Agent–Based Models and Human Subject Experiments*

John Duffy
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260 USA
Tel: 412-648-1733
Fax: 412-648-1793
Email: jduffy@pitt.edu
Web: http://www.pitt.edu/~jduffy/

Revised Final Draft
March 2005

Contents

*Forthcoming in K.L. Judd and L. Tesfatsion (editors), *Handbook of Computational Economics Volume 2: Agent-Based Computational Economics*, Handbooks in Economics Series, North-Holland, 2005. I thank Jasmina Arifovic, Thomas Brenner, Sean Crockett, Cars Hommes, Thomas Riechmann, Shyam Sunder, Leigh Tesfatsion and Utku Ünver for helpful comments on earlier drafts.

Abstract

This chapter examines the relationship between agent-based modeling and economic decision-making experiments with human subjects. Both approaches exploit controlled "laboratory" conditions as a means of isolating the sources of aggregate phenomena. Research findings from laboratory studies of human subject behavior have inspired studies using artificial agents in "computational laboratories" and vice versa. In certain cases, both methods have been used to examine the same phenomenon. The focus of this chapter is on the empirical validity of agent-based modeling approaches in terms of explaining data from human subject experiments. We also point out synergies between the two methodologies that have been exploited as well as promising new possibilities.

Keywords

agent–based modeling, human subject experiments.

JEL classification numbers: B4, C6, C9.

1 Introduction

The advent of fast and cheap computing power has led to the parallel development of two new technologies for doing economic research - the computational and the experimental laboratory. Agent-based modeling using computational laboratories grew out of frustration with the highly centralized, top-down, deductive approach that continues to characterize much of mainstream, neoclassical economic-theorizing.[1] This standard approach favors models where agents do not vary much in their type, beliefs or endowments, and where great effort is devoted to deriving closed-form, analytic solutions and associated comparative static exercises. By contrast, agent-based computational economic (ACE) researchers consider decentralized, dynamic environments with populations of evolving, heterogeneous, boundedly rational agents who interact with one another, typically locally. These models do not usually give rise to closed-form solutions and so results are obtained using simulations. ACE researchers are interested in the aggregate outcomes or norms of behavior that emerge and are sustained over time as the artificial agents make decisions and react to the consequences of those decisions.

Controlled laboratory experimentation with human subjects has a longer history than agent-based modeling as the experimental methodology does not *require* the use of laboratories with networked computers; indeed the experimental methodology predates the development of the personal computer.[2] However, computerization offers several advantages over the "paper–and–pencil" methodology for conducting experiments. These include lower costs, as fewer experimenters are needed, greater accuracy of data collection and greater control of the information and data revealed to subjects. Perhaps most importantly, computerization allows for more replications of an experimental treatment than are possible with paper–and–pencil, and with more replications, experimenters can more accurately assess whether players' behavior changes with experience. For all of these reasons, many human subject experiments are now computerized.

With advances in computing power, the possibility of combining the agent-based computational methodology with the human subject experimental methodology has been explored by a number of researchers, and this combination of methodologies serves as the subject of this survey chapter. Most of the studies combining the two approaches have used the agent-based methodology to understand results obtained from laboratory studies with human subjects; with a few notable exceptions, researchers have not sought to understand findings from agent-based simulations with follow-up experiments involving human subjects. The reasons for this pattern are straightforward. The economic environments explored by experimenters tend to be simpler than those explored by ACE researchers as there are limits to the number of different agent characteristics that one can hope to "induce" in an experimental laboratory and time and budget constraints limit the number of periods or replications of a treatment that can be considered in a human subject experiment; for instance, one has to worry about human subjects becoming bored! As human subject experiments impose more constraints on what a researcher can do than do agent-based modeling simulations, it seems quite natural that agent-based models would be employed to understand laboratory findings and not the other way around.

There is, however, a second explanation for why the ACE methodology has been used to understand experimental findings with human subjects. Once a human subject experimental design has been computerized, it is a relatively simple matter to replace some or all of the human subjects with "robot" agents. Indeed, one could make the case that some of the earliest ACE researchers were researchers conducting experiments with human subjects. For instance, Roth and Murnighan (1978) had individual human subjects play repeated prisoner's dilemma games of various expected durations against artificial "programmed opponents" in order to more clearly assess the effect of variations in the expected duration of the game on the human subjects' behavior. Similarly, Coursey, Issac, Luke

[1]See, e.g., Axelrod and Tesfatsion (2005) or Batten (2000) for introductions to the ACE methodology.

[2]See, Davis and Holt (1993) and Roth (1995) for histories of the experimental methodology.

and Smith (1984) and Brown–Kruse (1991) tested contestable market theories with human subjects in the role of sellers and robots in the role of buyers. The robots were programmed to fully reveal their market valuations and were introduced after human subject buyers were found to be playing strategically, in violation of the theory being tested. Gode and Sunder (1993) were the first researchers to "go all the way" and completely replace the human subject buyers and sellers in the experimental laboratory double auction environment with artificial agents, whom they dubbed "zero–intelligence" agents. Their approach, discussed in greater detail below, serves as the starting point for our survey. Subsequently, many researchers have devised a variety of agent-based models in an effort to explain, understand and sometimes to predict behavior in human subject experiments.[3]

Of course, the great majority of ACE researchers, following the lead of Schelling (1978), Axelrod (1984), or Epstein and Axtel (1996), do not feel constrained in any way by the results of human subject experiments or other behavioral research in their ACE modeling exercises. These researchers endow their artificial agents with certain preferences and what they perceive to be simple, adaptive learning rules. As these artificial agents interact with one another and their environment, adaptation takes place at the individual level, or at the population level via relative fitness considerations, or both. The details of how agents adapt are less important than the aggregate outcomes that emerge from repeated interactions among these artificial agents.

ACE researchers contend that these emergent outcomes cannot be deduced without resorting to simulation exercises, and that is the reason to abandon standard neoclassical approaches.[4] But it is not always clear when ACE approaches are preferred over standard, deductive economic theorizing. As Lucas (1986, p. 218) observed,

> "It would be useful, though, if we could say something in a general way about the characteristics of social science prediction problems where models emphasizing adaptive aspects of behavior are likely to be successful versus those where the non-adaptive or equilibrium models of economic theory are more promising."

Lucas went on to suggest that experiments with human subjects might serve to resolve such questions, and gave several examples. Of course, economic experiments are not without problems of their own. ACE researchers (e.g., Gode and Sunder (1993), Chan et al. (1999)) have argued that agent-based modeling permits greater control over the preferences and information-processing capabilities of agents than is possible in laboratory experiments, where human subjects often vary in their learning abilities or preferences (e.g. in their attitudes towards risk), despite careful efforts to control some of these differences by experimenters. Further, one can question the external validity of the behavior of the human subjects, who are often inexperienced with the task under examination and who may earn payments that do not accurately approximate "real-world" incentives.[5]

In addition to questioning when the ACE methodology is appropriate, one can also question the external validity of ACE modeling assumptions and simulation findings. Many ACE researchers, following the lead of Epstein and Axtell (1996) adopt the "generative approach" to understanding empirical phenomena. This involves pointing to some empirical phenomenon, for example, skewed wealth distributions, and asking: "can you grow it?" In other words, can you specify a multi-agent complex adaptive system that generates the empirical phenomenon.

[3]See Mirowski (2002) for an engaging history of the emergence of economics as a "cyborg science," and, in particular, the role played by experimentalists. See also Miller (2002) for a history of experimental analyses of financial markets.

[4]Batten (2000) offers some advice as to when ACE models are appropriate and when old-fashioned analytic methods are preferred.

[5]However, as Smith (1982, p. 930) observes, "...there can be no doubt that control and measurement can be and are much more precise in the laboratory than in the field experiment or in a body of Department of Commerce data."

While the ability to generate a particular empirical phenomenon via an ACE simulation exercise does represent a certain kind of understanding of the empirical phenomenon, ACE researchers could do more to increase our confidence in this understanding. Indeed, the empirical phenomena under study are often the result of some casual empiricism on the part of the ACE researcher. More precise and careful empirical support, using field data or other observations could be brought to bear in support of a particular phenomenon, but this is not (yet) the standard practice. Further, the processes by which agents in ACE models form expectations, choose actions or otherwise adapt to a changing environment is not typically based on any specific micro evidence; the empirical comparisons that most interest ACE researchers are between the simulated aggregate outcomes and the empirical phenomenon of interest. The shortcomings of such an approach have not gone unnoticed. Simon (1982) for example, writes:

> Armchair speculation about expectations, rational or other, is not a satisfactory substitute for factual knowledge as to how human beings go about anticipating the future, what factors they take into account, and how these factors, rather than others, come within the range of their attention.

As I argue in this chapter, data from human subject experiments provide a ready-made source of empirical regularities that can be used to calibrate or test ACE models of individual decision-making and belief or expectation formation. Explaining the aggregate findings of a human subject experiment might also serve as the goal of an agent-based modelling exercise.

The main behavioral principle that ACE researchers use in modeling individual artificial agent behavior is, what Axelrod (1997) has termed, the "keep-it-simple-stupid" (KISS) principle. The rationale behind this folksy maxim is that the phenomena that emerge from simulation exercises should be the result of multi-agent interactions and adaptation, and not because of complex assumptions about individual behavior and/or the presence of "too many" free parameters. Of course, there are many different ways to adhere to the KISS principle. Choosing simple, parsimonious adaptive learning rules that also compare favorably with the behavior of human subjects in controlled laboratory settings would seem to be a highly reasonable selection criterion.

Experimental economists and ACE researchers are natural allies, as both are interested in dynamic, decentralized inductive reasoning processes and both appreciate the importance of heterogeneity in agent types. Further, the economic environments designed for human subject experiments provide an important testbed for agent-based modelers. The results of human subject experiments are useful for evaluating the external validity of agent-based models at the two different levels mentioned above. At the aggregate level, researchers can and have asked whether agent-based models give rise to the same aggregate findings that are obtained in human subject experiments. For instance, do artificial adaptive agents achieve the same outcome or convention that human subjects achieve? Is this outcome an equilibrium outcome in some fully rational, optimizing framework or something different? At the individual level, ACE researchers can and have considered the external validity of the adaptive rules they assign to their artificial agents by comparing the behavior of individual human subjects in laboratory environments with the behavior of individual artificial agents placed in the same environments. Achieving some kind of external validity, at either the aggregate or the individual level, should enable agent-based modelers to feel more confident in their simulation findings. They may then choose to abandon, with even greater justification, the constraints associated with the experimental methodology or those of standard, deductive economic theorizing.

This chapter surveys and critiques three main areas in which agent-based models have been used to study findings from human subject experiments. In the next section, we explore what has been termed the "zero–intelligent" agent approach, which consists of a set of agent-based models with very low rationality constraints. In the following section, we explore a set of agent-based models that employ somewhat more sophisticated individual behaviors, ranging from simple stimulus-response learning to more complicated belief-based learning approaches. Finally, in the last section, we explore agent-based models where individual learning is even more complicated, as in a classifier system, or is controlled by population-wide selection criteria as in genetic algorithms. In all cases, we compare the findings of human subject experiments with those of agent-based simulations.

2 Zero–Intelligence Agents

The zero-intelligence agent trading model was developed to explain findings from laboratory double auction experiments with human subjects. We therefore begin with a discussion of the double auction environment and laboratory findings.

2.1 The Double Auction Environment

The double auction is one of the most celebrated market institutions, and is widely used in all kinds of markets including stock exchanges and business-to-business e-commerce. The convergence and efficiency properties of the double auction institution have been the subject of intense interest among experimental economists, beginning with the work of Smith (1962), who built on the early work of Chamberlin (1948). Altering Chamberlin's design so that information on bids and asks was centralized as in a stock market, Smith (1962) was able to demonstrate that experimental markets operating under double auction rules yielded prices and trading volumes consistent with competitive equilibrium predictions, despite limited knowledge on the part of participants of the reserve values of other participants.

The double auction markets studied by Smith and subsequently by other experimentalists and ACE researchers can be described using a simple, one-good environment, though multi-good environments are also studied. The single good can be bought and sold over a fixed sequence of trading periods, each of finite length. The N participants are often divided up between buyers or sellers (in some environments agents can play either role). Buyer i has valuation for unit $j = 1, 2, ...$ of the good, v_{ij}, where the valuations satisfy the principle of diminishing marginal utility in that $v_{ij} \geq v_{ik}$ for all $j < k$. Similarly, seller i has a cost of selling unit $j = 1, 2, ...$ of the good, c_{ij}, which satisfies the principle of increasing marginal cost, $c_{ij} \leq c_{ik}$ for all $j < k$. Sorting the individual valuations from highest to lowest gives us a step-level market demand curve, and sorting the individual costs from lowest to highest gives us a step-level market supply curve. The intersection of these two curves, if there is one, reveals the competitive equilibrium price and quantity. The left panel of Figure 1 taken from Smith (1962), provides an illustration. [Figure 1 here]. In this figure, the valuations of the 11 buyers (for a single unit) have been sorted from highest to lowest, and the costs to the 11 sellers (of a single unit) have been sorted from lowest to highest. The equilibrium price is \$2.00 and the equilibrium quantity is 6 units bought and sold.

In the experimental double auction markets, subjects are informed as to whether they will be buyers or sellers and they remain in this role for the duration of the session. Buyers are endowed with private values for a certain number of units and sellers are endowed with private costs for a certain number of units. No subject is informed of the valuations or costs of other participants. Buyers are instructed that their payoff from buying their j^{th} unit is equal to $v_{ij} - p_j$, where p_j is the price the buyer agrees to pay for the j^{th} unit. Similarly, sellers are instructed that their payoff from selling their j^{th} unit at price p_j is equal to $p_j - c_{ij}$. The double auction market rules vary somewhat across studies, but mainly consist of the following simple rules. During a trading period, buyers may post any bid order and sellers may post any ask order at any time. Further, buyers may accept any ask or sellers may accept any bid at any time. If a buyer and seller agree on a price, that unit is exchanged and is no longer available for (re)sale for the duration of the period. The buyer-seller pair earns the profit each realized on their transaction.

In many double auction experiments, the order book is cleared following each transaction, so that buyers and sellers have to resubmit bids and asks. It is also standard practice to assume a closed order book, meaning that subjects can only observe the best bid and ask price at any moment in time. To surplant the current best bid (ask) a buyer (seller) has to submit a bid (ask) that is higher (lower) than the best bid (ask); this is known as the standard bid/ask improvement rule. At all times, the current best bid-ask spread is known to all market participants. The entire history of market transaction prices is also public knowledge.

The striking result from applying these double auction rules in laboratory markets is the rapid convergence to the competitive equilibrium price and quantity. The right panel of Figure 1, shows the path of prices over five trading periods in session 1 of the Smith (1962) study. The first transacted price in period 1 is for \$1.70, the second for \$1.80, etc. Notice that the number of transacted prices in period 1 is 5, which is one short of the competitive equilibrium prediction, and these prices all lie below the competitive equilibrium price of \$2.00. As subjects gain experience over trading periods 2-5, however, the deviations of traded prices and quantities from the competitive equilibrium values steadily decrease. This main finding has been replicated in many subsequent experiments, and continues to hold even with small numbers of buyers and sellers (e.g. 3-5 of each).

2.2 Gode and Sunder's Zero-Intelligence Traders

Gode and Sunder (1993) were interested in assessing the source of this rapid convergence to competitive equilibrium in laboratory double auction markets. They hypothesized that the double auction rules alone might be responsible for the laboratory findings and so they chose to compare the behavior of human subject traders with that of programmed robot traders following simple rules. As these robot players chose bids and asks randomly, over some range, Gode and Sunder chose to label them "zero-intelligence" (or ZI) machine traders. This choice of terminology has stimulated much debate, despite Gode and Sunder's disclaimer that "ZI traders are not intended as descriptive models of individual behavior."

Gode and Sunder's 12 ZI traders were divided up equally into buyers and sellers. In the most basic environment, the buyer's bids and the seller's asks were random draws from a uniform distribution, $U[0, B]$, where the upper bound B, was chosen so as to exceed the highest valuation among all buyers. In particular, Gode and Sunder chose $B = 200$. Buyers' bids and sellers' asks were made without concern for whether the bids or asks were profitable. Gode and Sunder referred to these unconstrained traders as ZI-U traders. In the other, more restrictive environment they considered, buyer i's bid for unit j was a random draw from the uniform distribution, $U[0, v_{ij}]$ and seller i's ask for unit j was random draw from the uniform distribution $U[c_{ij}, B]$. As the traders in this environment were constrained from making unprofitable trades, they were referred to as ZI-C traders.

A trading period consisted of 30 seconds for the ZI traders and 4 minutes for a parallel human subject experiment. Within the 30 second period, the standard double auction rules applied: the best available bid is the one that is currently the highest of all bids submitted since the last transaction, while the best available ask is the one that is currently the lowest of all asks submitted since the last transaction. A transaction occurs if either a new bid is made that equals or exceeds the current-best ask, in which case the transaction occurs at the current-best ask price, or a new ask is made that equals or falls below the current-best bid, in which case the transaction occurs at the current-best bid price. Once a transaction occurs, all unaccepted bids/asks are cleared from the order book and, provided that the period has not ended, the process of bid/ask submission begins anew. Traders were further restricted to buying/selling their j^{th} unit before buying or selling their $j + 1^{th}$ unit. This sequencing restriction is not a double auction trading restriction, and it appears to be quite important to Gode and Sunder's results.[6] Of course, if every agent has a single inframarginal unit to buy or sell (those units to the left of the intersection of demand and supply) and one or more extramarginal units (units to the right of the intersection point), as is often the case in double auction experiments, then there is no sequencing issue.

The results from a simulation run of the ZI-U and ZI-C artificial trading environment and from a human subject experiment with 13 subjects (1 extra buyer) are shown in the three panels of Figure 2. [Figure 2 here]. The left panels show the induced demand and supply step-functions and the competitive equilibrium prediction (price =80, quantity=24) while the right panels show the path of transaction prices across the 6 trading periods. Gode and Sunder's striking finding is that the transaction price path with the budget constrained ZI-C traders bears some resemblance to the path of prices in the human subject experiment. In particular, prices remain close to the competitive equilibrium price, and within a trading period, the price volatility declines so that prices become even closer to the competitive equilibrium prediction. This finding stands in contrast to the ZI-U environment, where transaction prices are extremely volatile and there is no evidence of convergence to the competitive equilibrium. As the ZI-C or ZI-U agents have no memory regarding past prices, the difference in the simulation findings are entirely due to the difference in trading rules, namely the constraint imposed on ZI-C traders ruling out unprofitable trades. The dampened volatility in prices over the course of a trading period arises from the fact that units with the highest valuations or lowest costs tend to be traded earlier in the period, as the range over which ZI-C agents may submit bids or asks for these units is larger than for other units. After these units are traded, the bid and ask ranges of ZI-C agents with units left to trade become increasingly narrow, and consequently, the volatility of transaction prices becomes more damped.

Gode and Sunder also examine the "allocative efficiency" of their simulated and human subject markets, which is defined as the sum of total profit earned over all trading periods divided by the maximum possible profit, which is simply the sum of consumer and producer surplus (e.g., the shaded area in the left panel of Figure 1. They find that with the ZI-U traders, market efficiency averages 78.3 percent, while with ZI-C traders it averages 98.7 percent; the latter figure is slightly higher than the average efficiency achieved by human subjects, 97.6 percent! Gode and Sunder summarized their findings as follows:

> Our point is that imposing market discipline on random unintelligent behavior is sufficient to raise the efficiency from the baseline level [that attained using ZI-U agents] to almost 100 percent in a double auction. The effect of human motivations and cognitive abilities has a second-order magnitude at best."

One explanation for the high efficiency with the ZI-C agents is provided in Gode and Sunder (1997b). They consider the consequences for allocative efficiency of adding or subtracting various market rules and arrive at some very intuitive conclusions. First, they claim that voluntary exchange by agents who are sophisticated enough to avoid losses is necessary to eliminate one source of inefficiency, namely unprofitable trades. By voluntary exchange, they mean that agents are free to accept or reject offers. The second part of this observation, that agents are sophisticated enough to avoid losses, is the hallmark of the ZI-C agent model, but its empirical validity is not really addressed. We know from experimental auction markets, for example, where private values or costs are induced and subjects have perfect information about these values or costs, that subjects sometimes bid in excess of their private valuations (Kagel, Harstad and Levin (1987)). Gode and Sunder (1997a) are careful to note that they

[6] See, e.g., the discussion of Brewer et al. (2002) below.

"are not trying to accurately model human behavior," (p. 604) but the subtext of their research is that the no unprofitable trades assumption does not presume great sophistication; the traders are "zero-intelligence" but constrained. Perhaps the more restrictive assumption is that agents have perfect information about their valuations and costs and perfect recall about units they have already bought or sold. Absent such certainty, it might be harder to reconcile the assumption of no unprofitable trades with the observation that individuals and firms are sometimes forced to declare bankruptcy.

