actions are predicated upon goals so that, before deciding on an action, the agent must determine what the action is meant to achieve. In this model, aspiration levels are set for the various goal variables and this is done on the basis of the enterprise's observation of its own performance. Once the goals are defined, it is necessary to translate perceptions of the environment into some action. Since these perceptions are represented by mental models of the agents, an appropriate model must be selected. The selection will be from existing models and new models. The existing models are evaluated as they are used and the evaluations are remembered as endorsements. New models are defined on the basis of observations of the behaviour of the most successful (best endorsed) of the enterprise's trading partners. Consequently, the performance of these trading partners must be observed and evaluated.

Declarative knowledge is composed of facts that can be retrieved from one database or another by the agent and some relationships encoded as mental models. Procedural knowledge is composed of the means of mapping models into rules of action. There is also some common knowledge such as the impossibility of increasing outputs without increasing inputs that is declarative in the sense that agents could in principle communicate such knowledge to one another. Such knowledge is encoded as models that cannot be eliminated from the databases of an agent.

Models (apart from the common-knowledge models) and the memory of trading relations can be retrieved by agents with a probability related to their importance and the length of time since they were last retrieved. The scheme used to determine the probability of retrieval at need is taken from Anderson [2]. In particular, the odds in favour of retrieval are

$$\sum_{i=1}^{n} a t_i^d$$

where the t_j are the lags since the jth prior retrieval of the endorsement; d is a positive parameter determining the half-life of the influence of a prior retrieval on current retrieval. The value of a is determined in the simulation model by the endorsement values of models and the agents.

One virtue of this formula is its consistency with experimental data about memory. The choice of the d and a parameters is not determined by that experimental literature. Setting those values is an empirical issue to be considered presently. The precise specification of the PSA is also an empirical issue. In terms of validation, however, it seems much more robust to develop representations of cognition that, like this one, can be assessed on a basis that is independent of the application rather than to assess an arbitrary representation of cognition on the basis of its convergence to a rational expectations equilibrium.

4.3 Verification

While the PSA specified in 4.2 conforms to similar models developed for social simulations (see, for example, [38] or [19]), we have not yet sought independent evidence of its descriptive accuracy. The source of such evidence would naturally come from knowledge-elicitation exercises with enterprise managers in the Russian Federation. Such an exercise would also inform our assumptions about goals and goal-conflict resolution.

Parameterising the model is computationally expensive and requires reliable data series. Previous models produced by the Centre for Policy

Modeling have been parameterised using genetic programming algorithms. These extend genetic algorithms by representing data as tree structures that are subject to mutation and cross-over at points where usable programs result. Edmonds and Moss [10] report this technique in detail. The fitness functions are typically related inversely to root mean-squared errors and are also biased in favour of parameters that look reasonable to domain experts. In this way, we not only get results from application to hold-out data sets that are at least as good as ordinary regression methods applied to the same data but also results that reflect at the same time-domain expertise.

5 Conclusion

A necessary but not sufficient condition for simulation models to converge towards a rational expectations equilibrium is that agent perceptions converge so that mean expectations are close to model outcomes. The Sargent research programme seeks to demonstrate that non-homogeneous agents who search the whole possibility space will converge on perceptions that are effectively consistent with the outputs of the system model. The transition models described in this paper generate convergence in behaviour — specifically in price-setting behaviour — by local exploitation of declarative knowledge. These models were not implemented with rational expectations equilibria in mind. They were implemented to demonstrate that models that are well-validated with respect to both formal logics and independent theoretical structures relating to learning and decision making also lead to verifiably accurate model outputs.

We conjecture that any cognitive behaviour that entails the convergence of individuals' perceptions while these perceptions are changed when systematically wrong will also entail convergence towards expectations that are not wrong in any biased way. If so, then ultimate convergence towards a rational expectations equilibrium or something similar is a weak condition in the absence of structural change in the economy. However, the process of learning by global search and the process of learning by local exploitation might well yield very different results in relation to the effects of economic and social policies that are intended to effect particular structural changes. Before implementing policies based on either of these approaches, it would be important to look for reasons to believe in one or the other. Validation with respect to formalisms — mathematics or predicate or propositional logics — ensures rigour in the sense of a lack of ambiguity in specifications and implementations of models. Validation with respect to independent, experimentally or otherwise verifiable theories gives us confidence that our own models are better than just data-mining. We can see no reason to have confidence in the encoding of human cognition as genetic algorithms and classifiers simply because the resulting models converge to something that is not too different from conventional rational-expectations-type results. This scepticism stems from the absence of any independent reason to believe that agents learn by global search and the plethora of evidence that they learn by local exploitation of declarative knowledge conditioned by their procedural knowledge. Since procedural knowledge can only be acquired by experience, the identification and use of declarative knowledge must always be based on the experience of what the agent has done and this is itself an inherently local processing of declarative knowledge.

Experience indicates that economists are unlikely to be influenced by the very different style of process-centred disciplines such as cognitive science. Nonetheless, we have shown the feasibility of implementing models using concepts that have arisen independently in analyses of business history, the economic history of technical change and in the cognitive sciences to explain with empirical verification how learning and decision making actually take place.

References

[1] Alfandari G., Schaffer M. (1995), "On 'Arrears' in Russia", Paper presented at the Joint Conference of the World Bank and the Ministry of Economy of the Russian Federation, St.Petersburg, June 12-13.

[2] Anderson, J.R. (1993), *Rules of the Mind* (Hillsdale NJ: Lawrence Erlbaum Associates).

