

Binmore, Ken; Samuelson, Larry

**Working Paper**

## Evolutionary drift and equilibrium selection

Reihe Ökonomie / Economics Series, No. 26

**Provided in Cooperation with:**

Institute for Advanced Studies (IHS), Vienna

*Suggested Citation:* Binmore, Ken; Samuelson, Larry (1996) : Evolutionary drift and equilibrium selection, Reihe Ökonomie / Economics Series, No. 26, Institute for Advanced Studies (IHS), Vienna

This Version is available at:

<https://hdl.handle.net/10419/68638>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**Institut für Höhere Studien (IHS), Wien  
Institute for Advanced Studies, Vienna**

**Reihe Ökonomie / Economics Series**

**No. 26**

## **Evolutionary Drift and Equilibrium Selection**

**Ken Binmore, Larry Samuelson**



# **Evolutionary Drift and Equilibrium Selection**

**Ken Binmore, Larry Samuelson**

Reihe Ökonomie / Economics Series No. 26

**February 1996**

Ken Binmore  
Department of Economics  
University College London  
Gower Street  
London WC1E 6BT  
UNITED KINGDOM  
Phone: +44 171-387 7050  
e-mail: k.binmore@ucl.ac.uk

Larry Samuelson  
Department of Economics  
University of Wisconsin  
1180 Observatory Drive  
Madison, Wisconsin 53706  
U.S.A.  
Phone: +1-608-263-7791  
e-mail: larrysam@cournot.econ.wisc.edu

**Institut für Höhere Studien (IHS), Wien  
Institute for Advanced Studies, Vienna**

The Institute for Advanced Studies in Vienna is an independent center of postgraduate training and research in the social sciences. The **Economics Series** presents research done at the Economics Department of the Institute for Advanced Studies. Department members, guests, visitors, and other researchers are invited to contribute and to submit manuscripts to the editors. All papers are subjected to an internal refereeing process.

**Editorial**

*Main Editor:*

Robert M. Kunst (Econometrics)

*Associate Editors:*

Christian Helmenstein (Macroeconomics)

Arno Riedl (Microeconomics)

## **Abstract**

This paper develops an approach to equilibrium selection in game theory based on studying the equilibrating process through which equilibrium is achieved. The differential equations derived from models of interactive learning typically have stationary states that are not isolated. Instead, Nash equilibria that specify the same behavior on the equilibrium path, but different out-of-equilibrium behavior, appear in connected components of stationary states. The stability properties of these components often depend critically on the perturbations to which the system is subjected. We argue that it is then important to incorporate such *drift* into the model. A sufficient condition is provided for drift to create stationary states with strong stability properties near a component of equilibria. This result is used to derive comparative static predictions concerning common questions raised in the literature on refinements of Nash equilibrium.

## **Keywords**

Evolutionary games, cheap talk, stability, drift

## **JEL-Classifications**

C70, C72

**Comments**

Financial support from National Science Foundation grants SES-9122176 and SBR-9320678, the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, and the British Economic and Social Research 'Beliefs and Behaviour Programme' is gratefully acknowledged. We thank Drew Fudenberg, Tilman Börgers, Klaus Ritzberger, Karl Schlag and Jörgen Weibull for helpful discussions. We are grateful to the Department of Economics at the University of Bonn and the Institute for Advanced Studies at the Hebrew University of Jerusalem, where part of this work was done, for their hospitality.

# EVOLUTIONARY DRIFT AND EQUILIBRIUM SELECTION<sup>1</sup>

Ken Binmore  
Department of Economics  
University College London  
Gower Street  
London WC1E 6BT England

Larry Samuelson  
Department of Economics  
University of Wisconsin  
1180 Observatory Drive  
Madison, Wisconsin 53706 USA

December 22, 1995

<sup>1</sup>Financial support from National Science Foundation grants SES-9122176 and SBR-9320678, the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, and the British Economic and Social Research 'Beliefs and Behaviour Programme' is gratefully acknowledged. We thank Drew Fudenberg, Tilman Börgers, Klaus Ritzberger, Karl Schlag and Jörgen Weibull for helpful discussions. We are grateful to the Department of Economics at the University of Bonn and the Institute for Advanced Studies at the Hebrew University of Jerusalem, where part of this work was done, for their hospitality.



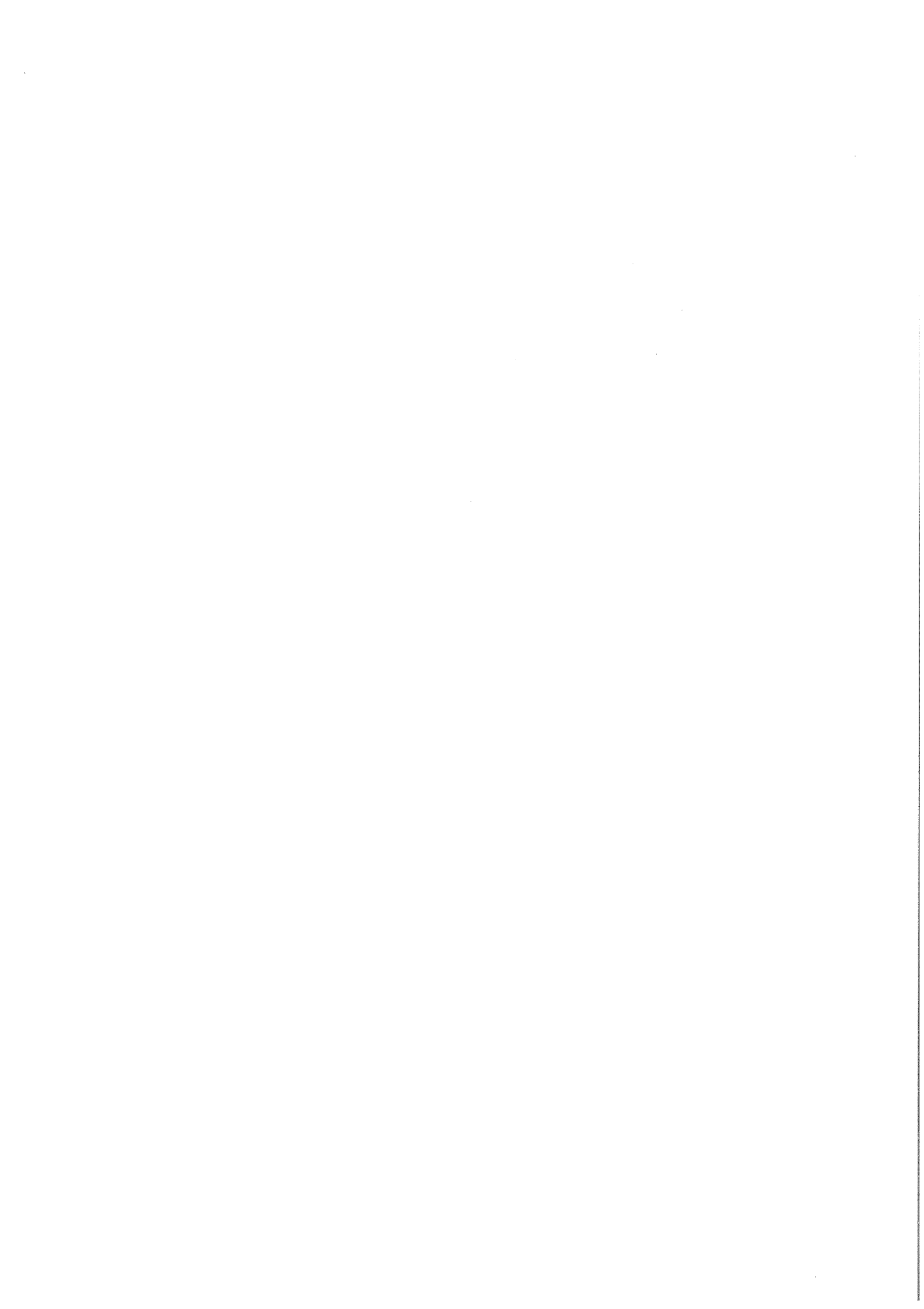


### Abstract

This paper develops an approach to equilibrium selection in game theory based on studying the equilibrating process through which equilibrium is achieved. The differential equations derived from models of interactive learning typically have stationary states that are not isolated. Instead, Nash equilibria that specify the same behavior on the equilibrium path, but different out-of-equilibrium behavior, appear in connected components of stationary states. The stability properties of these components often depend critically on the perturbations to which the system is subjected. We argue that it is then important to incorporate such *drift* into the model. A sufficient condition is provided for drift to create stationary states with strong stability properties near a component of equilibria. This result is used to derive comparative static predictions concerning common questions raised in the literature on refinements of Nash equilibrium

*Journal of Economic Literature* Classification Numbers C70, C72.

*Keywords:* Evolutionary Games, Cheap Talk, Stability, Drift.



# EVOLUTIONARY DRIFT AND EQUILIBRIUM SELECTION

by Ken Binmore and Larry Samuelson

## 1 Introduction

Game theory is a popular success as a theory, its methods having become a standard tool for economists. But its empirical success is more doubtful, with some of its simplest and clearest predictions having been consistently falsified in laboratory experiments.

One of the more striking empirical difficulties for game theory is posed by the problem of dominated strategies. It is generally held to be a fundamental principle of game theory that players should never use weakly dominated strategies.<sup>1</sup> Many canonical examples, such as the Ultimatum Game [21], the Dalek Game [29], and the Money-Burning Game [54], involve iterating the elimination of weakly dominated strategies to generate a unique prediction. But the experimental evidence is now strong that one cannot rely on predictions that depend on deleting weakly dominated strategies.<sup>2</sup>

Some critics take this experimental failure to imply that game theory is without predictive power. It is argued that people simply do not optimize, or that their utility functions incorporate exotic factors whose nature cannot easily be controlled in the laboratory (Bolton [11], Thaler [51], and Ochs and Roth [38]). We agree that players who face complicated games or who are poorly motivated are unlikely to optimize in the laboratory. Even in simple games played for significant stakes, it will take time for players to learn how to optimize. In the initial stages of the learning process, we must therefore expect their behavior to be driven by whatever social norm may have been triggered by the manner in which the game is framed. We also agree that players may be motivated by considerations other than the money payoffs offered to them by the experimenter.

---

<sup>1</sup>Game theory texts sometimes open by advocating this principle. The iterated elimination of weakly dominated strategies is identified by Kohlberg and Mertens [29] as a *sine qua non* for any satisfactory equilibrium concept. Dekel and Fudenberg [16, p.245] argue that the iterated elimination of weakly dominated strategies “clearly incorporates certain intuitive rationality postulates.” Nalebuff and Dixit [37, p.86] offer the avoidance of weakly dominated strategies as one of their four basic rules for playing games.

<sup>2</sup>See Güth, Schmittberger and Schwarze [21], Güth and Tietze [22], and Roth [41] for experimental studies of the Ultimatum Game. Balkenborg [1] studies the Dalek Game.

But none of these considerations justify the claims made by some critics that the optimization postulate of game theory is without value in the laboratory, or that the preferences of laboratory subjects have so little to do with monetary payoffs as to render the latter uninformative. Instead, we believe that laboratory behavior is often consistent with both the optimization paradigm and a naive interpretation of the players' preferences—provided that we recognize that *people must learn to optimize, and must do so in an imperfect world.*

In Binmore and Samuelson [6] and Binmore, Gale and Samuelson [4], for example, we showed that simple models of learning can direct players to Nash equilibria in the Ultimatum Game in which player *II* plays a weakly dominated strategy and receives a significant share of the surplus. But perhaps such outcomes are unstable? The world is rife with imperfections that we idealize away when formulating a game, but which must surely be present in the form of perturbations to any equilibrating process used to model how interactive learning leads subjects to an equilibrium of a game. Won't such perturbations inexorably eliminate weakly dominated strategies, and hence push the system to the subgame-perfect outcome of the Ultimatum Game, for the same reason that perfect equilibria emerge from Selten's [48] world of trembles? In [4] and [6] we show to the contrary that perturbations can play a key role in *stabilizing* Nash equilibria in weakly dominated strategies.