Other sources of inefficiency are that ZI-C traders fail to achieve any trades, and that extramarginal traders - traders whose valuations and costs lie to the right of the intersection of demand and supply – displace inframarginal traders whose valuations lie to the left of the intersection of demand and supply and who have the potential to realize gains from trade. Gode and Sunder (1997ab) define an expected efficiency metric based on a simplified model of induced demand and supply and show that inefficiencies arising from failure to trade can be reduced by having multiple rounds of trading. Inefficiencies arising from the displacement of inframarginal traders by extramarginal traders can depend on the "shape" of the extramarginal demand and supply, e.g., whether it is steep or not) and on the market rules, e.g., whether bids and asks are ranked and a single market clearing price is determined (as in a call market) or whether decentralized trading is allowed (as in the standard, double auction).

Gode and Sunder (2004) further consider the consequences of nonbinding price ceilings on transaction prices and allocative efficiency in double auctions with ZI-C traders (the analysis of price floors follows a symmetric logic). A nonbinding price ceiling is an upper bound on admissible bid and ask prices that lies above the competitive equilibrium price. If a submitted bid or ask exceeds the price ceiling it is either rejected or reset at the ceiling bound. Since the bound lies above the competitive equilibrium price, theoretically it should not matter. However, in experimental double-auction markets conducted by Isaac and Plott (1981) and Smith and Williams (1981), non-binding price ceilings work to depress transaction prices below the competitive equilibrium level relative to the case where such ceilings are absent. Gode and Sunder (2004) report a similar finding when ZI-C agents are placed in double auction environments with non-binding price ceilings similar to the environments examined in the experimental studies. Gode and Sunder explain their finding by noting that a price ceiling reduces the upper-bound on the bid ask range, and with ZI-C agents, this reduction immediately implies a reduction in the mean transaction price relative to the case without the price ceiling. Further they show that with ZI-C agents, a price ceiling reduces allocative efficiency as well (which is consistent with the experimental evidence) by making it more likely that extramarginal buyers are not outbid by inframarginal buyers, and by excluding extramarginal sellers with costs above the ceiling from playing any role.

Summing up, what Gode and Sunder (1993, 1997ab, 2004) have shown is that simple trading rules in combination with certain market institutions can generate data on transaction prices and allocative efficiency that approach or exceed those achieved by human actors operating in the same experimental environment. This research finding serves as an important behavioral foundation for the "KISS" principle that is widely adopted in agent-based modeling. However, agent-based modelers are not always as careful as Gode and Sunder to provide external validity (experimental or other evidence) for the simple rules they assigned to their artificial agents.

2.3 Reaction and Response

Not surprisingly, the Gode and Sunder (1993) paper provoked a reaction, especially by experimenters, who viewed the results as suggesting that market institutions were pre-eminent and that human rationality/cognition was unimportant. Of course, the various different market institutions are all of human construction, and are continually evolving, so the concern about the *source* of market efficiency (institutional or human behavior) seems misplaced.[7] Nonetheless, there is some experimental literature addressing what human subjects *can* do that Gode and Sunder-type ZI agents cannot.

Van Boening and Wilcox (1996) consider double auction environments where buyers all have the same market valuation for units of the good, and sellers do not have fixed or marginal costs for various units, but instead have large "avoidable costs" – costs they incur only if they decide to actively engage in exchange. In such environments, seller decisions to enter the market can be fraught with peril since they cannot anticipate the entry decisions of other sellers and consequently, supply, and a seller's average costs (avoidable cost divided by number of units sold) can be highly variable. Van Boening and Wilcox report that the efficiency of human subject traders in the more complex DA-avoidable costs environment is much lower than in the standard DA environment with pure marginal costs, but the efficiency of ZI traders in the DA-avoidable cost market is significantly worse than the human subject traders operating in the same environment.

Brewer et al. (2002) consider a different but similarly challenging variant of the double auction environment, where demand and supply conditions do not change within a trading period as exchanges between buyers and sellers remove units from trade, but where instead, market conditions remain invariant over each (and all) trading periods. This is accomplished by *continually refreshing* the units that all buyers (sellers) are able to buy (sell) following any trades, and Brewer et al. refer to this market environment as one with continuously refreshed supply and demand (CRSD).[8] Recall that the dampened volatility of prices over a trading period in the ZI-C simulations was owing to the greater likelihood that inframarginal units with the lowest marginal cost/highest reservation value would trade earlier than other inframarginal units where the difference between marginal cost and valuation was lower. In the continually refreshed design of Brewer et al. the forces working to dampen price adjustment over the course of a trading period are removed. Hence prices generated by ZI-C traders in the CRSD environment are quite random and exhibit no tendency toward convergence to any competitive equilibrium notion (Brewer et al. consider several). On the other hand, the human subject traders in the CRSD environment have no difficulty converging to the "velocity-based" competitive equilibrium, and are also able to adjust to occasional perturbations to this equilibrium.

Sadrieh (1998) studies the behavior of both human subjects and ZI agents in an "alternating" double-auction market, a discrete-time version of the continuous double-auction market that retains the double auction trading rules. The alternating DA is more conducive to a game-theoretic analysis but differs in some respects from the standard continuous DA in that only one side of the market (buyers or sellers) is active at once, the bids or asks submitted are sealed (made simultaneously), and there is complete information about values, costs and ex post offers of all players. The determination of the opening market side (buyers or sellers) is randomly determined, and then alternates over the course of a trading period. Sadrieh's game-theoretic prediction is that convergence to the competitive equilibrium price would be from above (below) when sellers (buyers) opened the market. By contrast, ZI simulations suggested that convergence to the market price would be from above (below) when the surplus accruing to buyers (sellers) in the competitive equilibrium was relatively larger than that accruing to sellers (buyers). Sadrieh's experimental findings, however, were at odds with both of these predictions; the most typical path for prices in an experimental session involves convergence to the competitive equilibrium from below, regardless of which side opens the market or the relative size of the surpluses. On the other hand, ZI simulations accurately predicted the extent of another of Sadrieh's findings, "the proposer's curse." The curse is that those submitting bids or asks tend to do so at levels

[7] Analogously, there was great outcry in May 1997 when Gary Kasparov, widely considered to be the greatest player in the history of chess, first lost a chess match to a machine nicknamed "Big Blue," even though Big Blue's hardware and algorithms were developed over many years by (human) researchers at IBM.

[8] A motivating example is housing or labor markets without entry or exit of participants. A worker attracted by a firm to fill a job vacancy, leaves another vacancy at his old firm, so that labor demand is effectively constant.

that yield them lower profits relative to the competitive equilibrium price; the additional gains go to the players accepting those bids or asks. Sadrieh reports that the frequency of proposer's curse among inexperienced subjects was comparable to that found in ZI simulations, though experienced subjects learned to avoid the curse.

Experimentalists are not the only ones to challenge Gode and Sunder's findings. AI researchers Cliff and Bruten (1997ab) have examined the sensitivity of Gode and Sunder's findings to the elasticity of supply and demand. In particular they examine DAs with four different types of induced demand and supply curves as shown in Figure 3. [Figure 3 here]. Of these four economies, simulations using ZI-C agents converge to the competitive equilibrium price, P_0 and quantity, Q_0 only in economies of type A, the same type that Gode and Sunder consider, and not in economies of type B, C or D. The intuitive reason for this finding (which Cliff and Bruten formalize) is that the probability density function (pdf) for transaction prices (a random variable with ZI agents) is symmetric about the competitive equilibrium price, P_0, only in the case of economy A; in the other economies, the transaction price pdf has P_0 as an upper or lower bound. Since the expected value of a random variable, such as the transaction price, is the "center of gravity" of the pdf, it follows that price convergence with ZI-C agents only occurs in economies of type A. Cliff and Bruten's simulations bear out this conclusion. It remains to be seen how human subject traders would fare in economies such as B, C and D. However, as a purely theoretical exercise, Cliff and Bruten suggest that an alternative algorithm, which they call "zero–intelligence plus" (ZIP), achieves convergence to competitive equilibrium in economies such as B, C, and D more reliably than does Gode and Sunder's ZI approach. By contrast with ZI agents, ZIP agents aim for a particular profit margin on each unit bought or sold, and this profit margin dictates the bid or ask they submit. Each agent's profit margin is adjusted in real time depending on several factors most of which concern properties of the most recent bids, asks and transactions made. Hence ZIP involves some memory though it is limited to the most recent data available. Comparisons of ZIP simulations with some of Smith's aggregate experimental findings are encouraging, though a more detailed analysis of the ZIP mechanism's profit margin adjustment dynamic with experimental data has yet to be performed.

As these critiques make clear, it is relatively easy to construct environments where human subjects outperform ZI agents or environments where ZI agents fail to converge to competitive equilibrium. However the broader point of Gode and Sunder's pioneering work is not that human cognitive skills are unimportant. Rather it is that, in certain market environments, aggregate allocation, price and efficiency outcomes can approach the predictions of models premised on high levels of individual rationality even when individual traders are only minimally rational. Understanding precisely the conditions under which such a mapping can be assured clearly requires parallel experiments with both human and artificial subjects.

2.4 Other Applications of the ZI methodology

In addition to Cliff and Bruten, several other researchers have begun the process of augmenting the basic ZI methodology in an effort to explain economic phenomena in various environments. The process of carefully building up an agent-based framework from a simple foundation, namely budget-constrained randomness, seems quite sensible, and indeed, is well under way.

Bosch-Doménech and Sunder (2001) expand the Gode and Sunder (1993) double auction environment to the case of m interlinked markets populated by dedicated buyers in market 1, by dedicated sellers in market m, and consisting exclusively of arbitrage traders operating in markets $i = 1, 2, ...m$. In the baseline model, arbitrageurs are prevented from holding any inventory between transactions. They operate in adjacent markets, simultaneously buying units in market $i + 1$ and selling them in market i. As market m is the only one with a positive net supply of the asset, trading necessarily begins there. Absent the possibility of inventories, a transaction in market m instantaneously ripples through the entire economy (the other $m-1$ markets) so that the good traded quickly ends up in the hands of one of the dedicated buyers in market 1. One interpretation of this set–up is that of a *supply–chain*, consisting of producers in market m, middlemen in markets m, $m-1$,...,1 and ultimate consumers in market 1. Bosch-Doménech and Sunder report simulations showing that regardless of whether the number of markets, m is 2, 5 or 10, prices and volume in each market quickly converge to the competitive equilibrium levels obtained by crossing demand in market 1 with supply in market m, and that market efficiency is close to 100%. Bosch-Doménech and Sunder further examine what happens when arbitrageurs can take long or short inventory positions. As the number of short or long positions that arbitrageurs can take is increased, and the number of markets, m, gets large, prices remain very close to the competitive equilibrium prediction in all m markets, but trading volume in the "middle" markets (populated only by arbitrageurs) increases well beyond the competitive equilibrium prediction and market efficiency declines. This finding is an argument for keeping supply chains short (or finding ways to "cut out the middleman"). An experimental test of this prediction remains to be conducted.[9]

Duffy and Ünver (2005) use the ZI methodology to understand asset price bubbles and crashes in laboratory market experiments of the type first examined by Smith, et al. (1988). In these laboratory markets there is a single "asset" that is traded in a finite number, T, of trading periods; unlike the previously described double auction experiments, players here can be either buyers or sellers, and so they are referred to as traders. Those holding units of the asset at the end of each trading period are entitled to a random dividend payment per unit, with expected value $\overline{d}$. The fundamental expected market value of a unit of the asset at the start of trading period $t \leq T$ is given by $D_t = \overline{d}(T-t+1) + D_{T+1}$, where D_{T+1} is the final buy-out value per unit of the asset held at the close of period T. All participants' initial endowments of the asset and money have the same expected value, though the allocation of assets and money differs across agents. Consequently, risk neutral traders should be indifferent between engaging in any trades or trading at the fundamental market value which is declining over time. With groups of inexperienced human subjects, the path of the mean transaction price tends to start below the fundamental value in the first trading periods, quickly soaring above this fundamental value in the middle trading periods before finally crashing back to or below fundamental value near to the final trading period T.

Duffy and Ünver show that such asset price bubbles and crashes can arise with ZI agents, who are a little more sophisticated than Gode and Sunder's ZI-C agents – Duffy and Ünver call them "near–zero intelligence agents" In particular, Duffy and Ünver's agents are not constrained from submitting bids or asks in excess of the fundamental market value of the asset as such a constraint would rule out the possibility of bubbles. As in Gode and Sunder (1993) there is an exogenously imposed range for bids and asks given by the interval $[0, \kappa D_t^T]$, where $\kappa > 0$. In addition, bids and asks are not entirely random. The ask of trader i in period t is given by $a_t^i = (1-\alpha)u_t^i + \alpha\overline{p}_{t-1}$, where u_t^i is a random draw from $[0, \kappa D_t^T]$ and $\overline{p}_{t-1}$ is the mean transaction price from the previous trading period; the weight given to the latter, α, if positive, introduces a simple herding effect, and further implies that ask prices must rise over the first few periods. A similar herding rule is used to determine bids. The random component to bids and asks serves to insure that some transactions take place. As in Gode and Sunder (1993), budget constraints are enforced; traders cannot sell units they do not own, nor can buyers submit bids in excess of their available cash balances. Finally, to account for the finite horizon, which was known to the human subjects, Duffy and Ünver endow their artificial agents with some *weak foresight*; specifically, the probability that a trader submits a bid (as opposed to an ask)

[9] See, however, the related work of Grossklags and Schmidt (2004), who add artifical arbitrage agents to a double auction experiment with human subjects.

is initially .5, and decreases over time, so, over time, there are more asks than bids being submitted reflecting the declining fundamental value of the asset. Standard double auction trading rules are in effect. Duffy and Ünver use a simulated method of moments procedure to calibrate the parameter choices of their model, e.g. κ, α, so as to minimize the mean squared deviations between the price and volume path of their simulated economies and the human subject markets of Smith et al. (1988). They are able to find calibrations that yield asset price bubbles and crashes comparable to those observed in the laboratory experiments and are able to match other, more subtle features of the data as well.

2.5 ZI Agents in General Equilibrium

The original Gode and Sunder (1993) study follows the Smith (1962) partial equilibrium laboratory design, where market demand and supply are exogenously given. In more recent work, zero-intelligence traders have been placed in general equilibrium settings, with the aim of exploring whether they might achieve competitive equilibrium in such environments. Gode, Spear and Sunder (2000) placed zero-intelligence traders, who could both buy and sell, in a two–good, pure exchange economy (an Edgeworth box). Traders are divided up into two types $i = 1,2$, that differ only in terms of the parameters of their Cobb-Douglas utility function defined over the two goods and their initial endowments of these two goods. The trading rules for ZI agents in the general equilibrium environment are similar to rules found in the partial equilibrium environment. In particular, in the general equilibrium environment, ZI agents's bids and asks are limited to utility improving allocations. Specifically, each agent of type i begins by calculating the slope of its indifference map at its current endowment point. The slope is calculated in terms of radians, r, where $0 \leq r \leq \frac{\pi}{2}$; this gives the number of units of good y the trader is willing to give up per unit of good x. Next, the agent picks two random numbers, $b \in [0,\ r]$ and $a \in [r, \pi/2]$, with the first representing its bid price for units of good y in terms of good x, and the second representing its ask price for units of good y in terms of good x. Finally, the unit of a transaction for simulation purposes involves a discrete step size in the quantity of both goods; otherwise, with an infinitesimal quantity exchanged each period, convergence could take a long time. A consequence of this discrete step size assumption is that an adjustment has to be made to the bid and ask ranges to account for the curvature of the indifference map. Given these trading restrictions, and the double auction rules, market transactions will be limited to lie in the set of Pareto improving reallocations, i.e., the area between the two indifference maps. Once an exchange occurs, endowments are updated, and the process described above begins anew.

Figure 4 (taken from Gode et al. (2000)) illustrates this process. [Figure 4 here]. The initial endowment is at point A, and the indifference maps of the two agent types intersect at this point. The ZI trading restrictions and discrete step size imply that the first transaction occurs along the arc BC. If this first round transaction occurs at, say, point D, this point becomes the new endowment point. The set of feasible trades in the subsequent period lie on the arc B'C', etc. Given this characterization of ZI trading rules it is clear – even without simulating the system – that this updating process must eventually converge to the contract curve, representing the set of all Pareto optimal allocations, and will then cease, as the bid-ask range shrinks to the null set.[10] And, indeed, this is precisely what Gode et al. (2000) find. Simulations of ZI agents operating according to the rules described above yield limiting allocations that lie on the contract curve, and so these allocations are Pareto optimal. However, these allocations do not necessarily correspond to the competitive equilibrium allocation,

[10]One consequence of studying ZI, directed random search processes is that once the environment is specified, actual simulation of the search process may be unnecessary. Still, the value of this approach lies in building the minimal, necessary restrictions on directed random search that achieve the desired outcome. The ZI approach aids in formulating these restrictions, by greatly simplifying agent behavior, allowing the researcher to concentrate on the institutional restrictions.

the point on the contract curve where the two price-offer curves of the two agent types intersect. So, by contrast with the findings in the partial equilibrium framework, ZI-trading rules turn out to be insufficient to guarantee convergence to competitive equilibrium in the two–good general equilibrium environment.

The nonconvergence of the ZI algorithm to competitive equilibrium is further addressed by Crockett, Spear and Sunder (CSS) (2004) who provide an answer to the question of "how much additional 'intelligence' is required" for ZI agents to find a competitive equilibrium in a general equilibrium setting with M agents and ℓ commodities. In their environment, ZI agents do not submit bids or asks. Rather a proposed allocation of the ℓ goods across the M agents is repeatedly made, corresponding to a random draw from an epsilon-cube centered at the current endowment point. Agent i compares the utility he gets from the proposed allocation with the utility he receives from the current endowment. If the utility from the proposed allocation is higher, agent i is willing to accept the proposal. If all M agents accept the proposal, the proposed allocation becomes the new endowment point. The random proposal generation process (directed search) then begins anew and continues until no further utility improvements are achieved. At this point the economy has reached a near-Pareto optimum (an allocation that lies approximately in the Pareto set, though not necessarily a competitive equilibrium; this outcome is analogous to the final outcome of the Gode, Spear and Sunder (2000) algorithm. Crockett, Spear and Sunder further assume that once agents have reached this approximate Pareto optimum (PO), they are able to calculate the common, normalized utility gradient at the PO allocation. The ZI agents are then able to determine whether this gradient passes through their initial endowment point (the condition for a competitive equilibrium) or not. If it does not, then, in the PO allocation, some agents are subsidizing other agents. Note that these assumptions endow the ZI agents with some calculation and recall abilities that are not provided (or necessary) in Gode and Sunder's partial equilibrium environment.

Consider for example, the two agent, two–good case. In this case, the normalized utility gradient corresponds to a price line through the tangency point of the two indifference curves (preferences must be convex), representing the relative price of good 2 in units of good 1 at the PO allocation. Suppose that at the end of trading period t, agent i's approximate PO allocation is $\hat{x}_i^t \in R_+^2$. Agent i's gain at this PO allocation can be written as:

$$\lambda_i^t = p^t(\hat{x}_i^t - \omega_i)$$

where p^t is the price line at the end of period t and $\omega_i \in R_+^2$ is agent i's initial endowment. Agent i is said to be subsidizing the other agent(s) if $\lambda_i < 0$. That is, at $p^t \gg 0$, agent i cannot afford to purchase his initial endowment. Crockett et al.'s innovation is to imagine that if agent i was a 'subsidizer' in trading period t, then in trading period $t+1$ he agrees to trade for only those allocations, x^{t+1} that increase his utility and that satisfy:

$$0 \geq p^t(x_i^{t+1} - \omega_i) \geq \lambda_i^t + \nu_i$$

where ν_i is a small, positive bound. With this additional constraint in place, the PO allocation achieved at the end of period $t+1$, $\hat{x}^{t+1}$, is associated with a larger gain for the subsidizing agent i, i.e. $\lambda_i^{t+1} > \lambda_i^t$, so he subsidizes less in period $t+1$ than in period t. When all i agents' gains satisfy a certain tolerance condition, convergence to a competitive equilibrium is declared. Crockett et al. show that while cycling is a possibility, it can only be a transitory phenomenon. Indeed, they provide a rigorous proof that their algorithm converges to the competitive equilibrium with probability 1.

This subsidization constraint puts to work the Second Welfare theorem – that every Pareto optimum is a competitive equilibrium for some reallocation of initial endowments. Here, of course, the initial endowment is not being reallocated. Instead, agents are learning over time to demand more (i.e. refuse trades that violate the subsidization constraint) if they have been subsidizing other agents in previous periods. The reallocation takes place in the amounts that agents agree to exchange with one another.

The appeal of Crockett et al.'s "ϵ-intelligent" learning algorithm is that it implements competitive equilibrium using only decentralized knowledge on the part of agent i, who only needs to know his own utility function and be able to calculate the normalized utility gradient at the PO allocation attained at the end of the previous period (or more simply, to observe immediate past prices). Using this information, he determines whether or not he was a subsidizer, and if so, he must abide by the subsidization constraint in the following period. The algorithm is simple enough so that one might expect that simulations of it would serve as a kind of lower bound on the speed with which agents actually learn competitive equilibrium in multi-good, multi-agent general equilibrium environments, analogous to Gode and Sunder's (1993) claim for ZI agents operating in the double auction.