[3] Arthur, W.B., J. Holland, B. LeBaron, R. Palmer and P. Tayler (1996), "Asset Pricing under Endogenous Expectations in an Artificial Stock Market", University of Wisconsin-Madison, Social Systems Research Institute 9625.

[4] Aukutsionek

[5] Belianova E. (1995) "Motivation and Behaviour of Russian Enterprises" [in Russian], *Voprosy ekonomiki*, no. 6, pp. 15-21.

[6] Chandler, A. (1975), *The Visible Hand*, (Cambridge MA: Harvard University Press).

[7] Cohen, P.R. (1985),

[8] Darley, V. and S. Kaufman, "Natural Rationality" in W. B. Arthur, et. al. (eds), Economy as an Evolving Complex Evolving System II (Reading, MA: Addison-Wesley).

[9] Debreu, G. (1974), "Excess Demand Functions", *Journal of Mathematical Economics*, 1, pp. 15-23.

[10] Edmonds, B. and S. Moss (1997), MS paper

[11] Edmonds, B., S. Moss and S. Wallis (1996), "Logic, reasoning and a programming language for simulating economic and business processes with artificially intelligent agents" in Ein-Dor, Phillip (ed.) *Artificial Intelligence in Economics and Management* (Boston: Kluwer Academic Publishers), pp. 221-230.

[12] Kirman, A.P. (1992), "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives*, 6, pp. 117-136.

[13] Konolige, K. (1988), "On the relation between default and autoepistemic logic", *Artificial Intelligence*, v. 35, pp. 343-382.

[14] Laird, J.E., A. Newell and P.S. Rosenbloom (1987), "Soar: An architecture for general intelligence", *Artificial Intelligence*, v. 33, pp. 1-64.

[15] Mantel, R.R. (1976), “Homothetic Preferences and Community Excess Demand Functions”, *Journal of Economic Theory*, 12, pp. 197-201.
[16] Marcet, A. and T.J. Sargent (1992), "The Convergence of Vector Autoregressions to Rational Expectations Equilibrium" in Vercelli, A. et al. (eds), *Macroeconomics: A Strategic Survey* (Oxford: Oxford University Press).
[17] Marimon, Ramon, Ellen McGrattan, and Thomas J. Sargent. (1990), “Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents”, *Journal of Economic Dynamics and Control*, v. 14, pp. 329-374.
[18] Moss and Edmonds (1997), "Modelling economic learning as modelling", *Cybernetics and Systems* (forthcoming).
[19] Moss, Gaylard, Wallis and Edmonds (forthcoming), “SDML: A multi-agent language for organisational modelling”, *Computational and Mathematical Organization Theory*.
[20] Moss, Scott (1995), “Control Metaphors in … “, *Computational Economics*, v. x, pp. xxx-xxx.
[21] Moss, Scott and Olga Kuznetsova (1996), “Modelling the Process of Market Emergence”. In Owsinski, J. W. and Nahorski, Z. (eds.), *Modelling and Analysing Economies in Transition*, (Warsaw: MODEST) pp. 125-138.
[22] Nelson, R.. and S. Winter 1982, *An Evolutionary Theory of Economic Change* (Cambridge, MA: Harvard University Press).
[23] Newell, A. and H. Simon (1972), *Human Problem Solving*, (Englewood Cliffs: Prentice-Hall).
[24] Palmer, R.G., W.B. Arthur, J.H. Holland, B. LeBaron, and P. Tayler (1994), “Artificial Economic Life: A Simple Model of a Market”, *Physica D*, v.75, pp. 264-274.
[25] Penrose, E. (1959), *The Theory of the Growth of the Firm* (Oxford: Basil Blackwell).
[26] Rosenberg, N. (1975), *Perspectives on Technology* (Cambridge: Cambridge University Press).
[27] Rosenberg, N. (1980), *Inside the Black Box* (Cambridge: Cambridge University Press).
[28] Sargent, T.J. (1993), *Bounded Rationality in Macroeconomics* (Oxford: Clarendon Press).
[29] Sent, E.-M. (1997) , "Sargent versus Simon: Bounded Rationality Unbound", *Cambridge Journal of Economics*, v. 21, pp. 323-338.
[30] Sent, E.-M. (forthcoming), *The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent’s Achievements* (Cambridge: Cambridge University Press).
[31] Simon, H.A. (1975), "The Functional Equivalent of Problem-Solving Skills", *Cognitive Psychology*, v. 8, pp. 268-288.
[32] Simon, H.A. (1976), “From Substantive to Procedural Rationality”. In Latsis, S.J. (ed.), *Method and Appraisal in Economics*. (Cambridge: Cambridge University Press), pp. 129-148.
[33] Simon, H.A. (1993): “The Human Mind: The Symbolic Level”, *Proceedings of the American Philosophical Society*, v. 137, pp. 638-647.

[34] Sonnenschein, H. (1972), "Market Excess Demand Functions", *Econometrica*, 40, pp. 549-563.
[35] Tambe, M. and P.S. Rosenbloom, (1996) "Architectures for Agents that Track Other Agents in Multi-agent Worlds", *Intelligent Agents, II, Springer Verlag Lecture Notes in Artificial Intelligence* (LNAI 1037).
[36] Williamson, O.E. (1975), *Markets and Hierarchies* (New York: Macmillan).
[37] Yavlinsky G. (1994) "The Lessons of the Economic Reform" [in Russian], *Oktiabr'*, no.9, pp.136-149.
[38] Ye, M. and K.E. Carley (1995), "Radar Soar: Towards an Artificial Organization Composed of Intelligent Agents", *Journal of Mathematical Sociology*, v. 20, pp. 219-246.