In this paper, we broaden the scope of our study of perturbed equilibrating processes. By analogy with the biologist Kimura [28], who stresses the importance of genetic drift between behaviors of equal fitness, we use the term *drift* to summarize all the imperfections that may perturb an equilibrating process. We then ask: when does drift matter? What can we say about equilibrium selection when drift does matter?

The key to our analysis lies in modeling the drift phenomena directly, rather than attempting to embody its results in the abstractly formulated equilibrium concepts of the refinements literature. As documented in Binmore and Samuelson [8], such a program for modelling drift has already been pursued implicitly by those who approach equilibrium selection by souping-up the definition of an evolutionarily stable strategy (ESS). However, we remain skeptical of the rewards to be gained from refining the ESS concept for much the same reasons that many game theorists have become skeptical of the literature on refining Nash equilibrium. As in the work of Kandori, Mailath and Rob [26] and Young [56], and in our previous work ([4, 7, 9]), we think it safer to operate with models in which the adjustment dynamics are specified explicitly. In this paper, we follow the lead of Nachbar [36],

Samuelson [43] and Samuelson and Zhang [44] in working with models based on deterministic differential equations.

We find that drift plays an important role whenever the problem of refining a Nash equilibrium is at issue. The alternative best replies that give rise to questions of equilibrium selection lead to components of stationary states of the adjustment dynamics. The learning dynamics typically lead toward some parts of such a component and away from other parts. As a result, the stability properties of the component depend crucially on the small shocks that cause the system to drift between equilibria within the component. If this drift pushes the system toward unstable equilibria, then the component as a whole will not be stable. If the drift pushes the system toward stable equilibria, then the component will have robust stability properties and deserves our attention as a long-run prediction of how the game will be played. Even arbitrarily small amounts of drift can be relevant in this context.

The finding that drift can be important might appear to be a death knell for empirical applications of game theory. How can we hope to make use of a theory whose implications depend upon the details of an arbitrarily small drift process? To answer this question, we investigate the implications of holding fixed the process by which players learn and the specification of drift, while manipulating the payoffs of the game. This leads to comparative-statics predictions concerning how the long-run outcome of play varies as the payoffs of the game vary. We view such predictions as a foundation for experimental work. Far from being the end of empirical applications, we think that an understanding of drift may provide the key to such work.

Section 2 introduces the model. Section 3 investigates conditions under which equilibrium selection results do not depend upon drift, which can therefore be ignored. Section 4 provides drift-driven, equilibrium-selection results for cases those cases in which outcomes depend upon drift. Section 5 applies these results to the Chain-Store Game, the Dalek Game, the Money-Burning Game, and a cheap talk game. Our aim in analyzing these examples is to demonstrate that Nash equilibria that are rejected by traditional equilibrium selection criteria may nevertheless be relevant to long-run prediction. Section 6 derives some comparative statics results to demonstrate that our conclusions are experimentally refutable in spite of their dependence on unobservably small levels of drift.

## 2 Drift

**The Model.** We consider an  $n$ -player game  $G$ , which we think of as being played by  $n$  populations of agents. We will speak of “players” when referring to the game  $G$ , and “agents” when referring to the members of the populations in the evolutionary model. However, we will find it convenient to use phrases like “player  $I$  plays  $B$  with probability  $\frac{3}{4}$ ” and “ $\frac{3}{4}$  of the agents in population  $I$  play  $B$ ” interchangeably.

A state  $z_\ell$  for population  $\ell$  is an  $|S_\ell| - 1$  dimensional vector of nonnegative numbers whose sum does not exceed one, where  $S_\ell$  is the strategy set of player  $\ell$ . We interpret such a vector as listing the fraction of agents in population  $\ell$  playing each of the first  $|S_\ell| - 1$  pure strategies in  $S_\ell$  (with the residual probability attached to the  $|S_\ell|$ th strategy). A state  $z$  of the entire system is a vector  $(z_1, z_2, \dots, z_n)$  identifying the state of each population.

**Dynamics.** Let  $z(t)$  be the population state at time  $t$ . The evolution of the state is described by a deterministic differential equation:

$$\frac{dz}{dt} = f(z) + \lambda g(z). \quad (1)$$

This differential equation is defined on the  $\prod_{\ell=1}^n (|S_\ell| - 1)$ -dimensional set given by the product of the  $n$  simplexes  $S_\ell$ . We typically denote this state space by  $Z$  (but occasionally also denote it by  $W$ ). We assume that  $f$  and  $g$  are continuously differentiable (and hence Lipschitz continuous) on an open set containing the state space  $Z$ . This ensures that there exists a unique, continuously differentiable solution  $z = z(t, z(0))$ , specifying the state at time  $t$  given initial condition  $z(0)$ , that satisfies (1) (Hale [23, chapter 1], Hirsch and Smale [24, chapter 9]). We assume that the state space  $Z$  is forward invariant under this solution, so that once the solution is an element of  $Z$ , it never leaves  $Z$ . Coupled with differentiability and the compactness of  $Z$ , this ensures that we encounter no boundary problems when working with the dynamic. Similarly, there exists a unique solution to the equation  $dz/dt = f(z)$ , which we again assume renders the state space  $Z$  forward invariant.<sup>3</sup>

To interpret (1), we think of  $f$  as capturing the important forces that govern agents’ strategy revisions. We refer to  $f$  as the “selection process.” In a biological context,  $f$  models a process of natural selection driven by

---

<sup>3</sup>Common examples such as the replicator dynamics satisfy these assumptions.

differences in fitness. In the models of Young [56] and Kandori, Mailath and Rob [26],  $f$  models a best-response learning process driven by differences in payoffs. Like any model, however, the selection process is an approximation, designed to capture the essential features of a problem while excluding a host of supposedly insignificant considerations. These latter forces are described by  $g$ . In a biological model,  $g$  models mutations, which are random alterations in the genetic structure of an individual. In the models of Young [56] and Kandori, Mailath and Rob [26],  $g$  models random alterations in agents' strategies. We refer to  $g$  as drift.

If our model is well constructed, meaning that the payoffs are a good representation of preferences and  $f$  captures the important forces behind strategy revisions, then we expect  $f$  to be closely linked to payoffs. The drift term  $g$ , however, may well have little or nothing to do with the payoffs of the game. Instead, considerations excluded from the model when specifying the game and its payoffs may play a major role in shaping  $g$ . In some cases, drift may be completely unrelated to payoffs. This is the case in many biological models of mutation as well as in the models of Young [56] and Kandori, Mailath and Rob [26].

The relative importance of drift is measured by  $\lambda$ , which we refer to as the "drift level." We will think of  $\lambda$  as being small, reflecting the belief that important considerations have been captured by the selection process  $f$ . Again, we follow in the footsteps of biological models and especially the models of Young and Kandori, Mailath and Rob.

**Selection.** How can we talk of mutations, or random alterations to strategies, while working with the deterministic model of (1)? We answer this question by briefly sketching the foundations of (1).

We begin with a stochastic model of the process by which agents choose strategies. Agents drawn from finite populations are repeatedly matched to play the game. In light of their experience, they adjust their strategies. These changes in strategy are governed by a Markov process. The state in which the system will next be found is a random variable that depends only upon the current state. Strategy adjustments are noisy, in the sense that knowing the state at time  $t$  allows us to identify only a probability measure describing the likely identity of the next state.

Let the *expected* state at time  $t+\tau$  given that  $z(t) = z$  be  $\mathcal{E}\{z(t+1)|z\} = F(z, \tau) + \lambda G(z, \tau)$ , where  $F$  represents the selection process and  $G$  represents mutations or other noise that may perturb the system. A Taylor expansion



of this expression gives:

$$\mathcal{E}\{z(t + \tau), z\} = z + \tau[f(z) + \lambda g(z)] + O(\tau^2) \quad (2)$$

or

$$\frac{\mathcal{E}\{z(t + \tau), z\} - z}{\tau} = f(z) + \lambda g(z) + O(\tau). \quad (3)$$

Strategy adjustment processes satisfying (2) are derived directly from explicit models of behavior in Binmore, Samuelson and Vaughan [9] in a biological context and by Binmore and Samuelson [7] in a learning context.<sup>4</sup>

The step from (3) to (1) now apparently involves nothing more than taking the limit as  $\tau \rightarrow 0$  and removing the expectation on the left side of (3) so that the result can be interpreted as  $dz/dt$ . Such a step is often justified informally, with a statement that the expectation can be removed because interest is directed to the case of a large population, so that a law of large numbers argument applies.<sup>5</sup>

When can the link between (3) and (1) be established formally? This depends on the span of time in which we are interested. Binmore and Samuelson [6] introduce a distinction between the medium run, the long run and the ultralong run. The medium run is a period of time long enough for selection to occur but too short for this selection to yield convergence to an equilibrium. The long run is a period of time long enough for selection to lead the system to the vicinity of an equilibrium of the game  $G$ .<sup>6</sup> However, this equilibrium need not be the final resting point of the system, as the noise in the model may occasionally produce a sufficiently large shock to bounce the system away from one equilibrium and into the basin of attraction of another. The ultralong run is a period of time long enough that sufficiently

---

<sup>4</sup>Notice that we write the original stochastic process as  $F + \lambda G$  and hence we do not interpret the drift term  $g$  as the remainder from a Taylor expansion of the underlying selection process  $F$ . In particular, we do not think that  $F$  is an exact model of the underlying stochastic selection process, with errors arising only from our use of local approximations. Instead, we view the underlying model selection model  $F$  itself as an approximation.

<sup>5</sup>See, for example, van Damme [52, ch. 9.4] and Hofbauer and Sigmund [25, ch. 16.1]. An alternative approach, based on stochastic differential equations, is pursued by Cabrales [15], Foster and Young [18, 57] and Fudenberg and Harris [19].

<sup>6</sup>There is an implicit convergence assumption here. We work only with simple games to avoid running afoul of convergence problems. Like the short and long runs of the conventional theory of the firm, our medium, long and ultralong runs are economic rather than calendar concepts of time and will be applicable in some but not all cases.

many such shocks will have occurred as to produce a stationary distribution over states of the model.<sup>7</sup>

Our concern in this paper is with the long run. Binmore, Samuelson and Vaughan [9, Theorem 1], using techniques introduced by Börgers and Sarin [12] and Boylan [14, 13], show that the deterministic differential equation (1) provides an arbitrarily good long-run approximation of the behavior of the stochastic process by which strategies are adjusted, so long as we are interested in large populations.<sup>8</sup>

Two assumptions are crucial in the development of this model. First, the underlying stochastic selection process is a Markov process, with strategy revisions depending on only the current state. This is a strong assumption, as it may require players to forsake a long history of experience in order to react to an idiosyncratic experience in their most recent play. At this point we follow much of the learning literature in focussing on Markov models because of their tractability, but consider the relaxation of this assumption an important topic for future work. Second, the resulting differential equation (1) must be smooth (continuously differentiable). Hence, pure best-response behavior, where even an arbitrarily small difference in payoffs suffices to switch all agents to the high-payoff strategy, is excluded. We consider this a realistic assumption. We do not think that dramatic changes in behavior are prompted by arbitrarily small differences in payoffs. Instead, we expect people to be more likely to switch strategies as the payoff differences from doing so increase, and expect this relationship to be reasonably approximated by a smooth learning process.

**Drift.** We reserve the term “noise” for the random elements modeled by  $G$  in the underlying stochastic process, and speak of “drift” when discussing the deterministic term  $g$  that this noise contributes to (1). Given that we are interested in the case in which drift is very small, two basic questions

---

<sup>7</sup>The equilibrium selection theories of Young [56] and Kandori, Mailath and Rob [26] are ultralong-run theories according to this definition.

<sup>8</sup>Binmore and Samuelson [7] study the ultralong run. Binmore, Samuelson and Vaughan’s [9] long-run approximation result is established for the special case of a one-dimensional state space, but their argument is easily adapted to our more general case. The result is that for any time  $T$  and any  $\epsilon$ , we can choose sufficiently large  $N$  and sufficiently small  $\tau$  (in particular, small enough that  $\tau N^2$  is small enough) that the behavior of the stochastic strategy adjustment model over the interval  $[0, T]$  is within  $\epsilon$  of the expected value given by (1) with probability at least  $1 - \epsilon$ . In [9], we show that (1) does not suffice for an ultralong-run analysis.

arise. When we can simply ignore the drift and replace (1) with

$$\frac{dz}{dt} = f(z)? \tag{4}$$

What can be said when drift cannot be ignored?