Indeed, Crocket (2004) has conducted an experiment with paid human subjects aimed precisely at testing this hypothesis. Crocket's experiment brings the ZI research agenda full circle; his experiment with human subjects is designed to provide external validity for a ZI, agent-based algorithm whereas the original Gode and Sunder (1993) ZI model was developed to better comprehend the ability of human subjects to achieve competitive equilibrium in Smith's double auction model. Crockett's study explores several different experimental treatments that vary in the number of subjects per economy and in the parameters of the CES utility function defined over the two goods. For each subject, a preference function was induced, and subjects were trained in their induced utility function, i.e. how to assess whether a proposed allocation was utility improving. Further, at the end of each trading period, Crocket calculated for subjects the end–of–period–t marginal rate of substitution, p^t, as well as the value of the end–of–period–t allocation, $p^t x_i$, but did not tell subjects what to do with that information, which remained on subjects' screens for the duration of the following period, $t+1$. Subjects could plot the end–of–period–t price line on their screens to determine whether or not it passed through their beginning–of–period–t endowment point. Thus, subjects had all the information necessary to behave in accordance with the CSS algorithm, that is, they knew what comprised a utility improving trade and they had the information necessary to construct and abide by the subsidization constraint.

The left panel of Figure 5 presents the median end–of–period allocation of CSS–ZI agents for a particular 2-player CES parameterization, over trading periods 1-10, depicted in an Edgeworth box (the competitive equilibrium is labeled CE). [Figure 5 here]. The right panel of Figure 5 presents comparable median end–of–period allocations from one of Crockett's human subject sessions conducted in the same environment. Support for the hypothesis that the CSS–ZI algorithm accurately characterizes the behavior of paid human subjects appears to be mixed. On the one hand, nearly all of the human subjects are able to recognize and adopt utility improving trades, so that end of period allocations typically lie on or very close to the contract curve. And, once the contract curve is achieved, in subsequent periods, the human subjects appear to be moving in the direction of the competitive equilibrium allocation, as evidenced by the change in the median allocation at the end of period 10 relative to the median at the end of period 1 in the right panel of Figure 5. On the other hand, simulations of the CSS–ZI algorithm (left panel of Figure 5) suggest that convergence to the competitive equilibrium should have been achieved by period 6.

The reason for the slow convergence is that most, though not all subjects in Crockett's experiments are not abiding by the subsidization constraint; most are content to simply accept utility improving trades, while a few behave as CSS-ZI agents. The median allocation masks these differences, though the presence of some "CSS-ZI-type" agents moves the median allocation towards the competitive equilibrium. Hence, there is some support for the CSS-ZI algorithm, though convergence to competitive equilibrium by the human subjects is far slower than predicted by the algorithm.

2.6 Summary

The ZI approach is a useful benchmark, agent-based model for assessing the marginal contribution of institutional features and of human cognition in experimental settings. Building up agent-based models starting from zero memory and random action choices seems quite sensible and is in accord with Axelrod's KISS principle. Using ZI as a baseline, the researcher can ask: what is the minimal additional structure or restrictions on agent behavior that are necessary to achieve a certain goal such as near convergence to a competitive equilibrium, or a better fit to human subject data.

Thus far, the ZI methodology has been largely restricted to understanding the process by which agents converge to competitive equilibrium in either the partial equilibrium double auction setting or in simple general equilibrium pure exchange economies. ZI models have achieved some success in characterizing the behavior of human subjects in these same environments. More complicated economic environments, e.g. production economies or labor search models would seem to be natural candidates for further applications of the ZI approach.

The ZI approach is perhaps best suited to competitive environments, where individuals are atomistic and, as a consequence, institutional features together with constraints on unprofitable trades will largely dictate the behavior that emerges. In environments where agents have some strategic power, so that beliefs about the behavior of others become important, the ZI approach is less likely to be a useful modeling strategy. In such environments - typically game-theoretic – somewhat more sophisticated learning algorithms may be called for. We turn our attention to such learning models in the next section.

3 Reinforcement and Belief–Based Models of Agent Behavior

Whether agents learn or adapt depends on the importance of the problem or choice that agents face. Assuming the problem commands agents' attention, e.g. because payoff differences are sufficiently salient, the *manner* in which agents learn is largely a function of the information they posses and of their cognitive abilities. If agents have little information about their environment and/or they are relatively unsophisticated, then we might expect simple, backward-looking adaptive processes to perform well as characterizations of learning behavior over time. On the other hand, if the environment is informationally rich and/or agents are cognitively sophisticated, we might expect more sophisticated, even forward-looking learning behavior to be the norm.

This distinction leads to two broad sets of learning processes that have appeared in the agent-based literature, which we refer to here as reinforcement and belief learning following Selten (1991). Both learning processes are distinct from the fully rational, deductive reasoning processes that economists assign to the agents who populate their models. The important difference is that both reinforcement and belief learning approaches are decentralized, inductive, real-time, on–line learning algorithms that are unique to each agent's history of play. In this sense, they comprise agent-based models of learning. Our purpose here is to discuss the use of these algorithms in the context of the experimental literature, with the particular aim of evaluating the empirical plausibility of these learning processes.

3.1 Reinforcement learning

The hallmark of "reinforcement," "stimulus–response" or "rote" learning is Thorndike's (1911) 'law of effect': that actions or strategies that have yielded relatively higher (lower) payoffs in the past are more (less) likely to be played in the future. Reinforcement learning involves an inductive discovery of these payoffs; actions that are not chosen initially, are, in the absence of sufficient experimentation, less likely to be played over time, and may in fact, never be played (recognized). Finally, reinforcement learning does not require any information about the play of other participants or even the recognition that the reinforcement learner may be participating in a market or playing a game with others in which strategic considerations might be important. Thus, reinforcement learning involves a very minimal level of rationality that is only somewhat greater than that possessed by ZI agents.

Reinforcement learning has a long history associated with behaviorist psychologists (such as B.F. Skinner), whose views dominated psychology from 1920 through the 1960s, until cognitive approaches gained ascendancy. Models of reinforcement learning first appeared in the mathematical psychology literature, e.g. Bush and Mosteller (1955) and Suppes and Atkinson (1960). Reinforcement learning was not imported into economics however, until very recently, perhaps owing to economists' long-held scepticism toward psychological methods or of limited–rationality heuristics.[11]

Brian Arthur (1991, 1993) was among the first economists to suggest modeling agent behavior using reinforcement–type learning algorithms and to calibrate the parameters of such learning models using data from human subject experiments. In his 1991 paper, Arthur asks whether it is possible to design a learning algorithm that mimics human behavior in a simple N-armed bandit problem. Toward this aim, Arthur used data from an individual–choice, psychology experiment – a 2–armed bandit problem – conducted by Laval Robillard four decades earlier in 1952–3 and reported in Bush and Mosteller (1955) to calibrate his model.[12]

In Arthur's model, an agent assigns initial "strength" s_0^i to each of the $i = 1, 2, ...N$ possible actions. The probability of choosing action i in period t is then $p_t^i = s_t^i/C_t$, where $C_t = \sum_i s_t^i$. Given that action i is chosen in period t, it's strength is then updated: $s_t^{i\prime} = s_t^i + \phi_t^i$, where $\phi_t^i \geq 0$ is the payoff that action i earned in period t. Finally, all of the strengths, including the updated $s_t^{i\prime}$ are renormalized so as to achieve a prespecified constant value for the sum of strengths in period t: $C_t = Ct^\nu$, where C and ν represent the two learning parameters. When $\nu = 0$ (as in Arthur's calibration) the speed of learning is constant and equal to $1/C$.

Arthur 'calibrated' his learning model to the experimental data by minimizing the sum of squared errors between simulations of the learning model (for different (C, ν) combinations) and the human subject data over all experimental treatments, which amounted to variations in the payoffs to the two arms of the bandit. He showed that regardless of the treatment, the calibrated model tracked the experimental data rather well. In subsequent work, (e.g. the Santa Fe Artificial Stock Market (Arthur et al. (1997) discussed in LeBaron's (2005) chapter), Arthur and associates appear to have given up on the idea of calibrating *individual* learning rules to experimental data in favor of model calibrations that yield aggregate data that are similar to relevant field data. Of course, for experimental economists, the relevant data remain those generated in the laboratory, and so much of the subsequent development of reinforcement and other types of inductive, individual learning routines in economic settings has been with the aim of exploring experimental data.

Roth and Erev (1995) and Erev and Roth (1998) go beyond Arthur's study of the individual-choice, N-armed bandit problem and examine how well reinforcement learning algorithms track experimental data across various different multi-player games that have been studied by experimental economists. The reinforcement model that Roth and Erev (1995) develop is similar to Arthur's, but there are some differences and important modifications that have mainly served to improve the fit of the model to experimental data. The general Roth–Erev model can be described as follows.

Suppose there are N actions/pure strategies. In round t, player i has a propensity $q_{ij}(t)$ to play the j^{th} pure strategy (propensities are equivalent to strengths in Arthur's model). Initial (round 1) propensities (among players in the same role) are equal, $q_{ij}(1) = q_{ik}(1)$ for all available strategies j,k,

[11] An even earlier effort, due to Cross (1983), is discussed in Brenner's (2005) chapter.

[12] Regarding the paucity at the time of available experimental data, Arthur (1991, pp. 355-56 wrote: "I would prefer to calibrate on more recent experiments but these have gone out of fashion among psychologists, and no recent more definitive results appear to be available." Of course, economists have recently taken to conducting many such experiments.

and $\sum_j q_{ij}(1) = S_i(1)$, where $S_i(1)$ is an initial strength parameter, equal to a constant that is the same for all players, $S_i(1) = S(1)$; the higher (lower) is $S(1)$ the slower (faster) is learning.

The probability that agent i plays strategy j in period t is made according to the linear choice rule:

$$p_{ij}(t) = \frac{q_{ij}(t)}{\sum_{j=1}^{n} q_{ij}(t)}.$$

Some researchers prefer to work with the exponential choice rule:

$$p_{ij}(t) = \frac{\exp[\lambda q_{ij}(t)]}{\sum_{j=1}^{n} \exp[\lambda q_{ij}(t)]},$$

where λ is an additional parameter that measures the sensitivity of probabilities to reinforcements. For now, however, we follow Roth and Erev (1995) and focus on the linear choice rule.

Suppose that, in round t, player i plays strategy k and receives a payoff of x. Let $R(x) = x - x_{\min}$, where $x_{\min}$ is the smallest possible payoff. Then i updates his propensity to play action j according to the rule:

$$\begin{aligned} q_{ij}(t+1) &= (1-\phi)q_{ij}(t) + E_k(j, R(x)), \\ E_k(j, R(x)) &= \begin{cases} (1-\epsilon)R(x) & \text{if } j = k, \\ (\epsilon/(N-1))R(x) & \text{otherwise.} \end{cases} \end{aligned}$$

This is a three–parameter learning model, where the parameters are (1) the initial strength parameter, $S(1)$, (2) a forgetting parameter ϕ that gradually reduces the role of past experience, and (3) an experimentation parameter ϵ that allows for some experimentation.[13] Notice that if $\phi = \epsilon = 0$ we have a version of Arthur's model, where the main difference is that the sum of the propensities is not being renormalized in every period to equal a fixed constant. This difference is important, as it implies that as the propensities grow, so too will the denominator in the linear choice rule and the impact of payoffs for the choice of strategies will become attenuated. Thus, one possibility is that certain strategies that earn relatively high payoffs initially get played more often, and over time, there is lock–in to these strategies; alternatively, the "learning curve" is initially steep and then flattens out, properties that are consistent with the experimental psychology literature (Blackburn's (1936) "Power Law of Practice").

The ability of reinforcement learning models to track or predict data from human subject experiments has been the subject of a large and growing literature. Roth and Erev (1995) compare the performance of various versions of their reinforcement learning model with experimental data from three different sequential games: a market game, a best-shot/weakest link game and the ultimatum bargaining game; in all of these games, the unique subgame perfect equilibrium calls for one player to capture all or nearly all of the gains, though the experimental evidence is much more varied, with evidence of convergence to the perfect equilibrium in the case of the market and best–shot games but not in the case of the ultimatum game. Roth and Erev's simulations with their reinforcement learning algorithm yield this same divergent result. Erev and Roth (1998) use simulations of two versions of their reinforcement model (a one parameter version where $\phi = \epsilon = 0$) and the three parameter version) to *predict* play in several repeated normal form games where the unique Nash equilibrium is in mixed strategies. They report that the one and three-parameter models are better at predicting experimental data as compared with the Nash equilibrium point predictions, and that the three-parameter model even outperforms a version of fictitious play (discussed in the next section).

[13] In certain contexts, the range of strategies over which experimentation is allowed is restricted to those strategies that are local to strategy k; in this case, the parameter ϵ can be regarded also as a 'generalization' parameter, as players generalize from their recent experience to similar strategies.

Figure 6 provides an illustration of the performance of the three models relative to human subject data from a simple matching pennies experiment conducted by Ochs (1995). [Figure 6 here]. This game is of the form

$$\begin{array}{cccc} & & \multicolumn{2}{c}{Player 2} \\ & & A2 & B2 \\ Player & A1 & x,0 & 0,1 \\ 1 & B1 & 0,1 & 1,0 \end{array}$$

where x is a payoff parameter that takes on different values in three treatments (x=1, 4 or 9). The unique mixed strategy equilibrium calls for player 1 to play A1 with probability .5, and player 2 to play A2 with probability $1/(1+x)$; these Nash equilibrium point predictions are illustrated in the figure, which shows results for the three different versions of the game (according to the value of x). The data shown in Figure 6 are the aggregate frequencies with which the two players play actions A over repeated plays of the game. The first column gives the experimental data, columns 2-3 give the results of the 1 and 3 parameter reinforcement learning models, while column 4 gives the result from a fictitious play-like learning model. The relatively better fit of the three-parameter model is determined on the basis of the deviation of the path of the experimental data from the path of the simulated data. Erev and Roth suggest that the success of reinforcement learning in predicting experimental data over Nash equilibrium point predictions is owing to the inductive, real–time nature of these algorithms as opposed to the deductive approach of game theory, with its assumptions of full rationality and common knowledge.

Other variants of reinforcement learning have been proposed with the aim of better explaining experimental data. Sarin and Vahid (1999, 2001), for instance, propose a simple deterministic reinforcement-type model where agents have "subjective assessments", $q_j(t)$, for each of the $j = 1, 2, ...N$ possible strategies. As in Roth and Erev's model, an agent's subjective assessment of strategy j gets updated only when strategy j is played: $q_j(t+1) = (1-\phi)q_j(t) + \phi\pi_j(t)$, where $\pi_j(t)$ is the payoff to strategy j at time t, and ϕ is the forgetting factor and sole parameter of their model. The main difference between Sarin and Vahid's model and Roth and Erev's is that the strategy an agent chooses at time t in Sarin and Vahid's model is the strategy with the maximum subjective assessment through period $t-1$. Thus, in Sarin and Vahid's model, agents are acting more like optimizers than in the probabilistic choice framework of Roth and Erev. Sarin and Vahid show that their one parameter model often performs well and sometimes better than Roth and Erev's 1 or 3–parameter, probabilistic choice reinforcement learning models in the same games that Erev and Roth (1998) explore.

Duffy and Feltovich (1999) modify Roth and Erev's (1995) model to capture the possibility that agents learn not only from their own experience, but also from the experience of other agents. Specifically, they imagine an environment where agent i plays a strategy r and learns his payoff in period t, $\pi^i_r(t)$ but also observes the strategy s played by another player j (of the same type as i) in period t and the payoff that player earned from playing strategy s, $\pi^j_s(t)$. Player i updates his propensity to play strategy r in the same manner as Roth and Erev, (with $\phi = \epsilon = 0$) but also updates his propensity to play strategy s: $q^i_s(t+1) = q^i_s(t) + \beta\pi^j_s(t)$, where $\beta \geq 0$ is the weight given to observed payoffs, or "second–hand" experience. Duffy and Feltovich set $\beta = .50$ and simulate behavior in two of the games studied in Roth and Erev (1995), the best–shot game and the ultimatum game. They then test their simulation predictions by conducting an experiment with human subjects; their reinforcement-based model of the effect of observation of others provides a very good prediction of the role that observation of others' actions and payoffs plays in the experiment.

Another modification of reinforcement learning is to suppose that agents have certain "aspiration levels" in payoff terms that they are trying to achieve. This idea has a long history in economics dating back to Simon's (1955) notion of satisficing. Aspiration learning has recently been resuscitated in game theory, e.g. by Karandikar, Mookherjee, Ray and Vega-Redondo (1998) and Börgers and Sarin (2000) among others. Bendor, Mookherjee and Ray (2001) provide an overview and additional references. The reinforcement learning models discussed above can be viewed as ones where a player's period aspiration level is constant and less than or equal to the minimum payoff a player earns from playing any action in the given strategy set, so that the aspiration level plays no role in learning behavior. More generally, one might imagine that an agent's aspiration level evolves along with the agent's probabilistic choice of strategies (or propensities), and this aspiration level lies above the minimum possible payoff. Thus, in aspiration-based reinforcement learning models, the state space is enlarged to include a player's aspiration level in period t, $a^i(t)$. Suppose player i chooses strategy j in period t yielding a payoff of $\pi^i_j(t)$. If $\pi^i_j(t) \geq a^i(t)$, then player i's propensity to play strategy j in subsequent periods is assumed to be (weakly) higher than before; precisely how this is modeled varies somewhat in the literature, but the end result is the same: i's probability of playing strategy $j$satisfies $p^i_j(t+1) \geq p^i_j(t)$. On the other hand, if $\pi^i_j(t) < a^i(t)$, then $p^i_j(t+1) < p^i_j(t)$. Finally, aspirations evolve according to:

$$a^i_t = \lambda a^i_t + (1-\lambda)\pi^i_j(t),$$

where $\lambda \in (0,1)$. This adjustment rule captures the idea that aspirations vary with an agent's history of play. The initial aspiration level a_0 as with the initial probabilities for choosing actions, are assumed to be exogenously given. Karandikar et al. (1998) also add a small noise term to the aspiration updating equation representing trembles. They show, for a class of 2×2 games that includes the prisoner's dilemma, that if these trembles are small, and aspiration updating is slow (λ is close to 1) that in the long-run, both players are cooperating most of the time.

There is some experimental evidence in support of aspiration learning. Bereby–Meyer and Erev (1998) studied behavior in a binary choice game where the probabilities of achieving a 'success' were exogenously fixed at .7 for choice 1 and .3 for choice 2. In one treatment, subject payoffs were set at 2 for a success and -2 for a failure, while in another treatment, the payoffs were 4 for a success and 0 for a failure, amounting to an addition of 2 to the payoffs in the first case. They found that learning of the optimal choice of strategies (choice 1) was significantly reduced when the payoffs were (4,0) relative to the case where the payoffs were (2,-2). Erev, Bereby–Meyer and Roth (1999) explain this result by presenting an adjustable reference point reinforcement learning model. In place of the assumption that $R(x) = x - x_{\text{min}}$ in the Roth-Erev model, they propose that $R(x,t) = x(t) - \rho(t)$, and let the reference point, $\rho(t)$ be a weighted average of the past reference point and current payoffs, where the weights depend on the difference between the payoff and the reference point; if payoffs are highly variable relative to the reference point, learning is slower than if payoffs are less variable; this is simply another version of aspiration learning. They report that this model tracks the difference in the experimental findings rather well.

Huck et al. (2002) find evidence of aspiration learning in a laboratory oligopoly experiment. They test the theoretical proposition that bilateral mergers in oligopoly markets with $n > 2$ firms, homogeneous goods and constant returns to scale are unprofitable; the profit share of the merged firm, $1/n-1$ is less than the total share of the two firms prior to the merger $2/n$ ($1/n$ each). In the experiment, $n > 2$ subjects make quantity decisions in a Cournot game and midway through a session, two of the subjects combine decision-making as a merged firm. The authors report that, contrary to theory, the subjects in the role of the merged firm produce significantly more output than the other unmerged firms and come close to sustaining total profit levels they would have achieved as unmerged firms. The authors argue that pre-merger aspiration-levels cause merged firms to increase output with the aim of maintaining total profits and the other firms acquiesce by reducing their output. They connect this finding with Cyert and March's (1956) observation that oligopoly firms are guided by "an acceptable–level profit norm" that is a function of market history.

Varieties of reinforcement learning algorithms have become a mainstay of agent-based modeling, perhaps because they accord with Axelrod's KISS principle. Other attractive features are the low level of history–dependent rationality, and relatively few parameters. Examples of the use of reinforcement learning in agent-based models are commonplace. Epstein and Axtell (1996) use several variants of reinforcement learning in their Sugarscape model. Nicolaisen, Petrov and Tesfatsion (2001) use Roth-Erev-type reinforcement learning to model buyer and seller price–quantity decisions in a computational model of the wholesale electricity market. Pemantle and Skyrms (2003) use reinforcement learning to study how groups of players play games in endogenously formed social networks. Franke (2003) uses reinforcement learning to study Arthur's (1994) El Farol Bar problem; Kutschinski, Uthmann and Polani (2003) use a reinforcement learning model to study buyer search and seller price setting behavior in a competitive market with induced demand and supply schedules. Bendor et al. (2003) use a reinforcement learning model with endogenous aspirations to model voter turn-out. Finally, Erev and Barron (2003) apply reinforcement learning to cognitive strategies, e.g., loss avoidance, hill–climbing, rather than to the direct strategies available to agents in simple, repeated decision problems.

There is also a parallel and much more voluminous literature on reinforcement learning in the machine learning literature. See, e.g., Kaelbling et al. (1996) and Sutton and Barto (1998) for surveys. A popular reinforcement learning model in this literature is Q–learning (Watkins, 1989), which is closely related to Bellman's approach to dynamic programming, but differs from the latter in being much less informationally demanding, e.g. the agent need not know the period payoff or state transition functions. (See, e.g. Mitchell (1997) for a good introduction the topic). Q–learning algorithms involve on-line estimation of an evaluation function, denoted $Q(s, a)$, representing the maximum expected discounted sum of future payoffs the agent earns from taking action a in state s. Starting from some random initialization of values, estimation of the Q function occurs in real–time using the history of states and payoffs earned by the agent from action choices in those states. To determine the action chosen, a probabilistic choice rule is used: actions with higher Q-values for the given state s and the current approximation of the Q–function, are more likely to be chosen than actions with lower Q-values. Thus, the main difference between Q–learning and the reinforcement-learning models studied by economists is that Q–learners are learning an *evaluation function* mapping from states to actions, analogous to the policy function of dynamic programming. An advantage of Q–learning over reinforcement learning algorithms studied by economists is that convergence results for Q–learning can be proved under certain assumptions, e.g. for simple Markov-decision processes. Surprisingly, the predictions of Q–learning models have yet to be compared with data from controlled laboratory experiments with human subjects – a good topic for future research.

3.2 Belief-based learning

The primary difference between belief-based learning algorithms and reinforcement learning algorithms is that in belief-learning models, players recognize they are playing a game or participating in a market with other players, and form beliefs about the likely play of these other players. Their choice of strategy is then a best response to their beliefs. By contrast, reinforcement learners do not form beliefs about other players and need not even realize that they are playing a game or participating in a market with others. Belief-based learning models range from naive, Cournot-type learning to slightly more sophisticated "fictitious play," to fully rational, Bayesian learning. Here we discuss the first two types of belief learning models.

Fictitious play was proposed by Brown (1951) as a model of how players form beliefs and best respond to them in two-person zero sum games. Fictitious play was originally proposed as a means of determining the value of a game; indeed, Robinson (1951) shows that fictitious play converges to equilibrium in 2×2 zero sum games, though Shapley shows via a counterexample that this result does not hold in more general games. Subsequently, fictitious play has come to serve as a model of boundedly rational learning: players form beliefs about their opponents based on the historical frequency of their opponent's actions choices and play myopic best responses to these beliefs; the best responses are myopic because agents do not anticipate that their opponent is behaving similarly toward them.

Cheung and Friedman (1997) propose a one–parameter class of learning rules that yields Cournot and fictitious play learning as special cases and thus serves to compactly illustrate the main difference between the two approaches. They suppose there are $i = 1, 2, ..N$ players, each of whom chooses an action a_i from the set of possible actions, A, in each period. Player i's payoff function is $\pi(a_i, s^{-i})$, where s^{-i} is a state vector representing the distribution of action choices chosen by all of i's opponents. It is assumed that each player i discounts past states using a constant discount factor, γ_i, and possesses some initial prior, $s^{-i}(1)$. Player i's belief about the state that will prevail in periods $t = 1, 2, ...$ is given by:

$$\hat{s}^{-i}(t+1) = \frac{s^{-i}(t) + \sum_{k=1}^{t-1} \gamma_i^k s^{-i}(t-k)}{1 + \sum_{k=1}^{t-1} \gamma_i^k}.$$

Cournot (naive) belief learning results from setting $\gamma_i = 0$ for all i; in this case, players hold the naive belief that $\hat{s}^{-i}(t+1) = s^{-i}(t)$. Fictitious play belief learning results from setting $\gamma_i = 1$ for all i; in this case, players' beliefs about the current state are simply the average of all past observed states. Weighted average, *adaptive* belief learning results from setting $0 < \gamma_i < 1$.[14] Given beliefs, a player's decision is to choose $a_i \in A$ so as to maximize his expected payoff (i.e., $\max_{a_i \in A} \pi(a_i,\ \hat{s}^{-i})$).

Consider by way of illustration, the class of 2 player, binary choice games that have been widely studied in the experimental literature. Let the 2×2 payoff matrix be given by $M = (m_{ij})$, and let us assign a '1' to the choice of action 1 and a '0' to the choice of action 2. With a single opponent per period, $s^{-i}(t) \in \{0,\ 1\}$ and $\hat{s}^{-i}(t) \in [0, 1]$ represents player i's belief about the likelihood that his opponent will play action 1 in period t.[15] Player i evaluates the expected payoff differential from choosing action 1 over action 2:

$$r_{i1} = R(\hat{s}^{-i}(t)) = (1, -1)M(\hat{s}^{-i}(t), 1 - \hat{s}^{-i}(t))'.$$

A deterministic best response in the binary choice game is to choose action 1 if $R(\hat{s}^{-i}(t)) > 0$ and to choose action 2 if $R(\hat{s}^{-i}(t)) < 0$. Some kind of tie-breaking rule is needed for the special case where $R(\hat{s}^{-i}(t)) = 0$. As Fudenberg and Levine (1998) note, fictitious play ($\gamma = 1$) is a form of Bayesian learning in the special case where a player's prior beliefs over the distribution of opponent strategies is Dirichlet.

As was the case under reinforcement learning, researchers examining the predictions of Cournot or fictitious play belief learning have added some kind of noise to the deterministic best response. Boylan and El-Gamal (1993) propose that agents play the deterministic best response with probability $1 - \epsilon$, and any of the available actions $a \in A$ with probability ϵ/A.

Fudenberg and Levine (1998) propose a stochastic approximation to deterministic fictitious play – smooth fictitious play – which can be implemented, as in Cheung and Friedman (1997), through the use of the logistic function:

$$p_{ij}(t) = \frac{1}{1 + e^{-x_i(t)}}, \text{ where } x_i(t) = \alpha_i + \beta_i r_{ij}(t),$$

[14] Other, less plausible possibilities include $\gamma > 1$, so that the past is given more weight than the present and $\gamma_i < 0$, which implies cycling.

[15] More generally, if player i faces up to $n \leq N - 1$ opponents in a binary action game, then $s^{-i}(t) = n^{-1} \sum_{j=1}^{n} I(a_j)$, where $I(j) = 1$ if $a_j = 1$ and $I(a_j) = 0$ otherwise.

where α_i is an individual specific fixed effect indicating individual i's bias for action j ($\alpha_i = 0$ reveals an unbiased choice) and β_i representing the sensitivity of choices to expected payoff differentials.

These stochastic versions of fictitious play have several advantages over deterministic fictitious play. First, they do not imply that behavior switches dramatically with small changes in the data agents use to form beliefs. Second, insisting that strategies remain probabilistic has certain advantages, e.g. when agents have achieved near convergence to a mixed strategy equilibrium and need to keep their opponent guessing even though the differences in utility from the various actions may be quite small. (See Fudenberg and Levine (1998) for a further discussion).

Boylan and El Gamal (1993) use a Bayesian approach to assess the likelihood that behavior in 9 different matrix game experiments (conducted by other researchers) is consistent with either the noisy-Cournot or the noisy-fictitious play hypothesis. They find that for some games, the Cournot belief hypothesis is favored while for other games the fictitious play hypothesis is favored. Their over all assessment of the relative validity of the two learning hypotheses is that fictitious play describes the experimental data better than Cournot learning.

Cheung and Friedman (1997) estimate their three parameter model (α, β, γ) on data from several different bimatrix games. Median estimates of α, β and γ are all significantly positive; the finding that $\gamma > 0$ rules out the Cournot belief hypothesis. Further they report they can reject the hypothesis that $\gamma = 1$(fictitious play). Indeed, their estimates of γ always lie between 0 and 1 indicating that subjects' belief updating process is neither Cournot or fictitious play, but is instead approximated best by some adaptive intermediate case.

In addition to asking which belief-based learning model best predicts experimental data, one can also explore the empirical validity of the belief formation process associated with these belief-based models. This can be simply accomplished by asking subjects to state, prior to play of the game, their beliefs about their opponent's play and comparing these stated beliefs with those predicted by belief-based learning models. Nyarko and Schotter (2002) have carried out such an exercise in a simple 2×2 matrix game where the unique Nash equilibrium prediction is in mixed strategies. The two strategies were labeled Green and Red, and the equilibrium calls on both players to play Green (Red) with probability .4 (.6). Nyarko and Schotter asked subjects to state the probability with which they thought their opponent would play Green prior to the play of each round. Subjects' compensation was determined in part by the accuracy of their stated beliefs and in part by the payoffs they received from playing the game.

Figure 7 plots stated beliefs against those predicted by fictitious play for a "typical subject" in Nyarko and Schotter's experiment. [Figure 7 here]. As is apparent, the variance in subject beliefs is much greater than predicted by fictitious play, and the differences do not decrease with experience. A similar difference is found in a comparison of the subjects' beliefs with Cournot beliefs. Nyarko and Schotter further conclude that best responses to subjects' *stated* beliefs provide a better account of the path of actions chosen by subjects than does reinforcement or a hybrid belief-reinforcement model discussed below. This evidence suggests both that subjects are following some kind of belief-learning process and that a good model of that belief formation process has yet to be developed.

Belief-based learning models also make strong predictions regarding equilibrium selection in environments with multiple, Pareto rankable equilibria. Essentially, belief-based models predict that if the initial conditions lie in the domain of attraction of a particular equilibria under the belief learning dynamic, then, with experience, agents will learn over time to coordinate on that equilibrium, regardless of its efficiency. This hypothesis has been experimentally tested by Van Huyck et al. (1997) and Battalio et al. (2001) in the context of simple coordination games where the domain of attraction of the two symmetric pure strategy equilibria is defined by the best response separatrix. Van Huyck et al. (1997) show that both Cournot and fictitious play learning dynamics predict different equilibrium outcomes depending on initial conditions in a median effort game (involving strategic complementarities), and their experimental findings are remarkably accurate on this score. If the initial condition (median effort) lies in the domain of attraction of the unique, payoff-dominant equilibrium, subjects subsequently coordinate on that equilibrium, otherwise they coordinate on the other symmetric Nash equilibrium. As Van Huyck et al. point out, this behavior is very different from deductive equilibrium selection principles, which might involve, for instance, calculation of all equilibria and selection of the payoff dominant one.

The use of belief-based learning models by economists is not limited to normal form games. Varieties of belief-based learning models have also been used to study bid and ask behavior in the double auction.[16] Gjerstad and Dickhaut (1998) provide a particularly elegant characterization of the DA and propose heuristic rules by which buyers and sellers assess and update the probability that their bids or asks will be accepted, given market history. Using these beliefs together with private information on valuations and costs, individual buyers or sellers propose bids or asks that maximize their (myopic) expected surplus. The main parameter in their model is the length of memory that players use in calculating probabilities. Using a stricter convergence criterion than Gode and Sunder adopt, Gjerstad and Dickhaut show via simulations that their heuristic belief-learning model can more reliably achieve convergence to competitive equilibrium than Gode and Sunder's ZI–C model, and the belief-learning model provides a better fit to the aggregate human subject data as well. Indeed, in their chapter in this handbook, Mackie-Mason and Wellman (2005) argue that this heuristic belief-learning model represents the best agent-based model of the DA. Still, the fit of this belief-learning model to *individual* human subject behavior remains to be examined.

Belief-based learning models are less common in the agent-based literature than are reinforcement learning models, perhaps for the simple reason that belief-based models require that agents possess more memory (e.g. the histories of their opponents). Still, some versions of belief–based learning can be found see, e.g. Kandori et al. (1993), Young (1993, 1998); naive Cournot best response behavior is also found see, e.g. Ellison (1993) or Morris (2000).

3.3 Comparisons of Reinforcement and Belief-Based Learning

A large literature is devoted to testing whether simple reinforcement or more complicated belief-based learning algorithms better characterize experimental data from a wide variety of different games. In addition to the papers of Roth and Erev and Cheung and Friedman mentioned above, other papers comparing versions of these two approaches to learning include Mookherjee and Sopher (1994, 1997), Camerer and Ho (1999), Feltovich (2000), Salmon (2001), Blume et al. (2002), Stahl (1999) and Haruvy and Stahl (2004) among others. In making these comparisons, researchers have adopted some kind of goodness–of–fit metric or made use of an econometric estimator to assess the fit of various candidate learning models to experimental data.

The findings from this literature are varied, but several conclusions appear to have wide support. First, the evidence is very strong that either reinforcement or belief-based learning models are better predictors of human subject behavior than are the static Nash equilibrium point predictions. This is strong evidence in favor of the bottom–up, inductive reasoning approaches used by ACE researchers as opposed to the top–down, forward–looking, deductive reasoning of fully rational players that gives rise to those equilibrium point predictions. Second, in the simple games that experimentalists have studied, reinforcement and belief-based learning models do not yield predictions that are all that distinct from one another and so identifying which rule performs well across a variety of different games leads to murky outcomes that appear sensitive to various particulars of the datasets or games examined (Feltovich (2000), Salmon (2001)). Given the lack of a clear bias in favor of reinforcement or

[16]Early efforts include Friedman (1991) and Easley and Ledyard (1993).

belief-based approaches over a wide variety of games, a natural approach is to adopt a hybrid model that allows for both reinforcement and belief-based learning as special cases, as well as mixtures of both. The hybrid modelling approach is taken e.g., by Camerer and Ho (1999), and discussed in Brenner's (2005) chapter. While this approach has had some success in explaining data from human subject experiments (see Camerer (2003) for an extensive and detailed assessment), the additional complexity of such models, e.g., more parameters to calibrate, may make this approach less appealing to ACE researchers.[17] Third, there is some evidence that if subjects' information is restricted to their own histories of play, that reinforcement learning models perform slightly better than belief-based learning models that use data on opponent's histories that was unavailable to subjects. Analogously, in environments where data on opponent's histories was made available, players appear to condition their expectations, in part, on those histories, in line with the predictions of belief-based models. (Blume et al. 2002). These findings are not so surprising, and, indeed, simply confirm that players use histories to form expectations. Finally, there is some evidence that the complexity of the game, the manner in which players are matched and the length of play are all important factors in the accuracy of learning models in predicting the play of human subjects.

On the latter point, much of the observed differences in the two approaches to modeling learning may be tied up with the relatively small periods of time over which individual human subject experiments are conducted. While experimentalists often give their subjects repeated experience with a game or decision, concerns about subject boredom or the salience of participation payments severely restrict the length of the time series that can be generated in the laboratory for any individual subject. By contrast, ACE researchers do not feel bound by such considerations, and think nothing of simulating their models out for very long periods of time. Asymptotically, the behavior of reinforcement and belief-based models may not be all that different. Hopkins (2002) shows that both reinforcement learning and stochastic fictitious play can be viewed as noisy versions of replicator dynamics (discussed later in section 4.1), and that the asymptotic predictions of these two models may be the same; roughly speaking if an equilibrium is locally stable under stochastic fictitious play, then the same holds true under reinforcement learning. Duffy and Hopkins (2005) conduct experiments with a longer than typical number of repetitions under various information conditions in an effort to test this prediction and find that it has some, qualified support. An implication of these findings for ACE researchers is that the kind of learning rule that agents are endowed with may not be of such great importance if the research interest lies in the long-run behavior of the agent-based system.

3.4 Summary

Unlike ZI agent models, reinforcement and belief-based learning models presume that agents have some memory. These models of inductive reasoning have been primarily studied in the context of simple two player games. Reinforcement learners condition their actions on their own histories of play and abide by the principle that actions that have yielded relatively high (low) payoffs in the past are more (less) likely to be played in subsequent periods. Belief-based learning models assume that players have history dependent beliefs over the actions their opponents are likely to play, and they choose actions that are myopic best responses to these beliefs. While there is no guarantee that either type of learning model converges to an equilibrium, these models have nevertheless proven useful in tracking the behavior of human subjects in controlled laboratory settings.

Reinforcement learning models have been widely used in the agent-based literature, perhaps for the simple reason that they require only information on an individual's own history (payoffs and actions). In complex, multi-agent settings, this parsimony of information may be an important consideration in

[17] See, however, a simpler, one-parameter version of their model given in Ho, Camerer and Chong (2002).

the modeling of agent learning. On the other hand, in settings with just a few agents, and especially in settings where agents interact with one another repeatedly, a belief-based learning approach may be more appropriate. Indeed, the available experimental evidence suggests that agents do condition their actions on both their own history of play and, when available, on information about the play of their opponents. However, the manner in which they do this does not appear to be strictly consistent with either reinforcement or belief-based learning models. Asymptotically, there may be little difference between the two approaches.

4 Evolutionary Algorithms as Models of Agent Behavior

In addition to directed random (ZI-agent) searches and individual learning approaches, agent-based researchers have used a variety of different *evolutionary algorithms* to characterize the behavior of populations of heterogenous, interacting, boundedly rational agents facing various economic decisions. Examples include replicator dynamics, genetic algorithms, classifier systems and genetic programming. These evolutionary algorithms differ from the learning processes considered so far in several respects. First, evolutionary algorithms were designed to mimic naturally occurring, biological processes. Not surprisingly, these algorithms can be difficult for social scientists to interpret and for experimentalists to test in the laboratory. Second, these methods are *population–based*, which is to say that the fitness of a particular individual or strategy (the distinction becomes blurred in this literature) is based on its performance relative to a certain population of individuals (or strategies). Thus, these algorithms presume that fitness values across individuals/strategies are readily and immediately available for comparison purposes; in this regard, they can be viewed as the most complex class of algorithms (or least decentralized) in the set of approaches considered in this chapter. Third, as with ZI or reinforcement learning, evolutionary algorithms are not belief-based; players are not aware that they are playing a game against other players and do not act strategically in any way. Fourth, some evolutionary algorithms, e.g., genetic algorithms and genetic programming, are employed in environments where strategies or equilibrium policy functions cannot be characterized analytically. This (alternative) use of evolutionary algorithms is owing to the performance of these algorithms as function optimizers in complex landscapes; indeed, genetic algorithms were developed for precisely this purpose. Finally, evolutionary algorithms may or may not be well-suited to modeling economic decision-making. Evolution is often a slow process and so algorithms that mimic this process tend to work best on an unchanging landscape. However, economic systems are often modeled as state dependent, and may also be subject to temporary shocks or more permanent structural shifts. In such environments, the performance of evolutionary algorithms may be degraded relative to the less volatile (natural) landscapes for which they were developed.

Despite these potential problems and shortcomings, evolutionary algorithms are widely used by agent-based modelers. By contrast with the other agent-based approaches we have discussed, evolutionary algorithms have not been developed or adapted to explain data from economic decision-making experiments. For the most part, the opposite has occurred; agent-based researchers have sought to validate the predictions of evolutionary algorithms by conducting experiments with human subjects placed in the same environments. In certain cases, the experimental environment has been modified to better approximate the evolutionary environment! These comparisons have met with some success, but as I will argue, some difficulties of interpretation remain, for example, the question of the appropriate time-frame for comparisons. It may simply be that evolutionary algorithms cannot be adequately tested using human subject experiments.

4.1 Replicator Dynamics

Replicator dynamics comprise the simplest class of evolutionary algorithms that economists have used to model the behavior of populations of players. See Hofbauer and Sigmund (1988, 1998) for a complete treatment. These models presume that the set of strategies (or phenotypes) does not evolve, and that reproduction is asexual. The assumption of a small strategy space is most likely to be satisfied in simple games, and so it is not surprising that replicator dynamics have mainly been employed by game theorists.