In understanding the answers we offer to these questions, it is important to understand that our interest lies in the long-run behavior of the system rather than its medium-run or ultralong-run behavior. To study the medium run, it is enough to fix a value of  $T > 0$  and to study the behavior of (1) for  $0 \leq t \leq T$ . Over such a restricted range, the solutions to (1) and (4) can be made arbitrarily close by taking  $\lambda$  small enough. Sánchez [45, Theorem 6.3.1] proves the following well-known continuity property of differential equations:

**Lemma 1** *Let  $z(t, z(0), \lambda)$  solve (1) given initial condition  $z(0)$  and drift level  $\lambda$ , and let  $z(t, z(0), 0)$  solve (4) given initial condition  $z(0)$ . Then for any  $T > 0$  and  $\epsilon > 0$ , there exists  $\lambda(T, \epsilon)$  such that for if  $\lambda < \lambda(T, \epsilon)$  and  $t < T$ , then  $\|z(t, z(0), 0) - z(t, z(0), \lambda)\| < \epsilon$ .*

However, Lemma 1 does not tell us that the long-run behavior of (1) and (4) are similar when  $\lambda$  is small. As the example of the Chain-Store Game studied shortly demonstrates, the asymptotic behavior of the solution to (4) need not approximate the asymptotic behavior of a solution to (1) even when  $\lambda$  is small. In brief, the limits  $t \rightarrow \infty$  and  $\lambda \rightarrow 0$  do not commute.

Just as it is important not to confuse medium-run and long-run considerations, so it is important not to confuse long-run and ultralong-run considerations. To study what happens in the ultralong run, one must examine the asymptotic behavior of the original Markov process directly—rather than studying the deterministic approximation (1) obtained by taking expectations. It is well known that the stationary distribution of a Markov process can be radically changed by minute changes in one of its transition probabilities, especially if this makes a zero probability positive. Young [56] and Kandori, Mailath and Rob [26] exploit precisely this dependence to obtain their strong, ultralong-run, equilibrium-selection results. But these results are obtained at the expense of expected waiting times that may be very long indeed (Ellison [17], Binmore and Samuelson [7], Binmore, Samuelson and Vaughan [9]). For many purposes, the equilibrium selected in the long run may therefore be of only limited interest. For this reason, this paper focuses on the equilibrium that first captures the process—the equilibrium selected in the long run. If the population size is sufficiently large, this equilibrium is predicted with high probability by the asymptotics of the deterministic process (1).

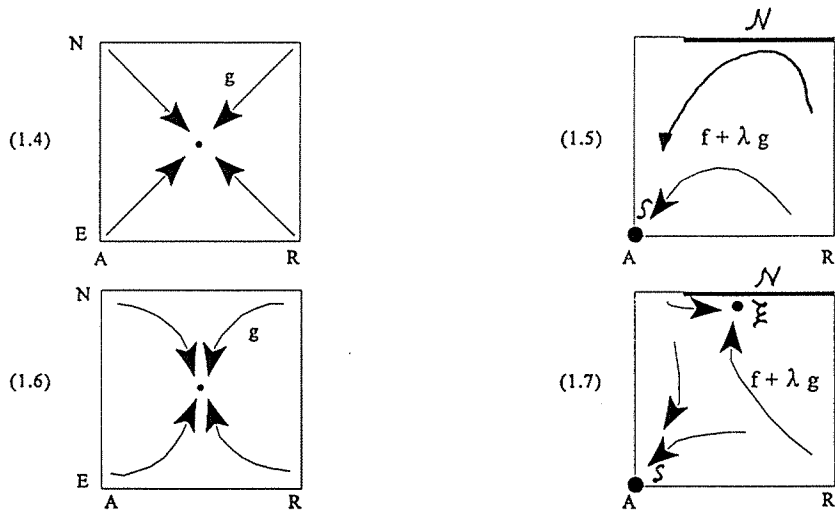
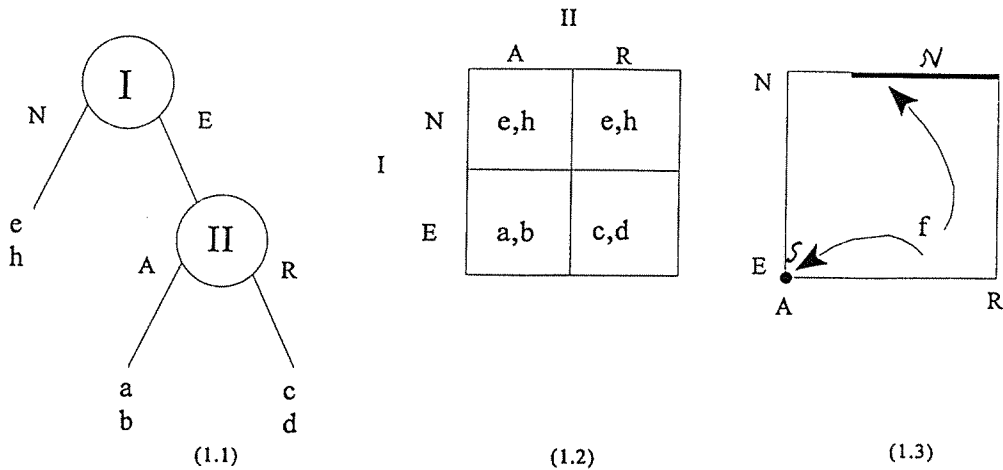


Figure 1: Chain-Store Game

**The Chain-Store Game.** Selten's [47] Chain-Store Game provides an example in which the asymptotic behavior of solution of (4) need not approximate that of (1) as  $\lambda \rightarrow 0$ . The extensive form of this game is shown in Figure 1.1. Player *I* moves first, choosing to enter (*E*) a market or not (*N*). If *I* enters, player *II* can acquiesce (*A*) or resist entry (*R*). The payoffs satisfy the inequalities  $a > e > c$ , so that the entrant prefers to enter if the chain store acquiesces but prefers to stay out if the chain store resists, and  $b > d$ , so that the chain store prefers acquiescing to resisting. Figure 1.2 shows the normal form of this game.

We assume that the dynamics  $\dot{z} = f(z)$  are regular and monotonic in the sense of Samuelson and Zhang [44] so that we have some idea of how to draw the phase diagram.<sup>9</sup> Figure 1.3 shows a phase diagram. The horizontal

<sup>9</sup> $f$  is monotonic if, for strategies  $i$  and  $j$  for player  $\ell$ ,  $\pi_i(z) > \pi_j(z) \iff f_i(z)/z_i > f_j(z)/z_j$ , where  $\pi_i(z)$  is the average payoff to strategy  $i$  in state  $z$  and  $z_i$  is the proportion

axis measures the proportion of agents in population  $II$  playing  $R$  while the vertical axis measures the proportion of population  $I$  agents playing  $N$ . We can see two types of equilibria in Figure 1.3. There is a subgame perfect equilibrium (denoted by  $S$ ) in which player  $I$  enters and  $II$  acquiesces. There is also a component of Nash equilibria (denoted by  $\mathcal{N}$ ) that are not subgame perfect, in which player  $I$  does not enter and player  $II$  resists entry with probability at least  $(a - e)/(a - c)$ .

Figure 1.3 shows that, depending upon the initial state, the dynamics will converge either to the subgame-perfect equilibrium or to the Nash equilibrium component  $\mathcal{N}$ . The subgame-perfect equilibrium is asymptotically stable, the Nash equilibria are not.<sup>10</sup> The interpretation of the model, and especially our assessment of the component  $\mathcal{N}$ , then hinges on the stability properties of the component  $\mathcal{N}$ . We address this question in two stages.

First, how does  $\mathcal{N}$  fare when faced with perturbations caused by factors excluded from the model that led to the selection dynamics? In the long run, these perturbations appear in the form of drift. Figure 1 shows two specifications of drift and the corresponding phase diagrams for the Chain Store Game. In Figure 1.5, the addition of drift yields a system that has a unique, asymptotically stable state that attracts the entire space: namely, the subgame-perfect equilibrium. In Figure 1.7, two asymptotically stable states exist, with one (denoted by  $\xi$ ) being a Nash equilibrium that is not subgame perfect. Drift can thus make a difference in the long-run behavior of the system.

Second, given that we are interested in the long run and not the ultralong run, why are these potentially different implications of drift relevant? Under the unperturbed dynamics, the long-run behavior is that trajectories from some initial conditions converge to  $\mathcal{N}$  and some do not. When the drift is as in Figure 1.6, we again have this conclusion. In Figure 1.5, some trajectories converge to  $S$  without coming near  $\mathcal{N}$ . Other trajectories do not converge to  $\mathcal{N}$ , but do come very close to  $\mathcal{N}$ . In addition, the rate of movement of the dynamic system is very slow near the component  $\mathcal{N}$ , because payoff

---

of the population playing strategy  $i$ .  $f$  is regular if the  $\lim_{z \rightarrow z^*} f_i(z)/z_i$  exists and is finite when  $z_i^* = 0$ . This in turn requires  $f_i(z^*) = 0$ , or, equivalently, that faces of the state space are forward invariant.

<sup>10</sup>Following Hofbauer and Sigmund [25, p. 51], a stationary state  $z$  of a dynamic process is *stable* if, for any open set  $V$  with  $z \in V$ , there is an open set  $U$  with  $z \in U \subset V$  such that any orbit beginning in  $U$  is contained in  $V$ . A stationary state  $z$  is *asymptotically stable* if it is stable and there is an open set  $W$  with  $z \in W$  such that any orbit beginning in  $W$  converges to  $z$ .

differences are small and hence learning forces are weak near a component of equilibria, and because drift is also small. As a result, the trajectories that come close to  $\mathcal{N}$  will spend a very long time near  $\mathcal{N}$ . Thus  $\mathcal{N}$  is interesting in the medium run, even when the equilibrium that will be selected in the long run is  $S$ . Since the medium run can be very long, why must we go further to a long-run analysis?<sup>11</sup>

The difference between the cases shown in Figures 1.5 and 1.7, and the relevance of drift, hinges on what is meant by small and by long run. In Figure 1.5, an argument for the relevance of  $\mathcal{N}$  over a long period of time must be driven by a belief that drift is sufficiently small and the initial condition is sufficiently close to state  $(1, 1)$ , since otherwise the system will leave the neighborhood of  $\mathcal{N}$  too quickly. The longer is the time period of interest, the more stringent are these requirements. In contrast, the conditions for the applicability of  $\mathcal{N}$  are much less demanding in Figure 1.7, requiring only that the initial condition lie in the basin of attraction of  $\xi$ , at which point we can be confident that the system will not stray from the vicinity of  $\mathcal{N}$  over any arbitrarily long time period.<sup>12</sup> Studying drift is important because drift can make an outcome such as  $\mathcal{N}$  a good prediction of long-run behavior under a much wider range of circumstances.

Do we have any reason to expect drift to look like Figure 1.6? The heart of Binmore, Gale and Samuelson's [4] analysis of the Ultimatum Game is an argument that drift may take an analogous form. This argument in turn depends upon the presumption that players or populations are likely to be less susceptible to drift when the payoff consequences of their actions are larger. McKelvey and Palfrey [33] similarly suggest modelling players as being more likely to make mistakes or experiment with new choices if the payoff implications are small. The idea goes back to Myerson's [35] proper equilibrium.

---

<sup>11</sup>The analysis of Roth and Erev [42] is based on speed-of-adjustment arguments of this type.

<sup>12</sup>The qualitative features of these phase diagrams, including the characteristics of the stationary states, are preserved no matter how small is the drift level  $\lambda$ . In particular, Proposition 2 below shows that the basin of attraction of  $\xi$  does not shrink as drift rates get small.

### 3 When can Drift be Ignored?

The Chain-Store Game shows that there are cases in which drift matters. We accordingly turn to the question of *when* it matters. We require no assumptions on the learning process in this section other than that the resulting differential equations be continuously differentiable and that the state space be forward invariant, though we also assume that the dynamics are monotonic and regular when drawing the phase diagrams in Figures 2–3.