To understand how replicator dynamics work, consider a game with N strategies, and let $s(t) \equiv (s_i(t))_{i=1,2,..N}$ be a vector representing the proportions of the N strategies in the population at time t; $\sum_i s_i(t) = 1$ for all t. The $N \times N$ payoff matrix $M = (m_{ij})$ here represents the payoff earned by each strategy in the population when matched against every other strategy, including itself. For illustration purposes, we focus here in the simplest cast where M is symmetric, known as the one-population model. The fitness of strategy i at time t is given by $M_i s(t)$, where M_i denotes the row of the payoff matrix corresponding to strategy i. The idea of assessing how a strategy fares against the entire population of strategies is what Maynard Smith termed "playing the field." The deterministic replicator dynamic posits that strategy i's representation in the population be updated as follows:

$$s_i(t+1) = \frac{s_i(t) M_i s(t)}{s'(t) M s(t)},$$

where the denominator can be interpreted as the average fitness level in the entire population of strategies, including strategy i. The idea of the replicator dynamic is that strategies with above average fitness see their proportion in the population increase while those with below average fitness see their proportion in the population decrease. Further, if $\hat{s}$ is a Nash equilibrium of the symmetric game M, then it is also a fixed point of the replicator dynamic. In the deterministic version of the replicator dynamic, the proportion of certain strategies can go to zero, i.e. extinction is possible. A stochastic version of replicator dynamics due to Foster and Young (1990) eliminates extinction, and can have quite different limiting dynamics than the deterministic version.

Friedman (1996) and Cheung and Friedman (1998) have examined the predictions of replicator dynamics using data from human subject experiments. Friedman studies the predictions of the replicator dynamic for equilibrium stability, and Cheung and Friedman compare replicator dynamic predictions with that of the individual, belief-based, stochastic fictitious play learning algorithm. Most of the games they study are two player, binary choice games with a unique Nash equilibrium in either mixed or pure strategies. In such games, the state, $s(t)' = (s_1(t), (1 - s_1(t))$, and the replicator dynamic for strategy s_1 is written as:

$$\frac{\Delta s_1(t+1)}{s_1(t)} = \beta \frac{[M_1 s(t) - s(t)' M s(t)]}{s(t)' M s(t)},$$

where $\beta > 0$ represents an adjustment parameter, and $\Delta s_1(t+1) = s_1(t+1) - s_1(t)$. Cheung and Friedman omit the denominator on the right hand side, $s(t)'Ms(t)$, which serves as a normalization device ensuring that proportions sum up to one; in the binary choice case this device is unnecessary, and furthermore, Cheung and Friedman report that the unnormalized version fits the data better.

In their experimental design, these authors make some accommodation for the "playing the field" nature of the replicator dynamic; in their "mean matching" treatment, each player is matched against all other players, receiving the average payoff from his choice of action against that of all others. The other matching treatment is the standard, random pairwise matching protocol. While game theory would treat these two environments very differently, with the first corresponding to an n-player repeated game and the second to a two-player, one-shot game, the only difference under the replicator dynamic lies in the greater variance in payoffs that players receive in the random pairwise matching protocol. Friedman and Cheung and Friedman are careful to address issues concerning group size, the length of play of a single game, and of the information that players receive, all of which are important to approximating the environment for which the replicator dynamic was devised.

Cheung and Friedman (1998) use experimental data from the two binary choice games they study to estimate the linear equation:

$$\Delta s_1(t+1)/s_1(t) = \alpha + \beta[M_1 s(t) - s(t)' M s(t)] + \gamma d_t + \epsilon$$

where $d_t = I(t)[M_1 s(t) - s(t)' M s(t)]$, $I(t) = 1$ if the mean matching treatment was used, and ϵ is an error term. They report that α is typically significantly different from zero, implying a persistent bias from the pure replicator dynamic, and that β is significantly positive as is γ. The latter finding suggests that the mean matching protocol aids in the speed of adjustment relative to random pairings. In a head-to-head comparison of the explanatory power of the replicator dynamic versus an individual, belief learning model – the three parameter weighted fictitious play model of Cheung and Friedman (1997) described in section 3.2 – Cheung and Friedman report that over the two games they study, the belief learning model outperforms the replicator dynamic, where performance is measured by either the root mean squared errors or the mean absolute deviations computed from the three parameter belief-learning or replicator dynamic model.

This finding suggests that there is some value to thinking of human players as playing best responses to beliefs about their opponents' actions rather than thinking of them as playing a game against nature. On the other hand, it is less clear that Cheung and Friedman have successfully implemented the evolutionary game environment germane to the use of replicator dynamics or that such an environment could be implemented in the laboratory, where budget and time constraints limit the number of subjects and replications of a treatment that are possible. Further work reconciling the replicator dynamic with human learning processes is needed.

4.2 Genetic Algorithms

Genetic algorithms (GAs) have been widely used by economists to model learning by populations of heterogenous, adaptive agents especially following Sargent's (1993) encouraging assessment and the subsequent use of GAs by his student, Jasmina Arifovic. These algorithms differ from replicator dynamics in that they allow for the development of new strategies or decisions that may not have been included in the initial population. As such, they are efficient sampling methods most appropriate to large decision or strategy spaces.

Indeed, genetic algorithms, originally developed by Holland (1975), are stochastic, directed search algorithms based on principles of population biology.[18] These algorithms have been demonstrated to perform well in large or "rugged" search spaces where classical methods, e.g. grid search or gradient descent, are either inefficient or susceptible to getting stuck at local optima. While there is wide variation in the specific details of genetic algorithms, there are some general principles and procedures that are regarded as relatively standard. First, the researcher must specify the objective function of the genetic algorithm search, the parameter values that will be used to maximize (or minimize) that objective, and the range of admissible parameter values allowed in the search for an optimum. Second, vectors of parameters, representing candidate solutions are encoded as strings of finite length L. The strings are intended to mimic chromosomes, with the individual elements of a string representing genes; hence the name genetic algorithm. In the earliest implementation of genetic algorithms (e.g.,

[18] For a complete treatment of genetic algorithms see, e.g., Goldberg (1989) or Michalewicz (1996). Dawid (1999a) provides a thorough discussion of genetic algorithms as applied to economic problems. See also Sargent (1993) and Judd (1998).

Goldberg (1989)), parameters were encoded using the binary {0,1} alphabet, and much of the theory of genetic algorithms as function optimizers is developed for binary encodings. However, more recently, researchers have made use of real–valued, character, or tree encodings in place of traditional binary encodings. Researchers typically work with a population of strings of some fixed size, N. Third, the performance of each string in the population is evaluated using the objective criterion – this is the string's fitness. Fourth, a new generation of N strings is determined using operations that mimic natural selection and naturally occurring biological processes.

The first step in a genetic algorithm, known as selection, is to randomly select N strings from the existing population in such a way that the fitness of the N randomly selected strings is on average higher than the average fitness of the population from which they were chosen. This selection operation can be accomplished in many ways, including the biased roulette wheel selection mechanism originally proposed by Holland, in which the likelihood of selecting a string is proportional to its relative population-wide fitness or other methods e.g. binary tournaments or rank order lists. The selection operation is intended to mimic Darwinian survival-of-the-fittest. Once a new set of N strings has been selected, these strings undergo two main biological operations that mimic genetic inheritance. The first, crossover, typically involves randomly pairing strings and, with some probability, p_c, randomly cutting the two strings at one or more points and swapping elements. Once crossover is applied to all strings, a second operator, mutation is applied, which involves randomly changing each element in a string with a (small) probability p_m, to some other value; in the case of binary strings, a '0' is flipped to a '1' and vice versa. After these operations are complete, the new generation of N strings is evaluated in fitness terms and the process of choosing a new generation begins again. The genetic algorithm is terminated after a set number of generations, G, or after some tolerance criterion based on the objective function has been satisfied. Some pseudo-code for a genetic algorithm is given in Figure 8. [Figure 8 here].

The main theoretical result for genetic algorithms is known as the schema theorem (Holland (1975)). The idea of a schema can be understood by the addition of a don't care character, *, to the binary alphabet that is typically used to encode strings. A schema is a template characterizing a set of chromosomes. For example, the schema of length 5, (*101*) characterizes the set of chromosomes {(11011), (11010), (01011), (11010) }. The order of a schema is the number of fixed positions; e.g., the order of the schema in our example is 3. The schema theorem (proved, e.g. in Goldberg (1989)) states that low-order, above-average (below-average) schema appear exponentially more often (less often) in subsequent generations of a genetic algorithm. This theorem follows directly from the operation of fitness-proportional selection. These low-order schema are sometimes referred to as "building blocks." Crossover plays the role of introducing new schemata and mutation also contributes to variability while at the same time preventing premature convergence to local optima.

How are the genetic operators to be interpreted when applied to economic systems? Several authors, e.g., Arifovic (1996), Bullard and Duffy (1998) Dawid (1999a) Riechmann (1999, 2001ab), have offered interpretations. One can think of the individual strings as representing the strategies/decisions of individual agents, so that the GA is made up of many interacting agents. Alternatively, one can imagine there is a single agent with the individual strings of the GA representing different decisions/strategies that agent might adopt. The selection operation is perhaps the easiest to defend; this operator just insures that agents or decisions that have worked well in the past are more likely to be chosen in the future while decisions that have fared poorly are more likely to be discarded. This probabilistic choice of decisions based on relative payoff or fitness success is similar to stochastic reinforcement learning or stochastic replicator dynamics. The turnover of population need not be interpreted so literally as one of birth and death; instead it can be interpreted as a turnover of decisions or ideas among players who are long-lived. The crossover/recombination operator is easiest to interpret if the population of strings is viewed as representing individual agents. In that case, crossover can be thought of as communication between pairs of agents, who exchange bits and piece of ideas, though the population as a whole retains core principles (low-order schema) that have yielded high payoffs in the past. Finally, the mutation operator can be viewed as representing trembles or experimentation.

A further issue concerns the choice of GA parameters: the number of strings, N, the string length, the mutation and crossover parameters, p_c, p_m, etc. Here, the practice has been to adopt parameterizations that computer scientists have found to perform well on test suites of difficult static optimization problems. These optimization problems are not ones that are so applicable to the dynamic settings studied by economists, and so further research into this issue would be of some value.

What about the external validity of simulations using genetic algorithms? Arifovic (1994) was the first to directly compare simulations of a genetic algorithm with the behavior of human subjects in a controlled laboratory experiment.[19] The economic environment studied was a textbook version of Ezekiel's (1938) "Cobweb" model of demand and supply for a single good. In this model, market demand in period t is a decreasing, linear function of current period price, p_t, while market supply in period t is an increasing, linear function of the market price that suppliers expected in period $t-1$ would prevail in period t, $E_{t-1}p_t$; the latter assumption captures the notion that it takes time (one-period) to produce the good, and makes the model dynamic. Arifovic followed experimental researchers, Carlson (1968) and Wellford (1989), who adopted Ezekiel's assumption of naive and homogeneous expectations, i.e. $E_{t-1}p_t = p_{t-1}$ as a benchmark assumption for expectation formation; in that case, the equilibrium is stable (unstable) if the ratio of the slope of the supply curve to the slope of the demand curve, in absolute value, is less than (greater than) unity. Bray and Savin (1986) have shown in a stochastic version of the linear cobweb model that adaptive learners, running regressions of prices on past prices, can learn the equilibrium price level in the stable case but not in the unstable case.[20] By contrast, a main finding of the experimental studies was that groups of subjects generally converged to a neighborhood of the unique equilibrium regardless of whether that equilibrium was stable or unstable under the naive expectations assumption. However, the variance of quantities or prices was much greater and more persistent in the unstable case as compared with the stable case.

Arifovic represented firms (suppliers) in two ways. In the single-population representation, each firm was represented as one of $N = 30$ strings in a single population. In the multiple population representation, each of the m firms is represented by a different population of 30 strings. In both cases, each string in a population represents a decision as to how much a firm might produce in the current period, $q_i(t) \in [0, \bar{q}]$, absent knowledge of the market price that will prevail. This decision was encoded as a string, of length 30, using a binary alphabet; initial 'bit' values were randomly determined. The fitness criterion used was the firm's current period profit; to evaluate fitness, strings had to be decoded to real quantities. In addition to using the standard genetic algorithm operations of selection, crossover and mutation on the binary strings, Arifovic adopted a fourth operator, which amounted to an augmented, elitist selection criterion which Arifovic called "election." Following crossover and mutation, which yields two new strings from two parent strings, the fitness of the new, offspring strings is evaluated and compared with the fitness of the parent strings; of this group of four strings, the two strings with the highest fitness values are allowed to enter the next generation of candidate solutions. This election operator simply allows the genetic algorithm to converge, asymptotically to a solution; without it, mutations would lead to persistent heterogeneity in the string population in the neighborhood of a solution. In the case of the single population representation, Arifovic reported the

[19]Similarly, Axelrod (1987) sought to determine whether the human-submitted 'tit-for-tat' strategy that won his (1984) prisoner's dilemma tournament would emerge in a simulation exercise that used a genetic algorithm to evolve strategies (it did).

[20]Hommes (1994) studies the more general case where demand is linear and supply is nonlinear. He provides conditions under which adaptive learning dynamics converge to limit cycles or chaos in the unstable case. Sonnemans et al. (2004) provide experimental evidence in support of Hommes' predictions.

average value of q_it in the population of 30 strings; in the case of the multiple population simulation, Arifovic imagined that each firm randomly chose one of its strings to determine its quantity decision in each period; she then reported the average of these m quantity decisions. In certain simulations, the model parameters were chosen to be the same as in one of Wellford's treatments, including number of periods, 30, and the number of firms, $m = 5$.

Figure 9 shows results for the unstable parameter case; the left panel shows the average quantity produced (with a 1-standard deviation band) for the human subject experiments and the right panel shows the same for a simulation of the multiple-population version of the genetic algorithm over the same number of periods. [Figure 9 here]. Both the human subjects and the genetic algorithm converges to a neighborhood of the equilibrium quantity of 14 though convergence takes longer and is more volatile in this 'unstable case' than in the stable case (not shown). However, the average quantity in the GA simulation appears to get very close to the equilibrium prediction beginning after period 10 while the same cannot be said of the experimental data. However, consistent with the experimental evidence, Arifovic is able to reject the null of difference between the volatility of prices in the stable and unstable cases using the simulation data. These findings provide some support for the reasonableness of genetic algorithms as models of adaptive processes.

Several papers explore GA learning in general equilibrium, overlapping generation models of money, and compare the results with experimental findings. Arifovic (1996) studies exchange rate volatility in a two-country, two-currency, two-period overlapping generations model due to Kareken and Wallace (1981). Details of this model are discussed in LeBaron's (2005) chapter. Arifovic's main conclusion is that, counter to the theoretical prediction derived under the rational expectations assumption, under genetic algorithm learning, the exchange rate displays persistent volatility, which is due to the persistence of mutation and the election operator.

By contrast, Arifovic (1995) shows that in a single country model, an equilibrium with valued fiat currency and low inflation is asymptotically stable under GA learning with persistent mutation and the election operator in place. The selection by the GA of the stationary, low inflation equilibrium, rather than another high inflation, stationary equilibrium is consistent with the laboratory findings of Marimon and Sunder (1993). Other, homogeneous an non-evolutionary learning algorithms, such as recursive least squares learning, fail to converge to the same low inflation equilibrium (see, e.g., Marcet and Sargent (1989)).

In Arifovic's work, the strings of the GA encode decisions that agents make, e.g. how much to consume in the first period. The GA then works to find the optimal decision, given feasibility and budget constraints. In Marimon and Sunder's (1993, 1994) overlapping generation experiments, subjects were not asked to make consumption/savings decisions as pilot studies suggested that subjects had a difficult time solving that kind of intertemporal optimization problem. Instead, Marimon and Sunder asked subjects to provide forecasts of the price level they expected would prevail in the next period. Given a subject's forecast, the computer program solved that subject's optimal consumption/savings allocation and determined market clearing prices. Bullard and Duffy (1999) adopted this same learning-how-to-forecast design in a GA-learning simulation of the environment studied by Arifovic (1995). They imagine that agents have some belief about how prices in period $t + 1$ will be related to prices in period t, and the strings of the GA encode this belief. Given the price forecast, the program optimally determines each agent's consumption/savings decision, along with market clearing prices. Bullard and Duffy (1999) show that this learning-how-to-forecast implementation of GA learning results in findings that are consistent with the experimental evidence of Marimon and Sunder (1994) and also with Arifovic (1995)'s learning-how-to-optimize implementation of GA learning.

Several papers use GAs to understand findings from auction experiments. A difficulty with auctions is that participants frequently fail to win an item or agree to a transaction, so that the fitness of strategies may need to be assessed over a longer period of time than is typical in other applications of

GAs.

Andreoni and Miller (1995) use genetic algorithms as a way of studying how close populations of adaptive agents might come to learn equilibrium bid functions in a variety of auction formats: first and second price affiliated-values auctions, first and second price private-values auctions, and common value auctions. The design of their simulation experiments is aligned with that of laboratory experiments with paid human subjects in several dimensions, e.g., the number of bidders in a group and the information available to these bidders. However, their 20 simulation runs of 1000 generations per auction format is more difficult to compare with the 20-30 auctions that human subjects participate in the typical experiment. In Andreoni and Miller's implementation, the genetic algorithm is employed to search over two parameters of a general linear bidding function of the form

$$b(x_i) = \beta_{i1} x_i + \beta_{i2} \epsilon,$$

where x_i is agent (string) i's valuation and ϵ is some distribution parameter that varies according to the knowledge that agents are assumed to have, e.g., whether valuations are private-independent, private-affiliated or common. This functional form nests (to an approximation) all the equilibrium bid functions that are predicted to obtain in the various auction formats. The binary strings of the GA encode the two parameters, β_1 and β_2. For the standard GA implementation, Andreoni and Miller report that the GA simulations come closest to learning the equilibrium bid functions in the affiliated private value, first or second price auction formats and have more difficulty achieving the equilibrium bid functions in the independent-private and common value formats. Consistent with evidence from human subject experiments, e.g. Cox et al. (1982), Kagel and Levin (1986), they find violations of revenue equivalence between first- and second- price auction formats, and they find that smaller groups of 4 rather than 8 bidders are less prone to the winner's curse in common value auctions.

Dawid (1999b) examines genetic algorithm learning in a sealed bid, double auction market. The N buyers' each have some value, v, from consuming a unit of the single good while the N sellers' have some cost, c, of producing a unit of the good, and $1 > v > c > 0$. The strings of the GA encode the buyers' bids and the sellers' asks. In each period, buyer and sellers are randomly paired. If a buyer's bid, p_b, exceeds a sellers' ask, p_a, a transaction occurs at price $p = (p_a + p_b)/2$; otherwise no transaction occurs. Profits are determined in the usual way, $v - p$ for buyers and $p - c$ for sellers, and the fitness of buyer/seller rules and application of genetic operators is assessed every m periods. Dawid shows analytically that the only locally stable equilibria under GA dynamics are those where all buyers (sellers) submit the same bid (ask) in the interval $[v, c]$. In 50 simulation runs where $v = 1$ and $c = 0$, he reports that the most common outcome is a single price equilibrium in a small neighborhood of .5. Interestingly, this finding is quite similar to that observed in an experiment conducted by Valley et al. (2002), where values of v and c are drawn randomly from $[0, 1]$ and after learning these values, pairs of players were allowed to communicate with one another prior to submitting bids/asks. The most common outcome, in cases where gains from trade are possible ($v > c$), was for both buyer and seller to name the same price. While this experimental finding may not be so surprising, the fact that the GA simulation delivers this same finding, without any explicit communication between populations of buyers and sellers, is quite interesting.

Finally, there are several papers comparing GA simulations with experimental findings in labor markets. Pingle and Tesfatsion (2001) examine the impact of varying levels of non-employment benefits on worker–employer matches and on-the-job cooperation using data from both human subject experiments and computational experiments that make use of genetic algorithms. The environment studied is a repeated two-stage game where in the first stage workers decide whether (and to which employer) to work or remain unemployed while employers decide whether to accept these offers or keep a position vacant. At the end of this first stage, unemployed workers and employers with vacancies receive a fixed non-employment benefit while matched workers and employers proceed to the second stage, which involves play of a prisoner's dilemma game, with strategies labeled 'shirk' and 'don't shirk.' The single treatment variable was the size of the non-employment benefit. The human subject experiments revealed that increases in the non-employment benefit both decreased the frequency with which relationships formed, and the frequency of mutual cooperation between worker-employer pairs, though this effect was not monotonic. Further, long-term relationships between the same worker and employer were rare. The computational labor market had four times as many workers and employers as the human subject experiment and was simulated for a much longer period of time: 1,000 generations. Each generation consisted up successive trade cycles followed by an evolutionary step that updated strategies; the genetic algorithm operates in the latter stage. A trade cycle consisted of both a matching process, which utilizes a reinforcement learning algorithm to determine the expected utility of potential partners, followed by a work-site interaction among matched players. The work-site interaction was governed by a finite state automaton, and the genetic algorithm was used to search for potentially better work-site rules in the evolution step. Among the findings from simulations of this model are that, consistent with the human subject experiments, the frequency of employment relationships decreases with increases in the non-employment benefit. On the other hand, by contrast with the human subject findings, in the computational experiment, nearly all employers and workers end up in long-term fixed relationships, and either mutual cooperation or mutual defection becomes the norm, depending on initial conditions. The authors suggest that these differences may be owing to differences in the design of the two experiments, in particular the different number of employers and workers in the computational versus the human subject experiments appears to have played an important role in the outcomes, though the different time-frames of analysis may also be a contributing factor.

Ünver (2001a) and Haruvy et al. (2002) use genetic algorithms to model the two-sided, worker-firm matching process in markets for medical intern and federal law clerks and compare these results with human subject experiments. These entry-level labor markets as well as others, have been susceptible to a phenomenon known as unraveling, in which the date at which firms and workers agree to contracts becomes increasingly earlier in time relative to the actual start-date of employment leading to possible inefficiencies in matches due to unavailability of relevant information. Some markets have sought to address this problem by having centralized clearinghouses that match workers with firms. Ünver studies three centralized matching mechanisms used in British medical-intern markets. Of these three, two are still in use, though only one of these two is stable in the Gale-Shapley (1962) sense. Ünver uses GA to encode and model the evolution of worker-firm strategies under these three mechanisms. Among other findings, he shows that the theoretically unstable, "linear programming" matching protocol may not be susceptible to unraveling under the GA adaptation, which is consistent with the continued use of this mechanism in the field. He is able to corroborate other findings of two-sided matching experiments conducted by Kagel and Roth (2000) and Ünver (2001b) that explore the unraveling in the British medical intern markets.

Haruvy et al. (2002) conduct a parallel experiment with human subjects and with artificial agents modeled using a genetic algorithm with the aim of studying two-sided matching in the market for federal law clerks. Applicants initially decide whether to submit applications to judges of varying qualities, and judges may in turn accept offers. The grades of applicants, affecting the payoff from a match, are only fully revealed later, during a centralized matching process. Matches not made by the end of the first two periods (years) are, in certain treatments, subject to a centralized match in period 3 using a stable matching protocol. In the 'idealized-centralized' treatment, applicants are not required to submit offers prior to the centralized match in order to participate in it, while in the coerced-centralized treatment they are required to submit offers prior to the match. In both cases, offers accepted prior to the centralized match date are binding, consistent with practice in this market, though in the idealized treatment, binding offers can be avoided by waiting for the centralized match. In the human subject experiments, the authors report that many more subjects in the role of applicants and judges wait for the centralized match under the 'idealized-centralized' treatment than do so under the coerced-centralized treatment, and given the additional information that can be obtained by waiting, welfare is higher in the former treatment than in the latter. In genetic algorithm simulations, where the strategies of applicants and judges co-evolve, a similar finding obtains. Haruvy et al. are careful to compare their findings for human subject experiments over the same time-scale used in the genetic algorithm simulations. They then carry out the genetic algorithm simulation exercise much further in time, and find that this difference becomes even more pronounced over time. This seems a reasonable merger of the two technologies they use to understand these matching markets. As they observe p. 3, "the computations will give us some assurance that our experimental results are not artifacts of slow learning in the laboratory, while experiments will assure us that the behavior produced by the genetic algorithms is in fact similar to human behavior."

The findings from all of these studies provide some support for the reasonableness of genetic algorithms as models of adaptive learning by populations of heterogenous agents. Genetic algorithms appear best suited for large, complex search spaces where it is more efficient to sample from the set of possible actions/strategies than to enumerate all possibilities and consider their relative fitness at every decision step. At the same time, most of the studies treat the genetic algorithm as a kind of black box generator of new-and-improved decisions or strategies, without much regard to the interpretation of genetic operators, or how they compare with actual human decision-making processes. Toward this goal, it would be of interest to consider the marginal contribution of each of the genetic operators in explaining data from human subjects, an exercise akin to adding additional structure to ZI-algorithms or moving from reinforcement to hypothetical reinforcement (belief) learning models.

4.3 Comparisons Between Genetic Algorithm and Reinforcement Learning

Two papers have compared the performance of genetic algorithm learning and reinforcement learning in terms of explaining data from human subject experiments. Haruvy and Ünver (2003) study matching behavior in procurement-type markets where the matching decision is consequential to both the seller and the buyer. They are interested in the question of whether buyers and sellers achieve a stable outcome, á la Gale and Shapley (1962) and if so, whether the stable matching is optimal for the party who initiates a proposed match (buyers or sellers). As the strategy space in the repeated game they consider is highly complex, and there are multiple stable outcomes, deductive reasoning is not very useful and so they turn to inductive reasoning processes, in particular, reinforcement learning and genetic algorithm learning, to predict what will happen in the experiments they conduct with human subjects. Both the reinforcement and genetic algorithm learning simulations predict that in seller– (buyer–) proposing markets, sellers (buyers) are most likely to achieve the seller– (buyer–) optimal stable outcome, and this prediction is consistent with the experimental findings. Aside from the observation that the two learning models yield the same prediction however, Haruvy and Ünver do not go into a deeper comparison of the performance of the two learning models.

By contrast, Arifovic and Ledyard (2004) look for a clear winner between reinforcement and genetic algorithm learning in the context of a repeated public good game that makes use of a Groves-Ledyard allocation mechanism. As the authors point out, this environment differs from those typically studied by learning researchers in that the strategy space is continuous. They compare the predictions of an "individual evolutionary learning" model (a GA–without–crossover for each individual's strategies) with Roth-Erev–style reinforcement learning and Camerer and Ho's (1999) hybrid reinforcement-belief learning algorithm in terms of the fit of simulations of these models to the experimental data. To facilitate a comparison, some discretization of the action space is necessary. They report that for two different ways of discretizing the strategy space, reinforcement learning fares substantially worse than the other two learning approaches in that it takes much longer to converge to the Nash equilibrium than does the human subjects. However, the version of reinforcement learning they use is not as general as Roth and Erev allow. For instance, there is no forgetting factor nor is there any spillover in the probability choice updating to nearby strategies. Given the large strategy space considered, it is not so surprising that the genetic algorithm appears to perform best for the reasons noted above. However, before concluding in favor of one approach over others, it would be useful to compare the predictions of evolutionary and reinforcement–type learning models on a broad range of games including those with both continuous and discrete strategy sets.

4.4 Classifier Systems

Classifier systems, first proposed by Holland (1986), are inductive, rule-based learning systems that combine reinforcement-type learning over a set of simple logical rules called classifiers, with occasional use of a genetic algorithm search for new classifiers. As with genetic algorithms, there are many variants, but a typical classifier system consists of four parts: 1) a set of if–then decision rules or classifiers, 2) an accounting system for assessing the strength of classifiers and for apportioning credit, 3) an auction system for determining which classifiers are invoked and 4) a genetic algorithm for the introduction of new classifiers. Classifier systems are perhaps best viewed as models of individual learning, akin to expert systems, while genetic algorithms, as typically modeled are often interpreted as models of population or social learning. As Vriend (2000) points out, simulations with classifier systems used to model social learning (mimicry) at the population level can yield outcomes that differ substantially from simulations with classifier systems used to model learning at the level of individual agents, especially in environments where strategic considerations come into play.[21]

The first use of a classifier system (or a genetic algorithm) in an economic application was due to Marimon, McGrattan and Sargent (1990), who used a classifier system to model behavior in Kiyotaki and Wright's (1989) model of money as a medium of exchange. That model has equal numbers of three types of agents who produce either good 1, 2 or 3, but who desire to consume another good, e.g. type 1 produces good 2, type 2 produces good 3, and type 3 produces good 1. Each agent may store a single unit of a good at a time, and the goods have different storage costs, with good 1 being the least costly to store and good 3 being the most costly to store. Agents receive utility from consumption of the good they desire in an amount that exceeds the highest storage cost. In each period, agents are randomly paired and decide whether to engage in trade with their match. Trades must be mutually agreed upon by both parties, in which case inventories of the two goods are swapped; otherwise, inventories of goods do not change. Agents earn utility only when they trade for the good they desire; in that case they immediately produce a new unit of their production good. In every period they incur storage costs base on the type of good they hold in inventory. The optimal trading strategy for a type 2 or 3 player is a fundamental, cost-reducing pure strategy in which they agree (refuse) to trade the good they hold in storage for less (more) costly–to–store goods in route to getting the good they desire to consume. On the other hand, depending on parameter values, type 1 players may find it optimal to adopt the fundamental strategy, or a speculative strategy in which they trade their production good 2 for the more costly to store good 3 with the rational expectation that speculating in the more costly to store good 3 will reduce the time it takes to acquire the good they desire, good 1.

In Marimon et al.'s implementation, there are two classifier systems for every agent, a set of trade and consumption classifiers represented by strings. The trade classifier takes as input the good an agent has in storage and the good that his match has in storage, and provides, as output, a decision

[21]For a further discussion of this issue see, e.g., Riechmann (2002) and Arifovic and Maschek (2004).

(or message) of whether to trade or not. The consumption classifier takes as input the good a player has in storage and provides as output, a decision (message) of whether or not to consume that good. Each classifier has a strength or fitness measure associated with it. In each period, the collection of classifiers that satisfy the current state for an agent, consisting of the good the agent holds in storage and the good in storage of the matched player, bid a fraction of their current strengths in an auction that determines which classifier the agent adopts; the highest bidding classifier of each type is chosen, its bid is deducted from its strength and its decision is implemented. The bid of the winning exchange classifier in the current period is paid to (added to the strength of) the previous period's winning consumption classifier, which determined the current good the agent holds in storage, while the bid of the winning consumption classifier is paid to current period winning exchange classifier, which determine the good the agent holds in storage. This payment system is what Holland termed a 'bucket brigade' wherein classifiers that are not necessarily active in the current period, but which were critical for activating classifiers that were active can still earn some share of credit and see their strengths improve. The current winning consumption classifier earns the 'external' payoff associated with its decision, which depends on whether the good in storage is the desired good or not. Finally a genetic algorithm is called on, with some decreasing frequency, to generate new classifiers, with the population of parent strings being selected from the population of classifiers according to relative strengths. The set of strings resulting from the genetic operators are assigned the strengths of the parent strings.

In simulations of this system, Marimon et al. report many interesting findings, but the main finding is that speculative trading strategies (e.g. by type 1 players) are not observed in environments where, in equilibrium, they would comprise a unique best response. Marimon et al. comment on this finding by observing that the behavior of the artificial agents, modeled using classifier systems, can be very myopic in the beginning, while it may take time for some optimal strategies, such as speculation, to achieve strengths that will sustain these strategies. They conclude that "the present algorithm seems defective in that it has too little experimentation to support the speculative equilibrium even in the long simulations we have run." [22]

Inspired by Marimon et al.'s simulation findings, Duffy and Ochs (1999, 2002) sought to test the Kiyotaki-Wright model in a laboratory experiment. They made an effort to provide subjects with all the information relevant in to making optimal decisions in the theoretical environment. Duffy and Ochs sought to induce a stationary infinite horizon, as the theory presumes, by having an indefinite end to a sequence of pairwise trading rounds. Such concerns with implementation of infinite horizons do not typically concern agent-based modelers, as the artificial agents in their models are not typically forward-looking, alleviating concerns about backward induction due to end-game effects. Finally, among the parameterizations they chose was one that was also used by Marimon et al. (1990). Though Duffy and Ochs had only 8 or 10 agents of each of the three types, while Marimon et al. had 50, the findings from the human subject experiments were quite similar to those obtained in the artificial agent simulations using classifier systems. In particular, Duffy and Ochs also find that subjects failed to adopt speculative trading strategies in environments where such strategies comprise an equilibrium best response.[23]

Duffy (2001) considers two alterations of the Kiyotaki-Wright model that might serve to promote

[22]Subsequent applications of classifier systems in economic applications, include Başçi (1999), Beltrametti, et al. (1997) and Vriend (2000). LeBaron's (2005) chapter discusses the Santa Fe artificial stock market (Arthur et al. (1997)) which makes use of a classifier system to model traders' decisions. See Lettau and Uhlig (1999) for a comparison between classifier/rule learning and dynamic programming.

[23]Brown (1996) conducted an experimental test of the Kiyotaki-Wright that was more narrowly focused on the speculative equilibrium prediction and came to the same conclusion: most subjects failed to adopt the speculative trading strategy.

the adoption of speculative strategies. In one version, agents whose optimal equilibrium strategy calls for speculation are given more encounters with situations where playing the speculative strategy results in higher expected utility. In the other, two of the three agent types are constrained to playing the strategies that are optimal for them in equilibrium. Duffy adopts a reinforcement learning model which is similar to the exchange classifier of Marimon et al. (1990), automates the consumption classifier and gets rid of the genetic algorithm. A similar model was found to provide a good fit to the experimental data of Duffy and Ochs (1999). Duffy uses this reinforcement model to simulate what will happen in the two alternative environments, and reports that both alternatives speed up the learning of speculative strategies. However, the adoption of speculative strategies is greater in the second alternative, where two thirds of the agent types are constrained to playing optimal strategies. He then conducts an experiment with human subjects designed to test these same alternatives. In the human subject experiment, the model parameters, the number of agents, and other features of the environment are kept as similar as possible to that of the simulated environments to facilitate comparisons. The human subject findings are largely consistent with the artificial agent findings. Duffy stresses that agent-based modeling exercises of this type can be a useful tool for experimental design, and at the same time, the results of human subject experiments might be useful in thinking about how to model the decisions of artificial agents.

4.5 Genetic Programming

Another variant of genetic algorithm learning is known as genetic programming developed by Koza (1992). In genetic programming, the same genetic operators of the GA are used to search over a population self-executing computer programs represented as decision trees (variable–length strings) in an effort to obtain an optimal functional relationship or program. This type of genetic search is well-suited to finding functional solutions to problems that do not readily yield closed-form solutions. Genetic programming has been mainly used by economists to study financial market phenomena, e.g. to uncover technical trading rules or to discover pricing formulas for financial derivatives. Chen (2002) provides a good survey.

However, the external validity of genetic programming has been assessed through a few comparisons with the results of human subject experiments. Perhaps the best known work is that of Chen and Yeh (1996), who revisit the unstable cobweb model studied by Arifovic (1994) and examined experimentally by Wellford (1989). Chen and Yeh note that it is more general to view agents as learning a *functional relationship* for prices, e.g. $E_{t-1}p_t = f(p_{t-1}, p_{t-2}...)$ than for them to be learning about what quantity to produce as in Arifovic's (1994) implementation, as the former approach allows for the possibility that the equilibrium is not a fixed point, e.g. it could be a limit cycle. Chen and Yeh apply a genetic programming algorithm to search over a class of price forecast functions. Essentially the algorithm allows for a wide range of linear and nonlinear functions mapping from observations on as many as 10 past prices to deliver a forecast for period t. These forecast functions determine quantities which subsequently determine actual market prices via the equilibrium market clearing condition. Fitness of individual forecast functions is then assessed, and genetic operations are applied to advance the search for better price forecast functions in a manner analogous to the genetic algorithm search. Chen and Yeh report that for the same unstable parameterization of the model considered by Arifovic and Wellford, (as well as for some even more egregious cases) their genetic programming algorithm has no difficulty yielding price predictions that were very close to the equilibrium price level without the need for an election operator to contain the effects of the mutation operator. The price forecasting functions are initially quite complex and difficult to interpret. However, as convergence to the equilibrium obtains, the price forecasting functions become quite simple, as prices cease to vary so much.[24]

[24] Chen et al. (2002) use a genetic programming algorithm to reach a similar conclusion in a median effort coordination

In a quite different application, Duffy and Engle-Warnick (2002) use genetic programming to infer the strategies that human subjects play in a simple bargaining game, given only the actions and histories of the players. This approach, which Koza (1992) termed "symbolic regression" involves evaluation of a population of computer programs in terms of their relative success in mapping from inputs, e.g. players' histories, to output, e.g. player's action choices. An advantage of this approach is that the user does not have to specify the functional form of the strategy model in advance, aside from specifying a set of model primitives; both the form and the coefficients of the computer programs are estimated simultaneously. Using this algorithm, Duffy and Engle-Warnick report that simple threshold strategies characterize the behavior of most of the human subject participants.

4.6 Summary

Evolutionary algorithms, by contrast with ZI and individual learning algorithms, are derived from principles of population biology. While the principle of survival and propagation based on relative fitness is similar to reinforcement learning, fitness assessments in evolutionary algorithms are not made on the basis of an individual agent or strategy's own history, but instead are based on population-wide measures. The biological models from which evolutionary algorithms derive lead to some difficulties of interpretation for social scientists. While some efforts have been made to interpret the operators of evolutionary algorithms, the more common approach has been to treat these algorithms as a kind of black box model of social learning, and focus on the similarity between aggregate outcomes in simulations and in human subject experiments.

Two main approaches in evolutionary models have been identified. With the replicator dynamic, the set of strategies or actions must be fully specified at the outset. Such an approach is reasonable in environments where the set of actions or strategies is small. In environments where the search space is larger, a genetic algorithm approach may be preferred. GAs are effective, population-based search algorithms that optimize on the tradeoff between finding new strategies, and exploiting strategies that have worked well in the past.

Comparisons between simulations using evolutionary algorithms and human subject experiments suggest that there is some support for the use of evolutionary algorithms as models of population learning. However, the time-frame and the number of agents used in simulation of evolution algorithms is often quite different from that adopted in human subject experiments.

5 Conclusions and Directions for the Future

Two parallel computer-based technologies, the experimental and the computational laboratory, have begun to have a major impact on economic research. While top-down, deductive theorizing with fully rational agents remains the standard in economics, the findings of experimentalists and ACE researchers using bottom–up, boundedly rational, inductive models of behavior are attracting increasing attention in the profession, as these models often provide a better fit to experimental (as well as to field) data, and operate without the centralized coordinating devices found in standard theory.

There are difficulties with the external validity of both approaches. Agent-based models have many degrees of freedom, while experimental methods are unable to perfectly induce or control subject behavior, etc. Still, the fact that findings from agent-based models and human subject experiments are often in agreement helps to allay concerns with either approach individually. Can an argument

game studied experimentally by Van Huyck et al. (1994). Chen et al. show that a steady state effort level that is theoretically unstable under a myopic, homogeneous best–response learning dynamic turns out to be stable under the genetic–programming–based learning system in accordance with Van Huyck et al.'s (1994) finding from human subject experiments.

be made for one approach over the other? Analogous to Judd's (1997) answer to the question of whether computational economics and economic theory are substitutes or complements, we have seen that agent-based models and humans subject experiments are sometimes nearly perfect substitutes (e.g., zero intelligent agents in certain versions of the double auction market) but are more often complements (e.g., the degree of sophistication in individual learning models can be calibrated based on experimental data).

There are several directions for future research. First, further comparisons of different agent-based models using a variety of experimental data sets are needed. "Horse-races" such as those between reinforcement learning and belief-learning and between belief-learning and replicator dynamics are important for choosing among agent-based modeling approaches. Second, further parallel experiments with human and artificial agents situated in the same environment are needed to better understand the external validity of agent-based models as well as to appropriately calibrate those models. These parallel experiments will necessarily involve more constraints on agent-based modeling exercises than on human subject designs owing to the stricter time and budget constraints of laboratory research. However, if agent-based models can accurately track the behavior of human subjects over the short-time frame of a human subject experiment, that finding would give the ACE researcher some license to carry out simulations of the model over a much longer time-frame, as might be necessary to achieve convergence to an equilibrium. Third, new agent-based models might be developed based on laboratory evidence.

There are at least two possibilities for attacking the latter goal. First, researchers could seek to determine how players go about analyzing the experimental environments in which they are placed. For example, the kind of information subjects consider, their cognitive skills and other characteristics that Costa-Gomes et al. (2001) have termed the players' *strategic sophistication*. Costa-Gomes et al.'s use of the Mouselab software which enables the researcher to capture and study the information that players consider in playing normal form games, as well as Camerer et al.'s (1993) use of the Mouselab software to study behavior in extensive form games, is very useful in identifying heterogeneity of player types, and testing cognitive concepts such as backward induction.

A second possibility for designing agent-based models more fully grounded in laboratory evidence is to make greater use of an experimental design known as the strategy method, first proposed by Selten (1967). The strategy method requires subjects to simultaneously specify, prior to the start of a game, the strategies they will play in that game, i.e. their action choice at every information set. Subjects' choices are then made for them based on the strategies they submit.[25] Unlike observing how players make decisions as a game unfolds in real-time and attempting to infer subjects' strategies from their action choices, the strategy method provides researchers with all the information necessary to program artificial agent strategies.[26] In more complex environments, it may be necessary to give subjects experience with the game prior to having them submit strategies. For instance, Selten et al.

[25] The counterpart of the strategy method in the agent-based literature is to hold a tournament á la Axelrod (1984), in which researchers submit computer code (strategies) characterizing the behavior of their gladiatorial-agent models. The tournament organizers then use some matching protocol or test suite of problems/data to determine a winning strategy/program. (See, e.g. the Trading Agent Competition, http://www.sics.se/tac/ or The Turing Tournament, http://turing.ssel.caltech.edu/). However, tournaments, especially the winner-take-all variety, may alter incentives so that the strategies/programs submitted do not reflect decisions agents would make in non-tournament environments, e.g. in random-pairwise interactions. For instance, winner-take-all tournaments might give rise to a higher variance in payoffs than simple random pair-wise interactions. If there is free entry/exit into a tournament (as is typically the case), then one might expect tournament participants to have a higher tolerance for risk than would agents interacting in random pair-wise encounters, so that tournament findings could be misleading for agent-based modelers. Further, tournaments are expensive to run, and are infrequently conducted more than once. By contrast, eliciting strategies from human subjects in non-tournament environments is relatively cheap and can be done repeatedly.