**Hyperbolic stationary states.** Our first observation is that we can ignore drift and work with  $\dot{z} = f(z)$  rather than  $\dot{z} = f(z) + \lambda g(z)$  when dealing with hyperbolic stationary states of  $f$ . A stationary state of a differential equation is *hyperbolic* if the Jacobian matrix of the differential equation, evaluated at the stationary state, has no eigenvalues with zero real parts. Hyperbolic stationary states are isolated and are either sources, saddles or sinks (Hirsch and Smale [24, ch. 9]).<sup>13</sup>

The first statement in the following Proposition is immediate from the continuity of  $f$  and  $g$  on the compact set  $Z$ , while the second statement follows from Hirsch and Smale [24, Theorems 1–2, p. 305]:<sup>14</sup>

#### Proposition 1

(1.1) *For any  $\epsilon > 0$ , there exists  $\lambda(\epsilon)$  such that for any  $\lambda < \lambda(\epsilon)$ , every stationary state of  $f + \lambda g$  lies within  $\epsilon$  of a stationary state of  $f$ .*

(1.2) *Let  $z$  be a hyperbolic stationary state of  $f$ . Then  $f + \lambda g$  has a hyperbolic stationary state  $z(\lambda)$  that converges to  $z$  as  $\lambda \rightarrow 0$  through an appropriate sequence of values, with each  $z(\lambda)$  being a sink (saddle) [source] if and only if  $z$  is a sink (saddle) [source].*

Proposition (1) indicates that, if we are working with hyperbolic stationary states of  $f$ , then we can ignore drift. The stationary states of  $f$  provide approximate information concerning stationary states of  $f + \lambda g$  that lie nearby and are of the same type. The approximation becomes arbitrarily sharp as  $\lambda$  gets small. Nonhyperbolic stationary states, however, do not have this “structural stability.” An arbitrarily small change in the dynamic

---

<sup>13</sup>Nonhyperbolic stationary states need not be isolated and isolated stationary states need not be hyperbolic.

<sup>14</sup>For convenience, we will often write simply  $f$  and  $f + \lambda g$  for  $\dot{z} = f(z)$  and  $\dot{z} = f(z) + \lambda g(z)$ .

system, or equivalently an arbitrarily small amount of drift, can completely change the nature of a nonhyperbolic stationary state.

This observation may appear to resolve the question, since it is often said that almost all dynamic systems have only hyperbolic stationary states.<sup>15</sup> However, the economics of the applications to which learning models are applied often force us to confront nonhyperbolic stationary states. The equilibria in the component  $\mathcal{N}$  of the Chain-Store Game are not hyperbolic stationary states. This is not exceptional. Every Nash equilibrium that does not reach every information set (excluding some games featuring fortuitous payoff ties) fails to be isolated under all of the familiar selection dynamics, and hence fails to be a hyperbolic stationary state. Drift matters in such cases, no matter how small it is.

Unreached information sets are notorious as the breeding ground for equilibrium refinements, the heart of a refinement concept lying in the restrictions imposed on what players do or believe at such information sets. Thus, when equilibrium refinements are at issue, drift matters.

**Asymptotically Stable Components.** If we are forced to deal with components of stationary states, we might hope that the component as a whole satisfies some stability property.

We begin with an example constructed by making two modifications to the payoffs of the Chain-Store Game. In the first modification, let the incumbent be one of Milgrom and Roberts' [34] or Kreps and Wilson's [30] "tough guys," who gets a higher payoff from fighting than from not fighting. Second, let the entrant be unprofitable, in the sense that the entrant prefers to stay out of the market even if the incumbent acquiesces. Then we obtain the version of the Chain-Store Game shown in Figure 2. Figure 2.3 shows the corresponding phase diagram for a regular, monotonic  $f$ . There is a component  $\mathcal{N}$  of Nash equilibria that are nonhyperbolic stationary states. As long as  $L$  is played with positive probability in the initial state, the system will converge to a point in  $\mathcal{N}$ . In the presence of drift, the system will converge to a point close to  $\mathcal{N}$ , and this point will be arbitrarily close for arbitrarily small drift levels. The component  $\mathcal{N}$  satisfies a set version of asymptotic stability, in that it attracts the trajectories from all nearby

---

<sup>15</sup>The Peoxito theorem (Hirsch and Smale [24, p. 314]) shows that for two-dimensional systems, there is a precise sense in which "almost all" dynamic systems have only hyperbolic stationary states. A similar result holds in higher dimensions for certain classes of dynamic systems, such as linear and gradient systems [24, pp. 313-315].



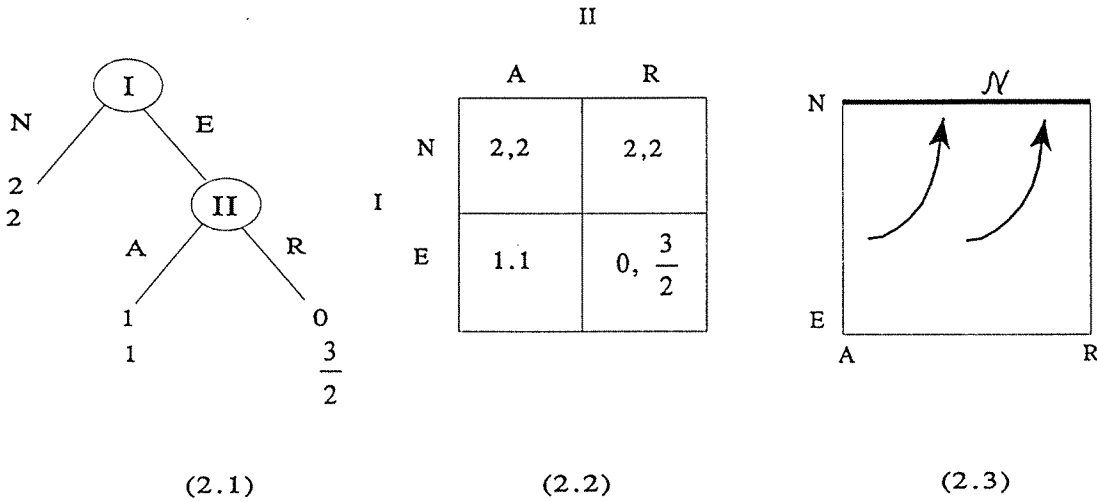


Figure 2: Chain-Store Game: unprofitable entrant and tough chain store

states.

The following definition is from Bhatia and Szegö [3, Def. 1.5, p. 58].<sup>16</sup>

**Definition 1** Let  $z(t, z(0))$  be the solution of a differential equation defined on  $Z$ . A closed set  $C \subset Z$  is asymptotically stable if, for every open set  $V \subset Z$  containing  $C$ , there is an open set  $U$  with  $C \subset U \subset V$  such that if  $z(0) \in U$ , then  $z(t, z(0)) \in V$  for  $t \geq 0$  and  $\lim_{t \rightarrow \infty} z(t, z(0)) \in C$ .

One's intuition is that drift should be irrelevant for the study of asymptotically stable sets in the same sense that drift can be neglected when studying hyperbolic stationary states. This intuition can be made precise by applying Proposition 2 below.

Guckenheimer and Holmes [20, pp. 258–259] advocate an approach to nonhyperbolic stationary states analogous to the recommendation that one

<sup>16</sup>The set  $C$  is restricted to be closed in this definition to ensure that  $C$  lies strictly inside the open set  $U$ , so that an arbitrarily small perturbation cannot take the system from  $C$  to outside the set  $U$ . Ritzberger and Weibull [40], Schlag [46] and Swinkels [50] examine components that satisfy a set version of asymptotic stability. A similar motivation but different techniques appear in Ritzberger [39], who introduces the idea of an essential component, where (very roughly) a component is essential if all nearby games have nearby equilibria.

direct attention to asymptotically stable sets. They suggest that if stationary states fail to be hyperbolic, then one should not necessarily abandon or embellish the model in a desperate attempt to achieve hyperbolicity. Instead, the existing model may well be the best description of the physical system to be studied, and the failure of hyperbolicity should serve only as a caution to place confidence only in those features of the model which are robust to all perturbations.

The asymptotic stability of  $\mathcal{N}$  in the Chain-Store Game of Figure 2 is a statement about equilibrium outcomes, namely that the entrant will not enter, rather than a statement about strategies. Nothing has been said about out-of-equilibrium behavior, or what the chain-store would do if entry occurred. In many cases, we will be unconcerned with out-of-equilibrium behavior because it is unobserved and has no economic consequences. In other cases, we may encounter asymptotically stable components with the property that the payoffs of at least some players vary across states in the component. We must then be concerned about which element in the component appears, for which there is no alternative to delving into the details of the drift process.

**Relatively Asymptotically Stable Components.** Asymptotic stability may be stronger than necessary for a component of equilibria to be deemed worthy of attention. Figure 3 shows a second modification of the Chain-Store Game. The chain store is again tough, but entry is profitable if the chain store acquiesces. There is a unique component  $\mathcal{N}$  of Nash equilibria in which the entrant does not enter and the chain store fights entry with probability at least  $\frac{1}{3}$ . This component is not asymptotically stable. Instead, any state in which no entry occurs is a stationary state, so that there exist stationary states arbitrarily close to  $\mathcal{N}$  that are not contained in  $\mathcal{N}$ . However, every initial condition that lies in the interior of the state space yields a trajectory that converges to  $\mathcal{N}$  (and that does not stray too far away if it starts nearby).

The following is a slight modification of Definition 5.1 of Bhatia and Szegö [3, p. 99].

**Definition 2** *Let  $z(t, z(0))$  be the solution of a differential equation on state space  $Z$  given initial condition  $z(0)$ . A closed set  $C \subset Z$  is asymptotically stable relative to  $W \subset Z$  if, for every open set  $V$  containing  $C$ , there is an open set  $U$  with  $C \subset U \subset V$  such that if  $z(0) \in U \cap W$ , then  $z(t, z(0)) \in V$  for  $t \geq 0$  and  $\lim_{t \rightarrow \infty} z(t, z(0)) \in C$ .*

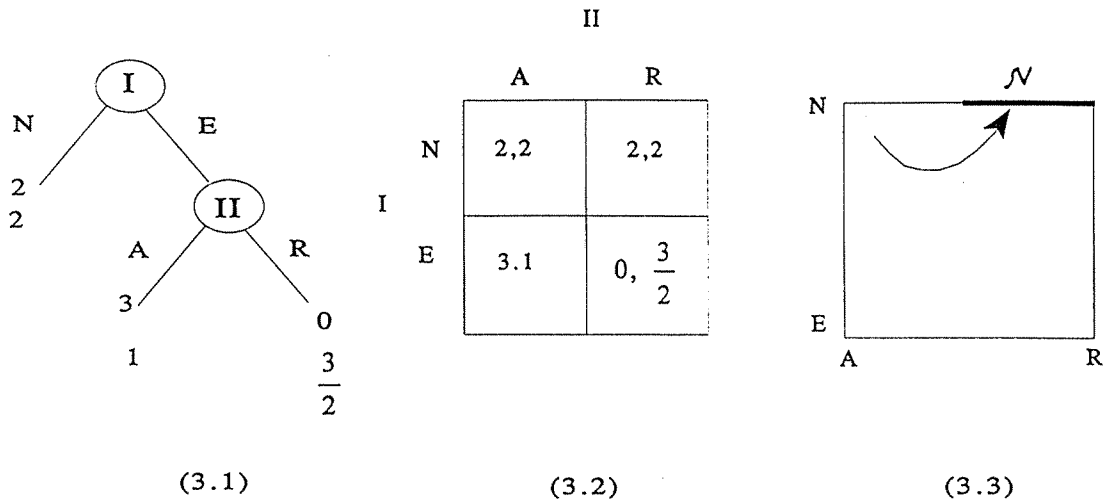


Figure 3: Chain-Store Game: tough chain store

In the case of the Chain-Store Game of Figure 3, the set  $W$  can be taken to be the interior of the state space, in which case we say that  $\mathcal{N}$  is *asymptotically stable relative to the interior*. Given the common presumption that drift points into the interior of the state space, this is an especially interesting case. Proposition 2 below can again be applied to make precise the sense in which the system with drift directs attention to components that are asymptotically stable relative to the interior under the original system.<sup>17</sup>

#### 4 When Drift Matters

Unfortunately, as demonstrated by the Chain-Store Game of Figure 1, we often encounter components of nonhyperbolic stationary states that are not asymptotically stable or asymptotically stable relative to the interior. In these cases, drift matters. Forces that have been excluded from the model

<sup>17</sup>One might suspect that if a component of Nash equilibria is asymptotically stable with respect to the interior, then the specification of drift, and hence Proposition 2, is irrelevant as long as drift is small and inward pointing. However, it is easy to construct an example of a component  $\mathcal{N}$  that is asymptotically stable with respect to the interior and a specification of (arbitrarily small) drift such that trajectories starting from points arbitrarily close to  $\mathcal{N}$  converge to points far away.