[26] Some researchers believe that the strategy method changes the way players play a game. The experimental evidence on this question is mixed. See, e.g. Brandts and Charness (2000) and Brosig et al. (2003).

(1997) have subjects play a Cournot duopoly game repeatedly and then ask them to program their strategies. The programmed strategies were then played against one another and the programmers were allowed to alter their strategies based on their performance. The adoption of such an approach might well lead to the development of new adaptive models with a greater claim to the term 'agent-based.'

References

Andreoni, J. and J.H. Miller (1995), "Auctions with artificial adaptive agents," Games and Economic Behavior 58:211–221.

Arifovic, J. (1994), "Genetic algorithm learning and the cobweb model," Journal of Economic Dynamics and Control 18:3-28.

Arifovic, J. (1995), "Genetic algorithms and inflationary economies," Journal of Monetary Economics 36:219-43.

Arifovic, J. (1996), "The behavior of the exchange rate in the genetic algorithm and experimental economies," Journal of Political Economy 104:510-41.

Arifovic, J. and J. Ledyard (2004), "Scaling up learning models in public good games," Journal of Public Economic Theory 6:203–238.

Arifovic, J. and M.K. Maschek (2004), "Social vs. individual learning – What makes a difference?" working paper, Simon Fraser University.

Arthur, W.B. (1991), "Designing economic agents that act like human agents: A behavioral approach to bounded rationality," American Economic Review Papers and Proceedings 81:353-59.

Arthur, W.B. (1993), "On designing economic agents that behave like human agents," Journal of Evolutionary Economics 3:1-22.

Arthur, W.B. (1994), "Inductive reasoning and bounded rationality" American Economic Review Papers and Proceedings 84:406–411.

Arthur, W.B., J.H. Holland, B. LeBaron, R. Palmer and P. Taylor (1997), "Asset pricing under endogenous expectations in an artificial stock market," in: W.B. Arthur, S.N. Durlauf and D.A. Lane, eds., The Economy as an Evolving Complex System II, Proceedings Volume XXVII, Santa Fe Institute Studies in the Science of Complexity, (Addison-Wesley, Reading, MA), 15–44.

Axelrod, R. (1984), The Evolution of Cooperation, (Basic Books, New York).

Axelrod, R. (1987), "The evolution of strategies in the iterated prisoner's dilemma," in L. Davis, ed., Genetic Algorithms and Simulated Annealing, (Morgan Kaufmann, Los Alamos, CA), 32-41.

Axelrod, R. (1997), The Complexity of Cooperation: Agent–Based Models of Competition and Collaboration, (Princeton University Press, Princeton, N.J.).

Axelrod, R. and L. Tesfatsion (2005), "A guide for newcomers to agent-based modeling in the social sciences," this handbook.

Başçi, E. (1999), "Learning by imitation," Journal of Economic Dynamics and Control, 23:1569-85.

Battalio, R., L. Samuelson and J. Van Huyck (2001), "Optimization incentives and coordination failure in laboratory stag hunt games," Econometrica 69:749–764.

Batten, D.F. (2000), Discovering Artificial Economics: How Agents Learn and Economies Evolve, (Westview Press, Boulder, CO).

Beltrametti, L. et al. (1997), “A learning-to-forecast experiment on the foreign exchange market with a classifier system,” Journal of Economic Dynamics and Control, 21:1543-75.

Bendor, J., D. Mookherjee and D. Ray (2001), “Aspiration-based reinforcement learning in repeated interaction games: An overview,” International Game Theory Review 3:159–174.

Bendor, J., D. Diermeir, and M. Ting (2003), “A behavioral model of turnout,” American Political Science Review 97:261–280.

Bereby–Meyer, Y. and I. Erev (1998), “On learning to become a successful loser: A comparison of alternative abstractions of learning in the loss domain,” Journal of Mathematical Psychology 42:266-86.

Blume, A., D.V. DeJong, G.R. Neumann, and N.E. Savin (2002), “Learning and communication in sender–receiver games: An econometric investigation,” Journal of Applied Econometrics 17:225–247.

Börgers and Sarin (2000), “Naive reinforcement learning with endogenous aspirations,” International Economic Review 41:921-50.

Bosch-Doménech, A. and S. Sunder (2001), “Tracking the invisible hand: Convergence of double auctions to competitive equilibrium,” Computational Economics 16:257-84.

Boylan, R.T. and M.A. El-Gamal (1993), “Fictitious play: A statistical study of multiple economic experiments,” Games and Economic Behavior 5:205–222.

Bray, M.M. and N.E. Savin (1986), “Rational expectations equilibria, learning, and model specification,” Econometrica 54:1129-60.

Brandts, J. and G. Charness (2000), “Hot vs. cold: Sequential responses and preference stability in experimental games,” Experimental Economics 2:227–238.

Brenner, T. (2005), “Agent Learning Representation: Advice on Modelling Economic Learning,” this handbook.

Brewer, P.J., M. Huang, B. Nelson, and C.R. Plott (2002), “On the behavioral foundations of the law of supply and demand: Human convergence and robot randomness,” Experimental Economics 5:179-208.

Brosig, J., J. Weimann and C–L. Yang (2003), “The hot versus cold effect in a simple bargaining experiment,” Experimental Economics 6:75–90.

Brown, G.W. (1951), “Iterative solution of games by fictitious play,” in T. Koopmans, ed., Activity Analysis of Production and Allocation, (Wiley, New York), 374-376.

Brown, P., (1996), “Experimental evidence on money as a medium of exchange,” Journal of Economic Dynamics and Control 20:583-600.

Brown-Kruse, J.L. (1991), “Contestability in the presence of an alternative market: An experimental examination,” Rand Journal of Economics 22:136–147.

Bullard, J. and J. Duffy (1998), “A model of learning and emulation with artificial adaptive agents,” Journal of Economic Dynamics and Control 22:179-207.

Bullard, J. and J. Duffy (1999), "Using genetic algorithms to model the evolution of heterogeneous beliefs," Computational Economics, 13:41-60.

Bush, R.R. and F. Mosteller (1955), Stochastic Models for Learning (John Wiley & Sons, New York).

Camerer, C. (2003), Behavioral Game Theory: Experiments in Strategic Interaction, (Princeton University Press, Princeton).

Camerer, C. and T-H. Ho (1999), "Experience-Weighted Attraction Learning in Normal Form Games," Econometrica 67:827-74.

Camerer, C. E. Johnson, T. Rymon and S. Sen (1993), "Cognition and framing in sequential bargaining for gains and losses," in K. Binmore, et al., eds., Frontiers of Game Theory, (MIT Press, Cambridge, MA), 27–47.

Carlson, J. (1968), "An invariably stable cobweb model," Review of Economic Studies 35:360-362.

Chamberlin, E.H. (1948), "An experimental imperfect market," Journal of Political Economy 56:95-108.

Chan, N.T., B. LeBaron, A.W. Lo and T. Poggio (1999), "Agent-based models of financial markets: A comparison with experimental markets," working paper.

Chen, S.H., J. Duffy and C.H. Yeh (2002), "Equilibrium selection via adaptation: Using genetic programming to model learning in a coordination game," The Electronic Journal of Evolutionary Modelling and Economic Dynamics, http://e-jemed.org article 1002.

Chen, S.H. and C.H. Yeh (1996), "Genetic programming learning and the cobweb model," in P.J. Angeline and K.E. Kinnear Jr., eds., Advances in Genetic Programming 2, (MIT Press, Cambridge, MA), 443–466.

Chen, S.H. (2002), Genetic Algorithms and Genetic Programming in Computational Finance, (Kluwer, Dordrecht).

Cheung, Y-W. and D. Friedman (1997), "Learning in evolutionary games: some laboratory results," Games and Economic Behavior 19, 46–76.

Cheung, Y-W. and D. Friedman (1998), "A comparison of learning and replicator dynamics using experimental data," Journal of Economic Behavior and Organization 35, 263–280.

Cliff, D. and J. Bruten (1997a), "Minimal–intelligence agents for bargaining behaviors in market-based environments," Technical report HP-97-91, Hewlett–Packard Research Labs, Bristol, England.

Cliff, D. and J. Bruten (1997b), "More than zero intelligence needed for continuous double-auction trading," Hewlett Packard Laboratories Paper HPL-97-157, Bristol, England.

Coursey, D., R.M. Issac, M. Luke and V.L. Smith (1984), "Market contestability in the presence of sunk (entry) costs," Rand Journal of Economics 15:69-84.

Costa-Gomes, M., V. Crawford and B. Broseta (2001), "Cognition and behavior in normal-form games: An experimental study," Econometrica 69, 1193–1235.

Crockett, S. (2004), “Learning competitive equilibrium in experimental exchange economies,” working paper, Carnegie–Mellon University.

Crockett, S., S. Spear, and S. Sunder (2004), “A simple decentralized institution for learning competitive equilibrium,” working paper, Carnegie–Mellon and Yale Universities.

Cross, J.G., (1983), A theory of adaptive economic behavior, (Cambridge University Press, Cambridge, UK).

Cyert, R.M. and J.G. March (1956), “Organizational factors in the theory of oligopoly,” Quarterly Journal of Economics 70, 44-64.

Davis, D.D. and C.A. Holt (1993), Experimental Economics, (Princeton University Press, Princeton).

Dawid, Herbert (1999a), Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models, second revised and enlarged edition, (Springer-Verlag, Berlin).

Dawid, H. (1999b), “On the convergence of genetic learning in a double auction market,” Journal of Economic Dynamics and Control 23:1545–1569.

Duffy, J. (2001), “Learning to speculate: Experiments with artificial and real agents,” Journal of Economic Dynamics and Control 25:295-319.

Duffy, J. and J. Engle-Warnick (2002), “Using symbolic regression to infer strategies from experimental data,” in S.H. Chen, ed., Evolutionary Computation in Economics and Finance, (Physica-Verlag, New York), 61–82.

Duffy, J. and N. Feltovich (1999), “Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study,” International Journal of Game Theory 28:131–152.

Duffy, J. and E. Hopkins (2005), “Learning, information and sorting in market entry games: Theory and evidence,” Games and Economic Behavior 51:31-62.

Duffy, J. and J. Ochs (1999), “Emergence of money as a medium of exchange: An experimental study,” American Economic Review 89:847-77.

Duffy, J. and J. Ochs (2002), “Intrinsically worthless objects as media of exchange: experimental evidence,” International Economic Review 43:637-673

Duffy, J. and M.U. Ünver (2005), “Asset price bubbles and crashes with near zero–intelligence traders,” forthcoming in Economic Theory.

Ellison, G. (1993), “Learning, local interaction and coordination,” Econometrica 61, 1047-1071.

Easley, D. and J.O. Ledyard (1993), “Theories of price formation and exchange in double oral auctions,” in D. Friedman and J. Rust, eds., The Double Auction Market: Institutions Theories, and Evidence, Santa Fe Institute Studies in the Sciences of Complexity, Vol. 14, (Addison–Wesley, Reading, MA), 63–97.

Epstein, J.M. and R. Axtell (1996), Growing Artificial Societies: Social Science from the Bottom Up, (MIT Press, Cambridge, MA).

Erev, I. and G. Barron (2003), “On adaptation, maximization and reinforcement learning among cognitive strategies,” working paper, Technion and Harvard Universities.

Erev, I., Y. Bereby–Meyer and A.E. Roth (1999), “The effect of adding a constant to all payoffs: experimental investigation and implications for reinforcement learning models,” Journal of Economic Behavior and Organization 39:111-128.

Erev, I. and A.E. Roth (1998), “Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria, American Economic Review 88:848–881.

Ezekiel, M. (1938), “The cobweb theorem,” Quarterly Journal of Economics 52:255-280.

Feltovich, N. (2000), “Reinforcement–based vs. belief–based learning models in experimental asymmetric information games,” Econometrica 68:605–641.

Franke, R. (2003), “Reinforcement learning in the El Farol model,” Journal of Economic Behavior and Organization 51:367-388.

Friedman (1991), “A simple testable model of double auction markets,” Journal of Economic Behavior and Organization 15:47-70.

Friedman (1996), “Equilibrium in evolutionary games: some experimental results,” Economic Journal 106:1–25.

Foster, D. and H.P. Young (1990), “Stochastic evolutionary game dynamics,” Theoretical Population Biology 38:219–232.

Fudenberg, D. and D.K. Levine (1998), The Theory of Learning in Games. (MIT Press, Cambridge, MA).

Gjerstad, S. and J. Dickhaut (1998), Price formation in double auctions, Games and Economic Behavior 22:1–29.

Gode, D.K. and S. Sunder (1993), “Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality,” Journal of Political Economy 101:119–137.

Gode, D.K. and S. Sunder (1997a), “What makes markets allocationally efficient?” Quarterly Journal of Economics 112:603–630.

Gode, D.K. and S. Sunder (1997b), “Lower bounds for efficiency of surplus extraction in double auctions,” in D. Friedman and J. Rust, eds., The Double Auction Market: Institutions Theories, and Evidence, Santa Fe Institute Studies in the Sciences of Complexity, Vol. 14, (Addison–Wesley, Reading, MA), 199–220.

Gode, D.K. and S. Sunder (2004), “Double auction dynamics: structural effects of non-binding price controls,” Journal of Economic Dynamics and Control 28:1707–1731.

Gode, D.K., S.E. Spear and S. Sunder (2000), “Convergence of double auctions to competitive equilibrium in an Edgeworth box,” working paper.

Goldberg, D.E. (1989), Genetic Algorithms in Search, Optimization, and Machine Learning. (Addison–Wesley, Reading, MA).

Grossklags, J. and C. Schmidt (2004), “Artificial software agents on thin double auction markets – A human trader experiment,” working paper, University of California, Berkeley.

Haruvy, E. and M.U. Ünver (2003), "Equilibrium selection in repeated B2B matching markets," working paper, University of Texas, Dallas and Koç University.

Haruvy, E., A.E. Roth and M.U. Ünver (2002), "The dynamics of law clerk matching: An experimental and computational investigation of proposals for reform of the market," working paper.

Haruvy, E. and D.O. Stahl (2004), "Deductive versus inductive equilibrium selection: Experimental results," Journal of Economic Behavior and Organization 53:319-331.

Ho, T-H., C. Camerer, J-K. Chong (2002), "Economic value of EWA lite: A functional theory of learning in games," Working paper.

Hofbauer, J. and K. Sigmund (1988), The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection, (Cambridge University Press, Cambridge, UK).

Hofbauer, J. and K. Sigmund (1998), Evolutionary Games and Population Dynamics, (Cambridge University Press. Cambridge, UK).

Holland, J.H. (1975), Adaptation in Natural and Artificial Systems, (University of Michigan Press, Ann Arbor, MI).

Holland, J.H. (1986), "Escaping brittleness: The possibilities of general purpose learning algorithms applied to parallel rule-based systems," in R.S. Michalski. J.G. Carbonell and T.M. Mitchell, eds., Machine Learning: An Artificial Intelligence Approach, (Morgan Kaufmann, Los Altos, CA).

Hommes, C.H. (1994), "Dynamics of the cobweb model with adaptive expectations and nonlinear supply and demand," Journal of Economic Behavior and Organization 24:315-335.

Hopkins, E. (2002), "Two competing models of how agents learn in games," Econometrica 70:2141–2166.

Huck, S., K.A. Konrad, W. Müller, and H-T. Normann (2002), "Mergers and the perception of market power," working paper, University College London.

Issac, R.M. and C.R. Plott (1981), "Price controls and the behavior of auction markets: An experimental evaluation," American Economic Review 71:448–459.

Judd, K.L. (1997), "Computational economics and economic theory: Substitutes or complements?" Journal of Economic Dynamics and Control 21:907-42.

Judd, K.L. (1998), Numerical Methods in Economics, (MIT Press, Cambridge, MA).

Kaelbling, L.P., M.L. Littman, and A.W. Moore (1996), "Reinforcement learning: A survey," Journal of Artificial Intelligence Research 4: 237–285.

Kagel, J.H., R.M. Harstad and D. Levin (1987), "Information impact and allocation rules in auctions with affiliated private values: A laboratory study," Econometrica 55:1275-1304.

Kagel, J.H. and A.E. Roth (1995), The Handbook of Experimental Economics, (Princeton University Press, Princeton, NJ).

Kagel, J.H. and A.E. Roth (2000), "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment," Quarterly Journal of Economics 115:201-35

Kandori, M., G.J. Mailath and R. Rob (1993), "Learning, mutation, and long run equilibria in games," Econometrica 61:29-56.

Karandikar, R., D. Mookherjee, D. Ray and F. Vega-Redondo (1998), "Evolving aspirations and cooperation," Journal of Economic Theory 80:292-331.

Kareken, J.H. and N. Wallace (1981), "On the indeterminacy of equilibrium exchange rates," Quarterly Journal of Economics 96:207-22.

Kiyotaki, N. and Wright, R. (1989), "On money as a medium of exchange," Journal of Political Economy 97:927–954.

Koza, J.R. (1992), Genetic Programming, (MIT Press, Cambridge, MA).

Kutschinski, E., T. Uthmann, and D. Polani (2003), "Learning competitive pricing strategies by multi-agent reinforcement learning," Journal of Economic Dynamics and Control 27:2207-2218.

LeBaron, B. (2005), "Agent-based computational finance," this Handbook.

Lettau, M. and H. Uhlig, (1999), "Rules of thumb versus dynamic programming," American Economic Review, 89:148-74.

Lucas, R.E. (1986), "Adaptive behavior and economic theory," Journal of Business 59:S401–S426.

Mackie-Mason, J. and M. Wellman (2005), "Automated markets and trading agents," this Handbook.

Marimon, R., E. McGrattan, and T.J. Sargent (1990), "Money as a medium of exchange in an economy with artificially intelligent agents," Journal of Economic Dynamics and Control 14:329–373.

Marimon, R. and S. Sunder (1994), "Expectations and learning under alternative monetary regimes: An experimental approach," Economic Theory 4:131-62.

Marimon, R. and S. Sunder (1993), "Indeterminacy of equilibria in a hyperinflationary world: Experimental evidence," Econometrica 61:1073-1107.

Michalewicz, Z. (1996), Genetic Algorithms + Data Structures = Evolution Programs, Third Edition, (Springer-Verlag, Berlin).

Miller, R.M. (2002), Paving Wall Street: Experimental Economics and the Quest for the Perfect Market, (John Wiley, New York).

Mirowski, P. (2002), Machine Dreams: Economics Becomes a Cyborg Science, (Cambridge University Press, Cambridge, UK).

Mitchell, T.M. (1997), Machine Learning, (McGraw Hill, New York).

Morris, S. (2000), "Contagion," Review of Economic Studies 67: 57–78.

Mookherjee, D. and B. Sopher (1997), "Learning and decision costs in experimental constant sum games," Games and Economic Behavior 19:97-132.

Mookherjee, D. and B. Sopher (1994), "Learning behavior in an experimental matching pennies game," Games and Economic Behavior 7:62-91.

Nicolaisen, J., V. Petrov and L. Tesfatsion (2001), “Market power and efficiency in a computational electricity market with discriminatory double auction pricing,” IEEE Transactions on Evolutionary Computation 5:504-523.

Nyarko, Y. and A. Schotter (2002), “An experimental study of belief learning using elicited beliefs,” Econometrica 70:971–1005.

Ochs, J. (1995), “Games with unique, mixed strategy equilibria: An experimental study,” Games and Economic Behavior 10:202-17.

Pemantle, R. and B. Skyrms (2003), “Network formation by reinforcement learning: The long and medium run,” Working Paper, UC Irivne.

Pingle, M. and L. Tesfatsion (2001), “Non-employment payoffs and the evolution of worker-employer cooperation: Experiments with real and computational agents,” Economic Report 55, Iowa State University.

Riechmann, T. (1999), “Learning and behavioral stability - An economic interpretation of genetic algorithms,” Journal of Evolutionary Economics 9:225-242.

Riechmann, T. (2001a), Learning in economics: Analysis and application of genetic algorithms, (Berlin, Springer-Verlag).

Riechmann, T. (2001b), “Genetic algorithm learning and evolutionary games,” Journal of Economic Dynamics and Control 25:1019-37.

Riechmann, T. (2002), “Cournot or Walras? Agent based learning, Rationality and long run results in oligopoly games,” Discussion Paper, University of Hannover.

Robinson, J. (1951), “An iterative method of solving a game,” The Annals of Mathematics 54:296-301

Roth (1995), “Introduction to experimental economics” in J. Kagel and A.E. Roth, eds., Handbook of Experimental Economics, (Princeton University Press, Princeton, NJ).

Roth, A.E. and J.K. Murnighan (1978), “Equilibrium behavior and repeated play of the prisoner’s dilemma,” Journal of Mathematical Psychology 17:189-198.