$\dot{z} = f(z)$ , on the grounds that they can be safely neglected, are then not negligible for the purposes of long-run prediction.<sup>18</sup> One then cannot evade studying the model  $\dot{z} = f(z) + \lambda g(z)$ .

We restrict our attention to the case when  $f$  and  $g$  are continuously differentiable and the stationary points of  $\dot{z} = f(z) + \lambda g(z)$  are hyperbolic for sufficiently small  $\lambda$ . If this were false, one would argue that yet more unmodeled sources of drift need to be incorporated into the model. We simplify further by assuming that the basin of attraction of the set of all stationary points is the whole space  $Z$ . Periodic orbits and other complicated trajectories are thereby excluded.<sup>19</sup>

Since drift matters, the unperturbed dynamic  $\dot{z} = f(z)$  has a component  $C$  of stationary states that are not hyperbolic. We assume that  $C$  is closed and ask the question: when can we guarantee finding stationary states of  $\dot{z} = f(z) + \lambda g(z)$  close to  $C$  for all sufficiently small  $\lambda > 0$ , and when do they have sufficiently strong stability properties to make them interesting long-run predictions? For a positive answer to the second question, we will require the stationary states to approach  $C$  as  $\lambda \rightarrow 0$  and to lie in the interior of a basin of attraction that does not shrink as  $\lambda \rightarrow 0$ . Figures 1.4 and 1.5 demonstrate that we cannot always ensure the existence of such stationary states.

We begin with a simple example that is generalized in Proposition 2.

**Example 1** For the purposes of this example, let the state space be denoted by  $W$  and a state by  $w$ . Let the continuously differentiable dynamics  $\tilde{f}(w)$  and  $\tilde{g}(w)$  on  $W$  have the property that there exists a component  $\Gamma$  of stationary states of  $\tilde{f}$  that is also a face of  $W$ . Let  $B_\delta(\Gamma)$  be the set of all points in  $W$  whose distance from  $\Gamma$  in the max norm is  $\delta$  or less.<sup>20</sup> Then  $B_0 = \Gamma$  and  $B_1 = W$ . Let  $D_\delta = \partial B_\delta \cap \partial(W \setminus B_\delta)$ . Suppose further that for all sufficiently small  $\delta > 0$ ,<sup>21</sup>

- (a) For all  $w \in B_\delta$ ,  $\tilde{f}(w)$  points into  $B_\delta \setminus D_\delta$  at  $w$ .

<sup>18</sup>We will rarely have the level of confidence suggested by Guckenheimer and Holmes [20, pp. 258–259], believing that we have literally captured the system of interest and that there is no drift to be added to the model.

<sup>19</sup>Complicated behavior of this sort is undoubtedly important in some cases. However, we think it is important to begin by understanding cases in which the dynamics are relatively well-behaved. We find that such simple cases have much to tell us about equilibrium selection.

<sup>20</sup>The max norm is defined by  $|w| = \max_i |w_i|$ .

<sup>21</sup>The sets  $S^\circ$  and  $\partial S$  are respectively the interior and boundary of  $S$ . To say that  $y$  points into  $S$  at  $x$  means that  $x + \epsilon y \in S$  for all small enough  $\epsilon > 0$ .

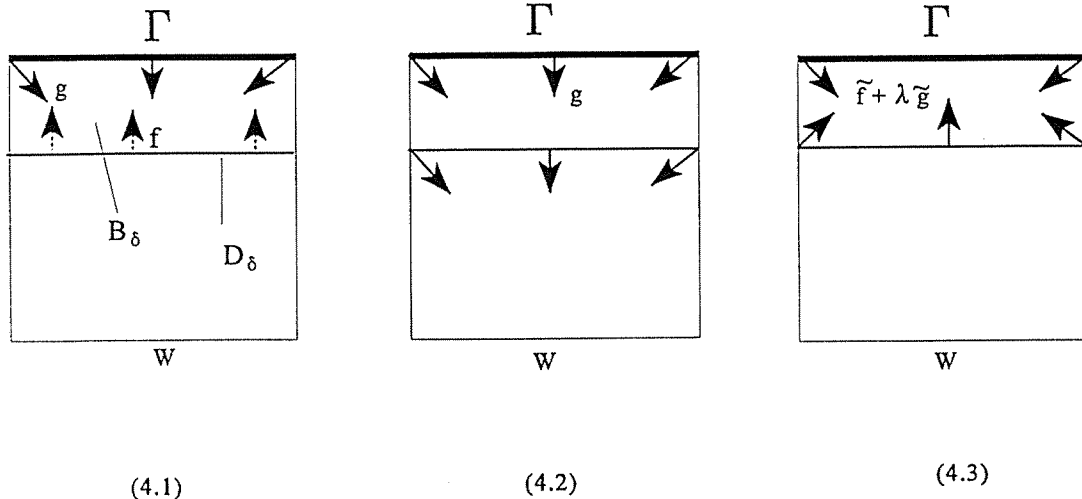


Figure 4: Illustration for Example 1

(b) For all  $w \in \Gamma$ ,  $g(w)$  points into  $B_\delta^o$  at  $w$ .

We illustrate conditions (a) and (b) in Figure 4.1. Since  $g$  is continuous, we deduce from (b) that

(c) There exists  $\delta^* > 0$  such that  $g(w)$  points into  $B_\delta^o$  for all  $w \in B_\delta \setminus D_\delta$  and points into  $W^o$  for all  $w \in B_\delta$ , provided  $0 \leq \delta \leq \delta^*$ .

Condition (c) is illustrated in Figure 4.2.

Conditions (a) and (b) are sufficient to demonstrate that there exists  $\lambda(\delta)$  such that stationary points of  $\dot{z} = \tilde{f}(w) + \lambda\tilde{g}(w)$  can be found in  $B_\delta^o$  whenever  $0 \leq \delta \leq \delta^*$  and  $0 \leq \lambda \leq \lambda(\delta)$ . This result is proved by showing that  $\tilde{f}(w) + \lambda\tilde{g}(w)$  points into  $B_\delta^o$  for all  $w \in B_\delta$  and then appealing to Brouwer's theorem. We illustrate this situation in Figure 4.3. As  $\lambda \rightarrow 0$ , these stationary states approach  $C$ . Since all trajectories through points of  $B_{\delta^*}$  do not leave  $B_{\delta^*}$  and exotic behavior is excluded by hypothesis, the basin of attraction of the set of all stationary states of the perturbed dynamics inside  $B_{\delta^*}$  contains  $B_{\delta^*}$ , no matter how small is  $\lambda$ . The stationary states accordingly satisfy our criteria for being interesting long-run predictions.

To verify that such a  $\lambda(\delta)$  exists, observe that  $\tilde{f}$  is continuous on  $W$  and, for sufficiently small  $\delta^*$ , has zeros in  $B_\delta$  only on  $\Gamma$ . By (a)–(b), we can

therefore find  $\lambda(\delta)$  such that  $\tilde{f}(w) + \lambda(\delta)\tilde{g}(w)$  points into  $B_\delta$  for  $w \in D_\delta$ . Elsewhere in  $B_\delta$ , it is enough to observe that (a) and (c) imply that  $\tilde{f}(w)$  and  $\tilde{g}(w)$  both point into  $B_\delta$ .

It remains to apply Brouwer's theorem to the function  $\Psi : B_\delta \rightarrow B_\delta$  defined, for small enough  $\epsilon$ , by  $\Phi(w) = w + \epsilon(\tilde{f}(w) + \lambda\tilde{g}(w))$ , where  $0 < \delta < \delta^*$  and  $0 < \lambda < \lambda(\delta)$ .<sup>22</sup> A fixed point  $w^*$  of  $\Psi(w)$  is a stationary state of  $\dot{w} = \tilde{f}(w) + \lambda\tilde{g}(w)$ .  $\square$

We now generalize the argument of this example to more complex situations. Our technique is to look for cases that are structurally equivalent to the example. As before, we let  $\dot{z} = f(z) + \lambda g(z)$  be continuously differentiable on the state space  $Z$ . Let  $C \subset Z$  be a closed component of stationary states of  $\dot{z} = f(z)$ . Let  $W$  and  $\Gamma$  be as in Example 1.

**Proposition 2** *Suppose there exists a differentiable injection  $\phi : W \rightarrow Z$  with differentiable inverse  $\Phi$  such that  $C = \phi(\Gamma)$  and such that  $\tilde{f}(w) = \Phi'(\phi(w))f(\phi(w))$  and  $\tilde{g}(w) = \Phi'(\phi(w))g(\phi(w))$  satisfy (a)–(b). Then there exists  $\delta^* > 0$  and  $\lambda(\delta) > 0$  for all  $0 < \delta < \delta^*$  such that stationary states of  $f(z) + \lambda g(z)$  can be found in  $\phi(B_\delta^o)$  whenever  $0 < \delta < \delta^*$  and  $0 < \lambda < \lambda(\delta)$ . The basin of attraction of the set of such stationary points contains  $\phi(B_{\delta^*})$ .*

**Proof** By Example 1, there exist  $\delta^* > 0$  and  $\lambda(\delta)$  such that  $\tilde{f} + \lambda\tilde{g}$  has a stationary state  $w^*$  in  $B_\delta^o$  for  $0 < \delta < \delta^*$  and  $0 < \lambda < \lambda(\delta)$ . But then  $z^* = \phi(w^*)$  is a stationary state of  $f + \lambda g$  because  $\phi$  is differentiable and hence  $\Phi'$  is nonsingular.  $\square$

In our applications, the mapping  $\phi$  will be chosen so that the set  $\phi(B_\delta)$ , which we will simply call  $B_\delta$ , will be the set of points in the basin of attraction of  $C$  that lie with distance  $\delta$  of  $C$ . The mapping  $\phi$  is a diffeomorphism (i.e., a differentiable bijection with differentiable inverse) between the state space and a set containing  $C$ . To say that  $\phi$  causes conditions (a) to hold is then to say that near the component  $C$ , the basin of attraction of  $C$  is nicely behaved in the sense that the basin locally looks like the state space  $Z$  and the trajectories in this basin approach  $C$  sufficiently directly, meaning that on the boundary of  $B_\delta(C)$ , the learning dynamics lead into  $B_\delta(C)$ . These are counterparts of the characteristics of a hyperbolic sink that are used to show that a perturbed dynamic must have a nearby sink. Condition (b) is straightforward, indicating that drift can push the system off the component

---

<sup>22</sup>  $\epsilon$  must be chosen sufficiently small here to ensure that  $\Psi(w) \in B_\delta$  for all  $w \in B_\delta$ .

$C$ , but in so doing must push the system into the basin of attraction of  $C$ .

The conditions of Proposition 2 are sufficient but not necessary, and we have refrained from seeking the weakest possible sufficient conditions in order to obtain conditions that are easily interpreted. In the examples of the next section a partial converse of this theorem holds, in that the selection dynamic  $f$  satisfies (a) in each case and the conclusions of Proposition 2 hold if and only if  $g$  satisfies (b).

In the examples that follow, we will speak simply of “condition (a)” holding and “condition (b)” holding, by which we mean that there exists a mapping  $\phi$  with the desired properties that causes these conditions to hold. We will illustrate these conditions for the dynamics  $f$  and  $g$  near the component  $C$  in state space  $Z$ , meaning again that we will illustrate dynamics for which an appropriate  $\phi$  exists. Figure 5 illustrates conditions (a)–(b) in the case of the component  $\mathcal{N}$  of Nash equilibria in the Chain Store Game of Figure 1. Figure 5.2 corresponds to the drift shown in Figure 1.6, in which case (a)–(b) hold. Figure 5.1 corresponds to the drift shown in Figure 1.4. In this case one easily finds sets  $C$  satisfying (a), but there is no set  $C$  for which (b) is satisfied.

It is important to note that in Figure 5.2, we have chosen  $C$  to be a *subset* of the Nash equilibrium component. It will typically be the case that conditions (a)–(b) are satisfied not by an entire component of Nash equilibria but by a subset of that component. But how can a subset  $C$  of a component  $\mathcal{N}$  have robust stability properties, since there must be other stationary points in  $\mathcal{N} \setminus C$  that are arbitrarily close? The component  $C$  has robust stability properties in the perturbed dynamic because the drift operates on  $C$  to push the system back into the basin of attraction of  $C$  and away from points in  $\mathcal{N} \setminus C$ . This is where drift plays its essential role.

What reason does this proposition give us for being interested in the set  $C$ ? If (a)–(b) hold, then for any small  $\delta$  and sufficiently small  $\lambda$ , the system  $f + \lambda g$  has stationary states that lie within distance  $\delta$  of  $C$ . In addition, these stationary states have a basin of attraction which contains  $B_{\delta^*}(C)$  and hence does not shrink as  $\lambda$  shrinks. Some subset of  $C$  thus provides a good approximation of the local limiting behavior of the dynamic  $f + \lambda g$  for small  $\lambda$ . In the examples of the next section, if (a)–(b) hold, then there is a *unique* stationary state near the component  $C$  and this stationary state is asymptotically stable.

If conditions (a)–(b) hold, then the conclusion of Proposition 2 holds no matter how small  $\lambda$  becomes, i.e., no matter how insignificant is drift. Why

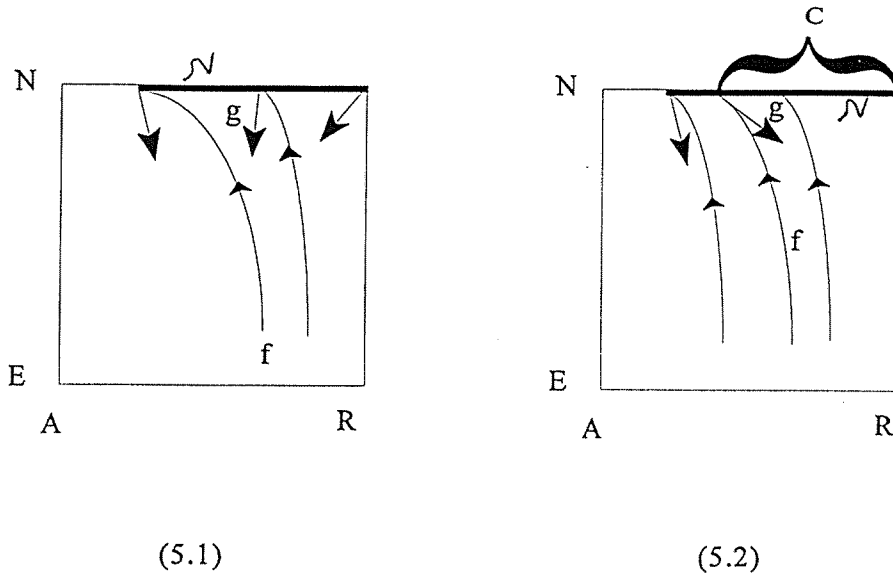


Figure 5: Conditions of Proposition 2

do we expect even very small amounts of drift to have a significant effect on the behavior of the dynamic system near  $C$ ? The answer is that the selection dynamics approach zero as we near  $C$ , because  $C$  is a component of stationary states. Even small amounts of drift can then overwhelm the selection process. The question is whether this drift tends to push the system toward or away from  $C$ . If (b) holds, then the drift pushes the system toward  $C$  and we have stationary states of the process with drift near  $C$ .

## 5 Examples

In this section, we illustrate the effect of drift with some examples. We assume that the selection process is monotonic and regular. Some assumption of this nature is required if there is to be any relationship between outcomes of the learning process and conventional equilibrium notions. We find this assumption appealing, though much work remains to be done before the links between such behavior and the underlying learning processes are clear. We also assume that  $g$  is completely mixed, or equivalently that it points inward on the boundary of the state space. We find this assumption especially natural, as the unmodeled forces captured by drift are likely to contain some



factors that have nothing to do with payoffs in the game and are capable of introducing any strategy.

### 5.1 Backward Induction

We first return to the Chain-Store Game of Figure 1. Let  $(r, n)$  denote a point in the unit square, where  $n$  is then the proportion of population  $I$  playing  $N$  (not enter) and  $r$  is the proportion of population  $II$  playing  $R$  (or resisting entry). For a point  $z \in Z$ , the selection process then specifies a pair  $(\dot{r}(z), \dot{n}(z)) = (f_r(z), f_n(z))$ .

To investigate the possibility of stationary states near the component  $C$ , we examine the components of the vector  $f$  at states  $z \in C$ . We will be particularly interested in the ratio  $f_n(z)/f_r(z)$ , and its relationship to the analogous ratio for  $g$ .<sup>23</sup>

**Proposition 3** *Let  $g$  be completely mixed and let  $f$  be the monotonic.*

(3.1) *Suppose there exists  $\theta \geq 1/3$  such that*

$$\frac{f_n(\theta, 1)}{f_r(\theta, 1)} < \frac{g_n(\theta, 1)}{g_r(\theta, 1)} < 0. \quad (5)$$

*Then there exists  $\phi$  satisfying the conditions of Proposition 2 such that (a)–(b) are satisfied with  $C = \{(r, 1) : r \geq \theta\} \equiv C_\theta$ .*

(3.2). *Suppose there is no  $\theta \geq 1/3$  such*

$$\frac{f_n(\theta, 1)}{f_r(\theta, 1)} \leq \frac{g_n(\theta, 1)}{g_r(\theta, 1)} \leq 0. \quad (6)$$

*Then for sufficiently small  $\lambda$ ,  $f + \lambda g$  has a unique stationary state that converges to  $(0, 0)$  as  $\lambda$  converges to zero.*

**Proof** (3.1) Let  $\theta > \frac{1}{3}$ . Let  $B(C_\theta)$  be the basin of attraction of  $C_\theta$  and let  $B_\delta$  be  $\{(r, n) \in B(C_\theta) : n > 1 - \delta\}$ . Such a set is illustrated in Figure 5.2. For sufficiently small  $\delta$ , there is a function  $\phi : (1 - \delta, 1) \rightarrow (\theta, 1)$  such that  $(\phi(n), n)$  is in the basin of attraction of  $(\theta, 1)$ . This function then describes the left side of the boundary of  $B_\delta$  (again, see Figure 5.2). Then the function

<sup>23</sup>Because  $z \in C$  is a stationary state of  $f$ , we must define  $f_n(z)/f_r(z) = \lim_{z_k \rightarrow z} f_n(z_k)/f_r(z_k)$  for some sequence along which  $f_r(z_k) \neq 0$ . Similarly, let  $g_n(z)/g_r(z) = \lim_{z_k \rightarrow z} g_n(z_k)/g_r(z_k)$ . The monotonicity of  $f$  and the inward-pointingness of  $g$ , along with the differentiability of  $f$  and  $g$ , ensures that this is well defined for  $x \in C$ , in the sense that the limits are independent of the sequences  $\{z_k\}$ .

$\Phi((r, n)) = ((n - \phi(n))/(1 - \phi(n)), (n - (1 - \bar{\delta}))/\bar{\delta})$  is a diffeomorphism for which (a) holds for any monotonic dynamic. The assumption that (5) holds then supplies condition (b), giving the result.

(3.2) Suppose the consequent of the statement fails. Then there must exist a sequence of values of  $\lambda_n$  approaching zero and a sequence of stationary points  $z_k$  converging to an element of  $C$ . Hence, there must be a sequence  $z_k$  such that  $f(z_k) + \lambda_k g(z_k) = 0$  with  $z_k$  converging to  $C$ . Because  $f(z_k) + \lambda g(z_k) = 0 \Rightarrow g_n(z_k)/g_r(z_k) = f_n(z_k)/f_r(z_k)$ , this contradicts the inability to satisfy (6).  $\square$

Condition (3.1) is the statement that we can find a subset  $C_\theta$  of Nash equilibria in which the entrant stays out, which has the property that, on this set, the drift dynamics  $g$  point into the basin of attraction of  $C_\theta$  under the selection process  $f$ . The set  $C_\theta$  is an interval connecting  $(\theta, 1)$  and  $(1, 1)$ , and the key to verifying that drift points into the basin of attraction under the selection process is verifying this property at the endpoint  $(\theta, 1)$ . The existence of such a set then hinges upon finding a value of  $\theta$  for which (5) holds. Figure 5.2 illustrates a case in which such a  $\theta$  exists. Condition (3.2) is the statement that such a  $\theta$  cannot be found and corresponds to Figure 5.1. Here, it is impossible to find a subset of  $\mathcal{N}$  on which drift points into the basin of attraction under the selection process, and there is no stationary point near  $\mathcal{N}$  in the presence of drift.

Proposition 3 thus indicates that when examining the Chain Store Game, we can simply check the relative slopes of the learning and drift processes on the component of Nash equilibria  $\mathcal{N}$ . The component  $\mathcal{N}$  is worthy of our attention if and only if the slope of the drift process is flatter, in the sense that the drift process points into the basin of attraction of a subset of the component of Nash equilibria, where “worthy of attention” means that there are nearby stationary states of the process with drift whose basin of attraction does not shrink as the drift level shrinks. For the case of the replicator dynamics, Binmore, Gale and Samuelson [4] show that if (3.1) holds then there is a unique stationary point close to  $C_\theta$ , which is a sink.

## 5.2 Outside Options

Consider the game shown in Figure 6. The shape of the extensive form of this game prompts us to refer to it as the “Dalek Game.”

The Dalek Game has two components of Nash equilibria, including a strict Nash equilibrium given by  $(M, L)$  with payoffs  $(9, 3)$  and a component  $\mathcal{N}$  of equilibria with payoffs  $(7, 4)$  in which player  $I$  takes the outside option

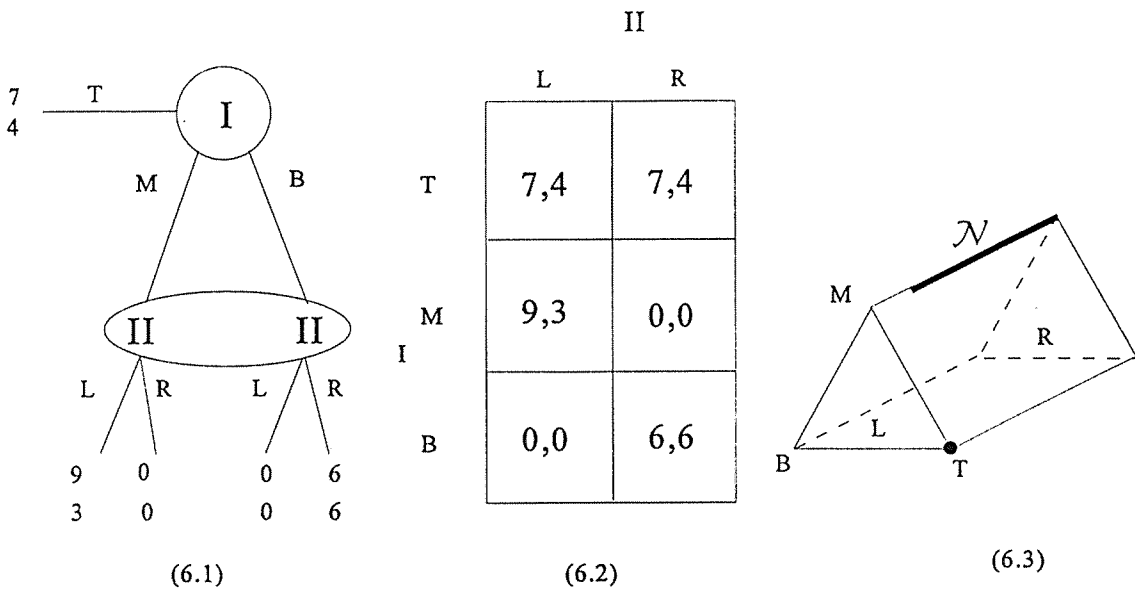


Figure 6: Dalek Game

(plays  $T$ ) and player  $II$  plays  $R$  with probability at least  $7/9$ . The former is a (hyperbolic) sink under regular, monotonic dynamics while the stationary states in the latter component are not hyperbolic.