Roth, A.E. and I. Erev (1995), “Learning in extensive–form games: Experimental data and simple dynamic models in the intermediate term,” Games and Economic Behavior 8:164–212.

Sadrieh, A. (1998), The Alternating Double Auction Market: A Game Theoretic and Experimental Investigation, Lecture Notes in Economics and Mathematical Systems, Vol. 466, (Springer, Berlin).

Salmon, T.C. (2001), “An evaluation of econometric models of adaptive learning,” Econometrica 69:1597-1628.

Sargent, T.J. (1993), Bounded Rationality in Macroeconomics, (Oxford University Press, Oxford).

Sarin, R. and F. Vahid (2001), “Predicting how people play games: A simple dynamic model of choice,” Games and Economic Behavior 34:104-22.

Sarin, R. and F. Vahid (1999), “Payoff assessments without probabilities: A simple dynamic model of choice,” Games and Economic Behavior 28:294-309.

Schelling, T.C. (1978), Micromotives and Macrobehavior, (W.W. Norton, New York).

Selten, R. (1967), "Die strategiemethdoe zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments," in H. Sauermann, ed., Beiträge zur Experimentellen Wirtschaftsforschung, (J.C.B. Mohr, Tübingen), 136–168.

Selten, R. (1991), "Evolution, learning and economic behavior," Games and Economic Behavior 3:3-24.

Selten, R. M. Mitzkewitz, and G.R. Uhlich (1997), "Duopoly strategies programmed by experienced players," Econometrica 65:517–555.

Simon, H.A. (1955), "A behavioral model of rational choice," Quarterly Journal of Economics 69:99–118.

Simon, H.A. (1982), Models of Bounded Rationality, (MIT Press, Cambridge, MA).

Smith, V.L. (1962), "An experimental study of competitive market behavior," Journal of Political Economy 70:111-137.

Smith, V.L. (1982), "Microeconomic systems as an experimental science," American Economic Review 72:923-955.

Smith, V.L., G.L. Suchanek, and A.W. Williams (1988), "Bubbles, crashes, and endogenous expectations in experimental spot asset markets," Econometrica 56:1119-51.

Sonnemans, J., C. Hommes, J. Tuinstra and H. v.d. Velden (2004), "The instability of a heterogeneous cobweb economy: A strategy experiment on expectation formation," Journal of Economic Behavior and Organization 54:453-481.

Stahl D., (1999), "A horse race among action-reinforcement learning models," working paper, University of Texas–Austin.

Suppes, P. and R. Atkinson (1960), Markov Learning Models for Multiperson Interactions, (Stanford University Press, Stanford).

Sutton, R.S. and A.G. Barto (1998), Reinforcement Learning: An Introduction, (MIT Press, Cambridge, MA).

Thorndike, E.L. (1911), Animal Intelligence, (Hafner Publishing, New York).

Ünver, M.U. (2001a), "Backward unraveling over time: The evolution of strategic behavior in the entry level British medical labor markets," Journal of Economic Dynamics and Control 25:1039-80.

Ünver, M.U. (2001b), "On the survival of some unstable two–sided matching mechanisms: An experimental and computational investigation of the stability hypothesis," working paper, Koç University.

Valley, K., L. Thompson, R. Gibbons, M.H. Bazerman (2002), "How communication improves efficiency in bargaining games," Games and Economic Behavior 38:127–155.

Van Boening, M.V. and Wilcox, N.T. (1996), "Avoidable cost: Ride a double auction roller coaster," American Economic Review 86:461-77.

Van Huyck, J.B., J.P. Cook, and R.C. Battalio (1994), "Selection dynamics, asymptotic stability, and adaptive behavior," Journal of Political Economy 102:975–1005.

Van Huyck, J.B., J.P. Cook, and R.C. Battalio (1997), "Adaptive behavior and coordination failure," Journal of Economic Behavior and Organization 32:483-503.

Vriend, N.J. (2000), "An illustration of the essential difference between individual and social learning, and its consequences for computational analyses," Journal of Economic Dynamics and Control 24:1-19.

Watkins C. (1989), Learning From Delayed Rewards (PhD Dissertation) King's College Cambridge, England.

Wellford, C.P. (1989), A laboratory analysis of price dynamics and expectations in the cobweb model," Discussion Paper 89-15, Department of Economics, University of Arizona.

Young, P.H. (1993), The Evolution of Conventions, Econometrica 61:57-84.

Young, P.H. (1998), Individual Strategy and Social Structure: An Evolutionary Theory of Institutions, (Princeton University Press, Princeton).

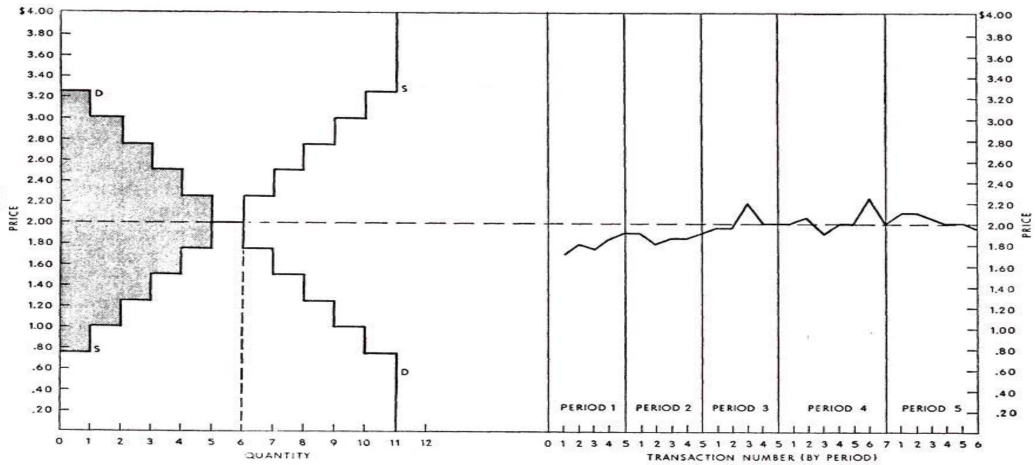

Figure 1: Values and costs induced in an experimental double auction design (left panel) and the path of prices achieved by human subjects (right panel) Source: Smith (1962, Chart 1).

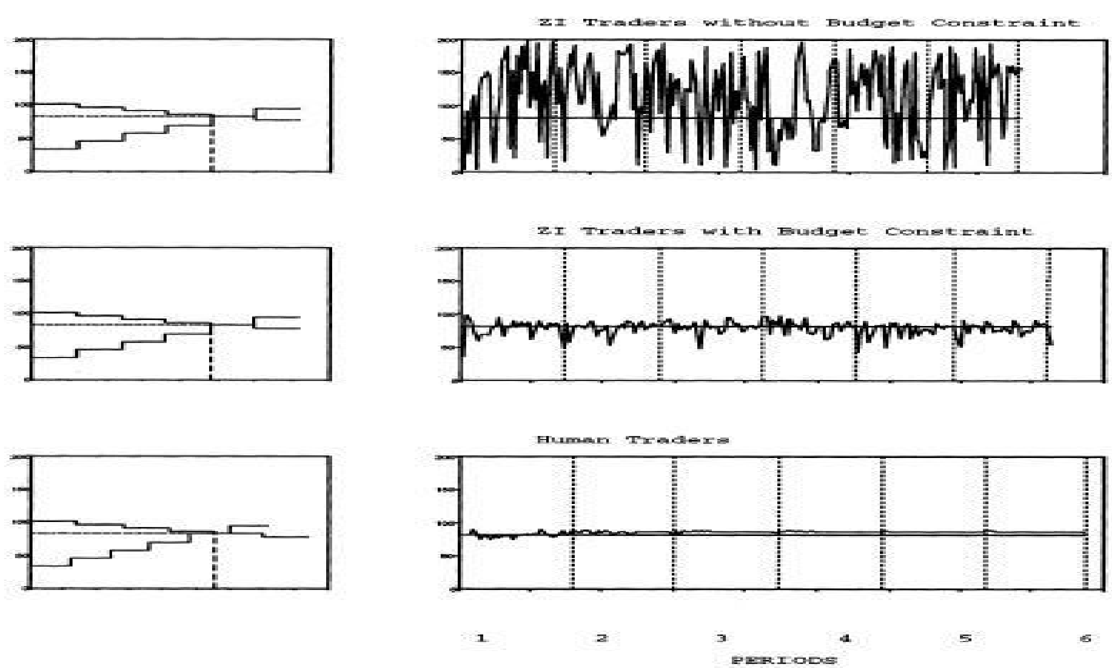

Figure 2: Competitive equilibrium prediction (left) and path of transaction prices (right) Source Gode and Sunder (1993, figure 1).

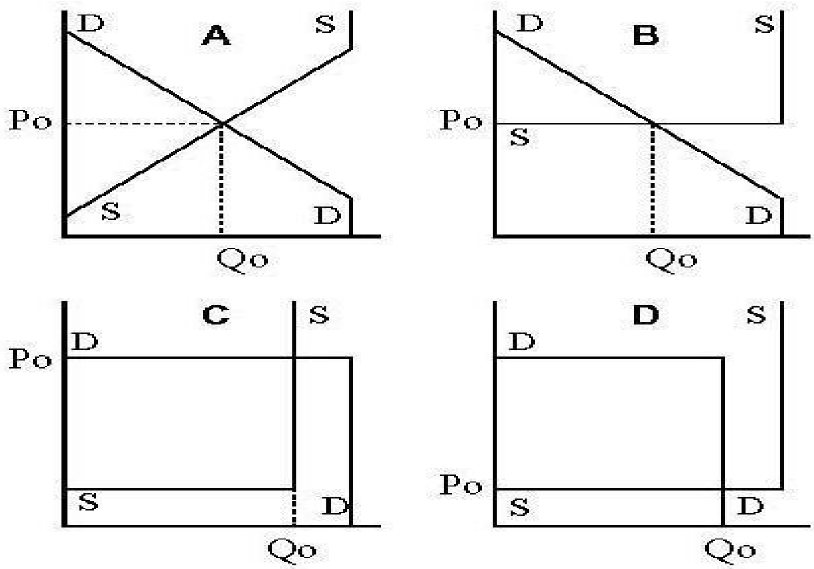

Figure 3: Demand (D) and Supply (S) Curves for Four Economies. Source: Cliff and Bruten (1997b).

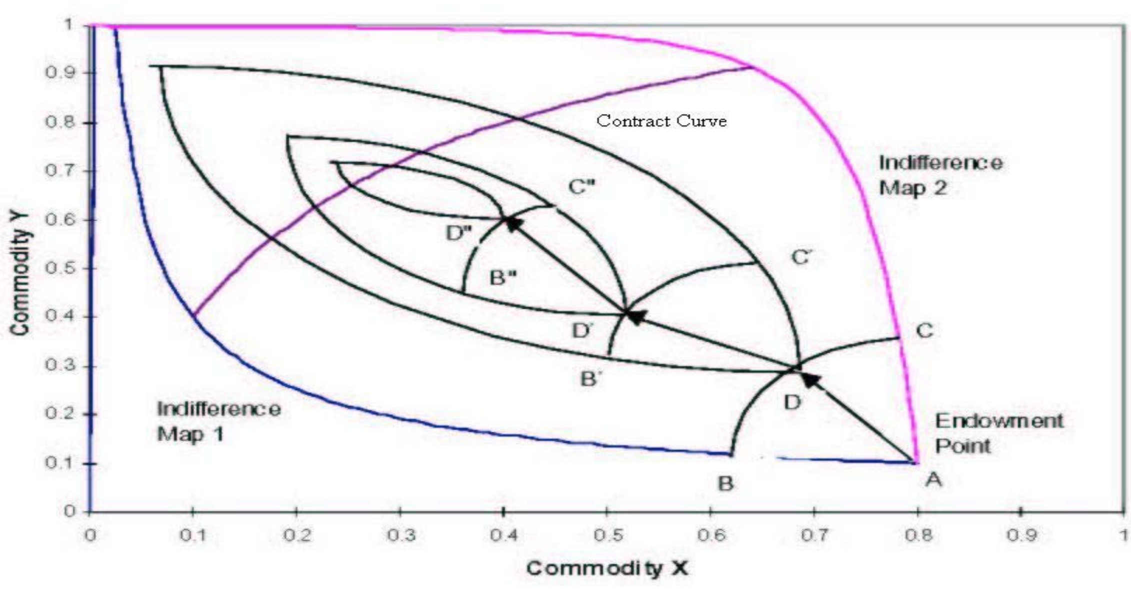

Figure 4: An illustration of the path of ZI transactions in an Edgeworth Box (Source: Gode, Spear and Sunder (2000)).

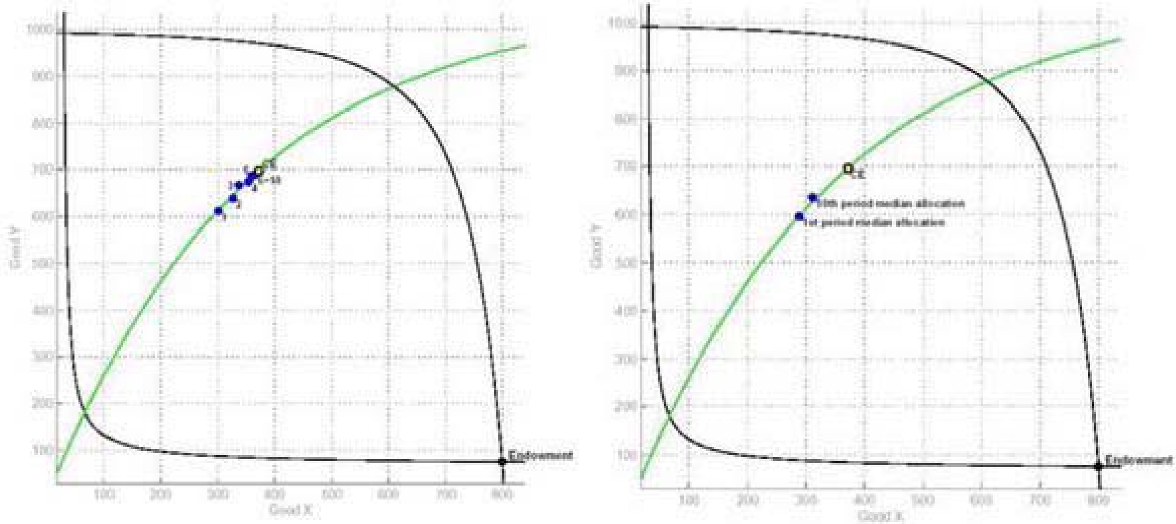

Figure 5: Median, end-of-period CSS-ZI allocations over periods 1-10 (left panel) versus median, end-of-period human subject allocations in periods 1 and 10 (right panel) in an Edgeworth box.

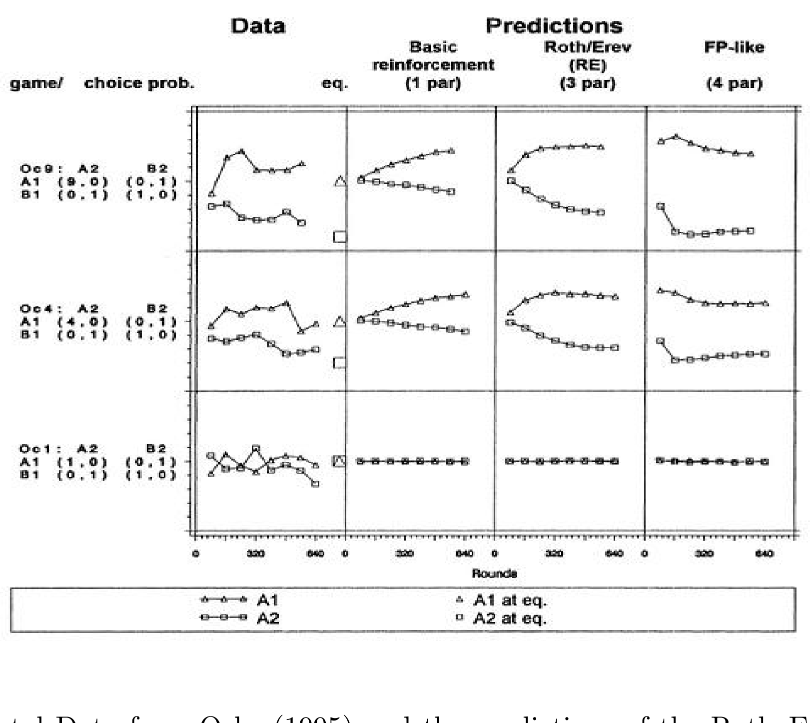

Figure 6: Experimental Data from Ochs (1995) and the predictions of the Roth–Erev and fictitious play learning models. (Source: Erev and Roth (1998).)

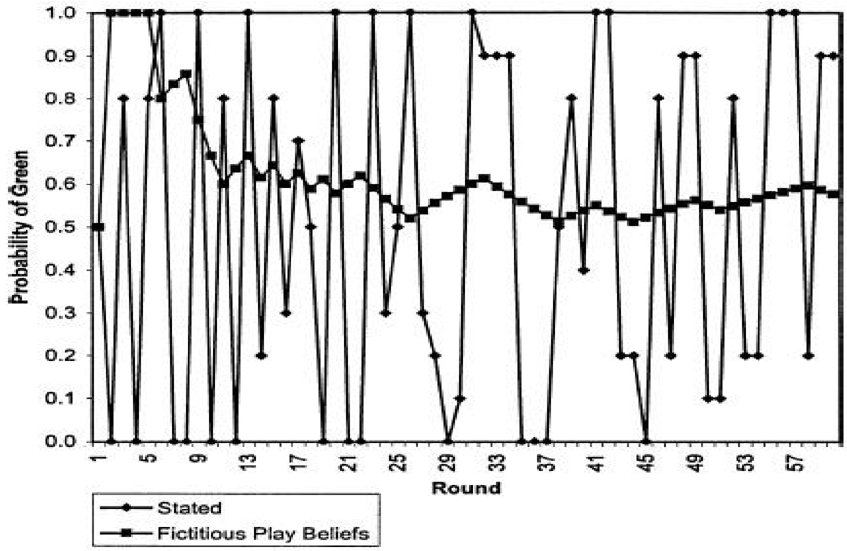

Figure 7: Stated versus Fictitious Play Beliefs of a Typical Subject in Nyarko and Schotter's Experiment (Source: Nyarko and Schotter (2002, fig. 2)).

$g = 0$
`initialize population of` N `strings, P(0)`
`while tolerance criterion remains unmet or` $g < G$
    `evaluate fitness of strings in` $P(g)$
    `select` N `strings for` $P(g+1)$ `based on relative fitness`
    `apply crossover to selected strings`
    `apply mutation to recombined strings`
    `evaluate tolerance criterion`
    $g = g + 1$
`end while`

Figure 8: Pseudo-code for a genetic algorithm

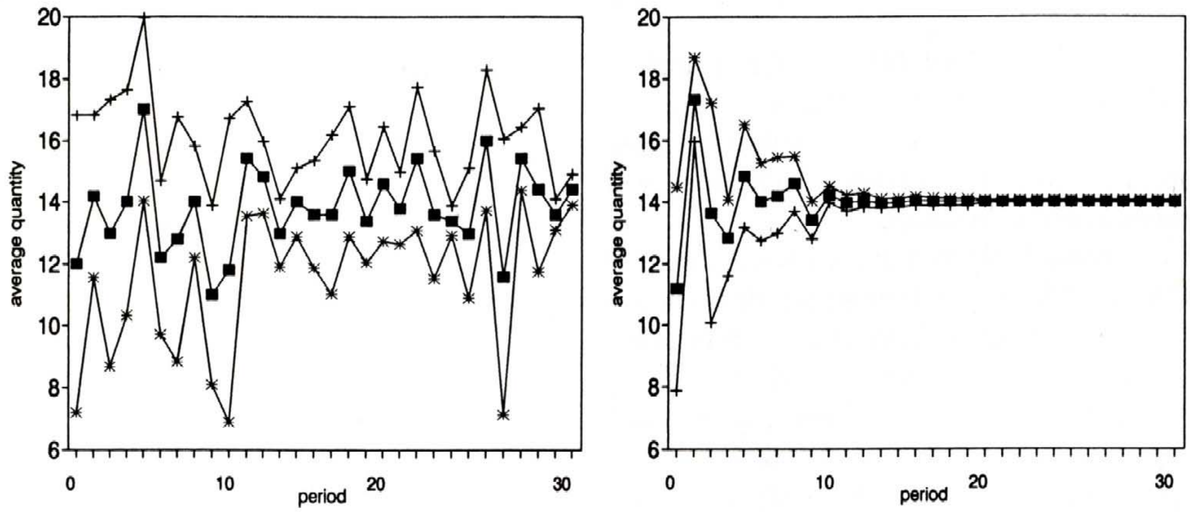

Figure 9: Average quantity in the Cobweb model, unstable case (plus/minus one st. dev.). Left panel: human subject data, right panel: multiple-population GA simulation (source: Arifovic 1994).