It is common to argue that forward induction restricts attention to the equilibrium  $(M, L)$  in this game.<sup>24</sup> How does this forward induction argument fare in our terms? First, it is straightforward to verify that condition (a) is satisfied in this case and hence it follows immediately from Proposition 2 that:

**Proposition 4** *If there is a state  $z^* \in \mathcal{N}$  such that  $g(z^*)$  points into the interior of the basin of attraction of the set  $C = \{z \in \mathcal{N} : z_R \geq z_R^*\}$ , then there exists a  $\phi$  satisfying (a)-(b).*

In the presence of drift, we thus have good reason to be interested in

<sup>24</sup>For example, we might appeal to the iterated elimination of weakly dominated strategies (cf. Kohlberg and Mertens [29]).  $B$  is strictly dominated for player  $I$ . Removing  $B$  causes  $R$  to be weakly dominated for player  $II$ , the removal of which causes  $T$  to be weakly dominated for player  $I$ , leaving  $(M, L)$ . Alternatively, we could appeal to the never-weak-best-response criterion (Kohlberg and Mertens [29]) or to the forward induction reasoning of van Damme [53, 54] (in an equivalent but different extensive form in which player  $I$  first chooses between  $T$  and  $\{M, B\}$  and then chooses between  $M$  and  $B$ ) or to the normal form variant of this reasoning given in Mailath, Samuelson and Swinkels [31].

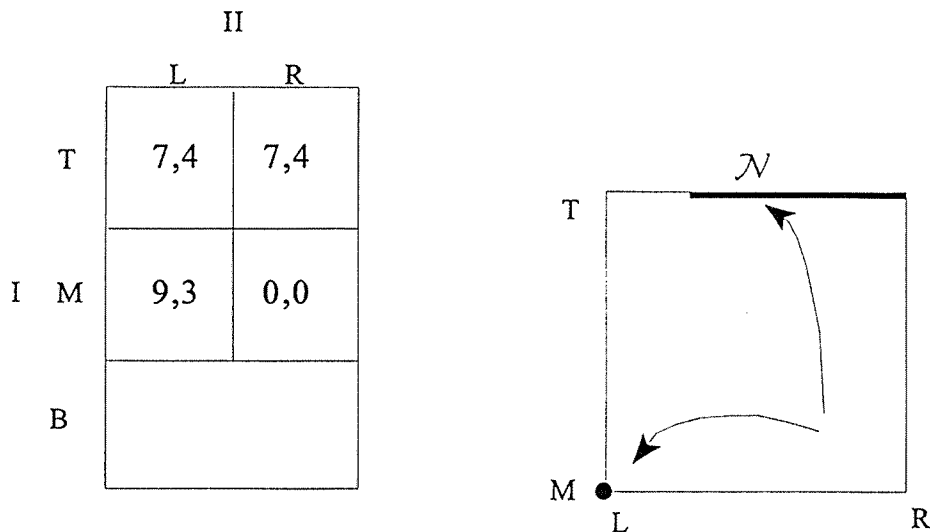


Figure 7: Dalek Game: undominated strategies

Nash equilibria that fail conventional forward induction arguments. However, we would like a condition that is easier to use than the requirement that “ $g(z^*)$  points into the interior of the basin of attraction of the set  $C = \{z \in N : z_R \geq z_R^*\}$ .” The key to finding such a condition is the observation that strategy  $B$  is strictly dominated by  $T$  for player  $I$  in Dalek. As a result, any monotonic dynamic will eliminate  $B$ . Hence, every orbit of the dynamics must approach the boundary face spanned by  $\{T, M\} \times \{L, R\} \equiv Z_{-B}$ . But on this face, as shown in Figure 7, the phase diagram is identical to that of the Chain-Store Game. Can we use what we know about the Chain-Store Game to identify sufficient conditions for stationary points near the component  $\mathcal{N}$  in the Dalek Game? In particular, suppose that condition (3.1) holds when the dynamics of the Dalek game are restricted to the face  $Z_{-B}$ . Can this fact be used to conclude that the Dalek Game has a stationary point close to  $\mathcal{N}$ ? The answer is yes, as long as drift does not create too powerful force toward strategy  $B$ .

If  $z$  is a point in the state space of this game, we let  $z_T$ ,  $z_B$ , and  $z_R$  denote the probabilities with which strategies  $T$ ,  $B$  and  $R$  are played at  $z$ , with the residual probabilities being attached to  $M$  and  $L$ . Similarly, we let  $f_T(z)$  be the element of the vector  $f$  corresponding to  $T$ , and so on. Then  $\mathcal{N} = \{z : z_T = 1, z_R \geq 2/9\} \subset Z_{-B}$  is the set of states corresponding to the

component of Nash equilibria that are not subgame perfect.

Let  $\hat{f}$ ,  $\hat{g}$ , and  $\hat{f} + \lambda\hat{g}$  be the dynamics  $f$ ,  $g$  and  $f + \lambda g$  on  $Z_{-B}$  defined by letting  $\hat{f} = f$  (this is possible because  $Z_{-B}$  is forward invariant under  $f$ ) and letting  $\hat{g}_T(z) = g_T(z)$  (and hence  $\hat{g}_M(z) = -\hat{g}_T(z)$ , making  $Z_{-B}$  forward invariant). We refer to  $\hat{f} + \lambda\hat{g}$  and the state space  $Z_{-B}$  as the *restricted dynamics*.

**Proposition 5** *Suppose that for all  $\delta > 0$ , there is a sufficiently small  $\lambda$  such that the restricted dynamics have a sink (saddle) [source] within  $\delta$  of  $\mathcal{N}$ . Then for all  $\delta > 0$ , there is a function  $k(\lambda) \mathbb{R} \rightarrow \mathbb{R}$  and a sufficiently small  $\lambda$  such that if  $\sup_{z \in Z} |g_B(z)| < k(\lambda)$ , then the unrestricted dynamics  $f + \lambda g$  have a sink (saddle) [saddle] within  $\delta$  of  $\mathcal{N}$  in the Dalek Game.*

To interpret this result, note that Proposition 3 provides sufficient conditions for the restricted dynamics to have stationary points close to the component of Nash equilibria  $\mathcal{N}$  in the state space  $Z_{-B}$ , as well as sufficient conditions for no such stationary point to exist. Proposition 5 indicates that if the former sufficient conditions are met, then we need look no further. As long as drift does not introduce the strategy  $B$  into the game with too much force, the unrestricted dynamics in the Dalek game also have a stationary point near  $\mathcal{N}$ .

**Proof** Let it be the case that for any  $\delta > 0$ , there is  $\lambda$  such that the restricted dynamics have a hyperbolic stationary state  $z(\delta)$  within  $\delta$  of  $\mathcal{N}$ . Then  $z(\delta)$  is also a stationary point of the dynamic  $f + \lambda\tilde{g}$  defined on  $Z$  by letting  $\tilde{g}_T(z) = g_T(z)$  and  $\tilde{g}_B(z) = 0$ . In addition, the Jacobian matrix of  $f + \lambda\tilde{g}$  at  $z(\delta)$  is given by:

$$\begin{bmatrix} \frac{\partial(f_T + \lambda g_T)}{\partial z_T} & \frac{\partial(f_T + \lambda g_T)}{\partial z_R} & \frac{\partial(f_T + \lambda g_T)}{\partial z_B} \\ \frac{\partial(f_R + \lambda g_R)}{\partial z_T} & \frac{\partial(f_R + \lambda g_R)}{\partial z_R} & \frac{\partial(f_R + \lambda g_R)}{\partial z_B} \\ \frac{\partial f_B}{\partial z_T} & \frac{\partial f_B}{\partial z_R} & \frac{\partial f_B}{\partial z_B} \end{bmatrix}.$$

Notice that, deleting the last row and column yields the Jacobian matrix of the restricted dynamics at  $z(\delta)$ . Because  $df_B(z)/dz_T = df_B(z)/z_R = 0$  when  $z_B = 0$  for a regular dynamic, the characteristic polynomial for the process  $f + \lambda\tilde{g}$  is given by  $(\partial f_B(z(\delta))/\partial z_B - \eta)\Lambda$ , where  $\eta$  is an eigenvalue and  $\Lambda$  is the characteristic polynomial of the restricted dynamics. Because  $z(\delta)_B = 0$ , we have  $\partial f_B(z(\delta))/\partial z_B < 0$ , where the inequality follows from monotonicity and the fact that  $z_T$  can be taken arbitrarily close to unity. The eigenvalues of the Jacobian of  $f + \lambda\tilde{g}$  thus consist of one negative eigenvalue and the

eigenvalues of the restricted dynamic at  $z(\delta)$ . If  $z(\delta)$  is a sink (saddle) [source] of the restricted process, then  $z(\delta)$  is a sink (saddle) [saddle] of the process  $f + \lambda\tilde{g}$ . The proof is now completed by noting that if  $\sup_{z \in Z} g_B(z) < k(\lambda)$  for some sufficiently small  $k(\lambda)$ , then  $f + \lambda g$  has a stationary point close to  $z(\delta)$  and of the same type as  $z(\delta)$  under the dynamic  $f + \lambda\tilde{g}$  (Hirsch and Smale [24, Theorems 1–2, p. 305]).  $\square$

### 5.3 Burning Money

We consider another common forward-induction example, the general form of which is due to van Damme [53, 54] and Ben Porath and Dekel [2]. The Battle of the Sexes game has two pure-strategy Nash equilibria and one mixed-strategy equilibrium. It is common to dismiss the mixed-strategy equilibrium. How do we choose between the two pure-strategy equilibria, given that the players have opposing preferences over these equilibria? Notice that all of these equilibria are hyperbolic stationary states, so that appealing to drift is no help in this game.

Suppose that before the game is played, player  $I$  has the option of burning two dollars, and let the payoffs in the Battle of the Sexes be such that the resulting game is shown in Figure 8. A strategy for player  $II$  in the Burning-Money Game identifies what player  $II$  will do if player  $I$  does and does not burn the two dollars. A strategy for player  $I$  indicates whether the money is burned and the subsequent choice of  $T$  or  $B$ . If the money is burned, then 2 must be subtracted from player  $I$ 's payoffs. The iterated elimination of weakly dominated strategies leads to a unique outcome for this game of  $(Not, T; LL)$  (the order of eliminations is shown in Figure 8), giving player  $I$  her preferred payoff.

Except among game theorists, this argument is typically regarded as preposterous. Its trivial first step is the observation that  $(Burn, B)$  is strictly dominated for player  $I$ , and will not be played. The remaining four steps are illustrated in Figure 9. We group these steps into two pairs, each of which corresponds to an argument about how the dynamics behave on a face of the state space where the phase diagram looks like that of the Chain-Store Game. The first two of these steps eliminate  $RR$  and  $LR$  for player  $II$  and  $(Not, B)$  for player  $I$ , and consists of the argument that  $(Burn, T; RL)$  will appear in game the game shown in Figure 9a. The second substantive part of the argument consists of the next two steps, which eliminate  $RL$  and  $(Burn, T)$ . These amount to the statement that  $(Not, T; LL)$  will be played in Figure 9b.

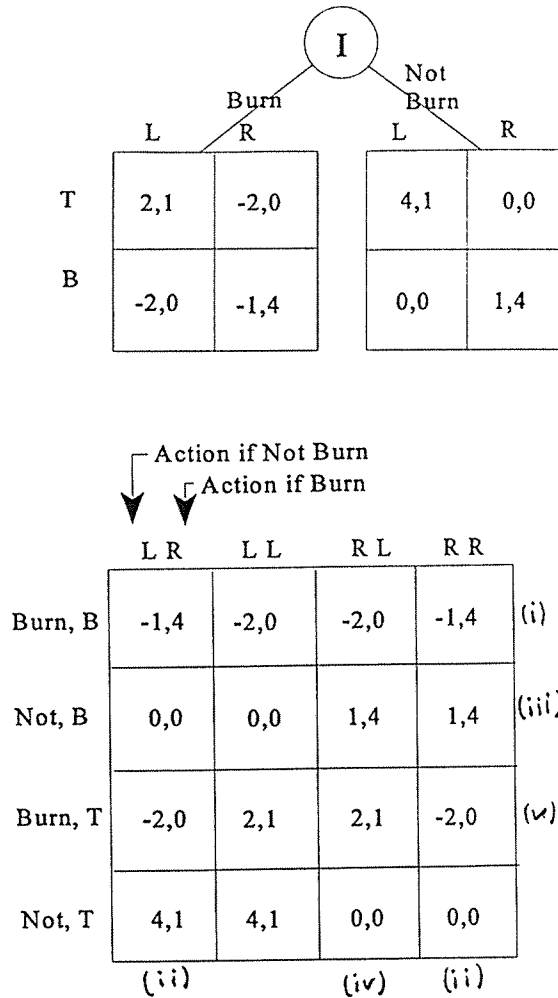


Figure 8: Burning-Money Game

With drift, the arguments that restrict attention to  $(Burn, T; RL)$  and  $(Not, T; LL)$  might fail. For each face, we can mimic the construction of the previous example to obtain a restricted dynamic  $\hat{f} + \lambda \hat{g}$ . Proposition 7 again supplies sufficient conditions for the restricted dynamics in each case to have a stationary state near the component  $\mathcal{N}$  of Nash equilibria in the restricted state space that are not subgame perfect. A proof analogous to that of Proposition 5 gives:

**Proposition 6** *Suppose that for all  $\delta > 0$ , there is a sufficiently small  $\lambda$  such that the restricted dynamics in the game of Figure 9a (9b) have a sink (saddle) [source] within  $\delta$  of  $\mathcal{N}$ . Then for all  $\delta > 0$ , there is a function  $k(\lambda)$  and a sufficiently small  $\lambda$  such that if  $\sup_{z \in Z} |g_i(z)| < k(\lambda)$ ,  $i \in \{Not, T; Burn, B; LL; LR\}$  ( $i \in \{Not, B; Burn, B; LR; RR\}$ ), then the unrestricted dynamics  $f + \lambda g$  have a sink (saddle) [saddle] in the Burning-Money Game within  $\delta$  of the component  $\mathcal{N}$ .*

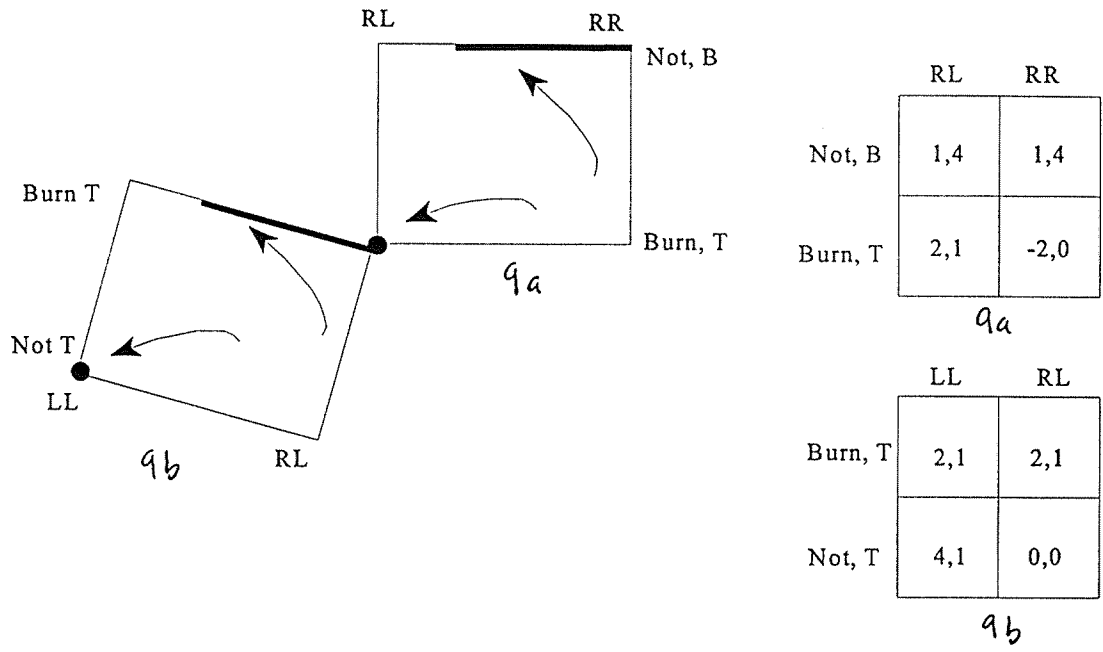


Figure 9: Strategy Eliminations

Hence, the dynamics  $f + \lambda g$  can yield an outcome in the Burning-Money Game in which player  $I$  does not burn the money and the outcome  $(B, R)$  appears in the Battle of the Sexes subgame, for payoffs  $(1, 4)$ , or an outcome in which player  $I$  does burn the money and  $(T, L)$  appears in the subgame for payoffs of  $(2, 1)$ . But this is precisely the intuition with which non-game-theorists greet this game—if, for whatever reason, an equilibrium  $(B, R)$  for payoffs  $(1, 4)$  has become established in the Battle of the Sexes, player  $I$ 's ability or threats to burn money are likely to be met with befuddlement or amusement by player  $II$ , and to have no effect on the outcome.

#### 5.4 Cheap Talk

One of the apparent successes of evolutionary game theory has been to use refinements of the evolutionarily stable strategy concept to examine issues of cheap talk (for example, Blume, Kim and Sobel [10], Kim and Sobel



		II	
		A	B
I	A	5, 5	0, 4
	B	4, 0	3, 3

Figure 10: Stag Hunt Game

[27], Matsui [32], Sobel [49], and Wärneryd [55]). Each paper establishes conditions under which the evolutionary process, operating on the cheap talk game, selects efficient equilibria of the underlying game. To see why such a result might be expected, consider an outcome in which everyone plays an inefficient equilibrium. Let a strategy appear in which some agents send the currently unused message  $\alpha$  and in which all agents play the efficient equilibrium whenever at least one agent sends message  $\alpha$ . The resulting dynamics will lead to an outcome with only the efficient equilibrium being played. Two steps are important in making this argument. The first is to establish that an unused message exists. The second is to ensure that agents react to this message by playing the efficient equilibrium. Both of these are essentially questions of drift.

To consider cheap talk, we begin with the Stag Hunt Game shown in Figure 10. This game has two Nash equilibria,  $(A, A)$  and  $(B, B)$ , with the former being-payoff dominant and the latter risk-dominant.

Now suppose that before playing the game, each player has an opportunity to announce either “A” or “B,” with the announcements being made simultaneously. We will interpret these as announcements of strategies that the agents claim they will play, but the announcements are “cheap talk” in the sense that they impose no restriction on the action that the player actually takes. A strategy is now an announcement and a specification of what the player will do for each possible announcement configuration. Hence, the strategy “ABA” is interpreted as “announce A, play B if the opponent announces A and play A if the opponent announces B.” The game with cheap talk is then given in Figure 11. We again think of this as a game played by a single population of players who are randomly chosen to be row or column players when matched.

The pure-strategy Nash equilibrium outcomes of the game are shown in boldface. These equilibria occur in two components of Nash equilibria, one yielding payoffs of  $(5, 5)$  (denoted by  $C_5$ ) and one yielding payoffs of  $(3, 3)$

	<i>AAA</i>	<i>AAB</i>	<i>BAA</i>	<i>BBA</i>	<i>ABA</i>	<i>BAB</i>	<i>ABB</i>	<i>BBB</i>
<i>AAA</i>	5, 5	5, 5	5, 5	0, 4	0, 4	5, 5	0, 4	0, 4
<i>AAB</i>	5, 5	5, 5	4, 0	3, 3	0, 4	4, 0	0, 4	3, 3
<i>BAA</i>	5, 5	0, 4	5, 5	5, 5	5, 5	0, 4	0, 4	0, 4
<i>BBA</i>	4, 0	3, 3	5, 5	5, 5	4, 0	0, 4	3, 3	0, 4
<i>ABA</i>	4, 0	4, 0	5, 5	0, 4	3, 3	5, 5	3, 3	0, 4
<i>BAB</i>	5, 5	0, 4	4, 0	4, 0	5, 5	3, 3	0, 4	3, 3
<i>ABB</i>	4, 0	4, 0	4, 0	3, 3	3, 3	4, 0	3, 3	3, 3
<i>BBB</i>	4, 0	3, 3	4, 0	4, 0	4, 0	3, 3	3, 3	3, 3

Figure 11: Cheap Talk Game

(denoted by  $C_3$ ). The component  $C_5$  is asymptotically stable. This is a reflection of the fact that 5 is the largest payoff available in the game.

The component  $C_3$  is not asymptotically stable and is not asymptotically stable with respect to the interior. For example, this component contains a state in which all agents play  $BBB$ , which we will refer to as “state  $BBB$ ,” as well as a state in which all agents play  $ABB$ , which we refer to as “state  $ABB$ .” If the system is in state  $BBB$ , with payoff (3, 3), then a slight perturbation that introduces strategy  $AAB$  leads to an outcome in which all agents play  $AAB$ , for a payoff of (5, 5). Notice that this transition reflects the cheap-talk intuition described above. We begin with a state in which all agents send message  $B$  and play  $B$ . The perturbation introduces agents who send message  $A$  and play  $A$  in any match in which both agents send  $A$ . This strategy earns a strictly higher payoff than  $BBB$ , since it always earns a payoff of at least 3 and sometimes earns 5.

Alternatively, if the system is in state  $ABB$ , then a slight perturbation that introduces strategy  $BBA$  yields a state in which all agents play  $BBA$  for a payoff of 5. In this case, agents are initially announcing an intention to play  $A$ , but always play  $B$ . The perturbation introduces a strategy in which players announce  $B$ , but play  $A$  whenever both players make such an announcement. The result is an outcome in which all players announce  $B$  but play  $A$ , for a payoff of 5. We see here that the evolutionary process attaches no importance to the nominal content of signals. Any signal can be used to prompt coordination on the good equilibrium. This is to be expected, as we could just as well have named our two messages “1” and “2.”

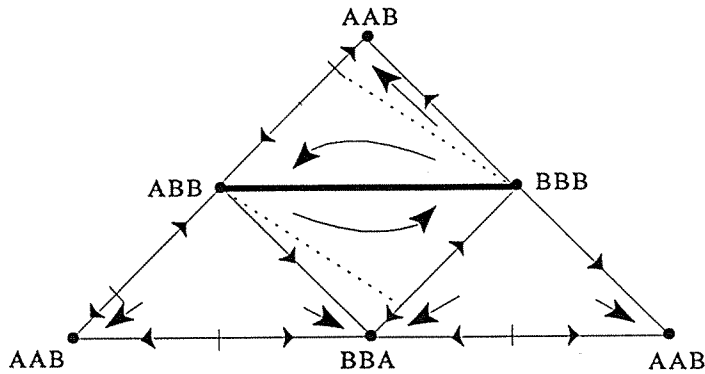


Figure 12: Partial phase diagram for Burning-Money Game

However, suppose that the current state divides agents in equal proportions between the strategies  $ABB$  and  $BBB$ . There is now no unused signal, and no small perturbation of the state can initiate learning dynamics that lead away from this state. We illustrate these considerations in Figure 12, which shows the phase diagram for four faces of the state space. The stability of component  $C_3$  thus hinges upon drift. If drift tends to push the system toward equal proportions of  $ABB$  and  $BBB$  whenever both strategies are in use, then the dynamics  $f + \lambda g$  may have a stationary state near the component  $C_3$ , and we cannot rule out inefficient payoffs.<sup>25</sup> This will be

<sup>25</sup>The second role for drift in cheap talk models is to ensure that if there is an unused signal that is seized upon by some agents as a device to coordinate on the good equilibrium, then other agents do not react to the signal by choosing some particularly disastrous action. In our Cheap Talk Game this is not a difficulty, because either of the actions an agent can take is part of some equilibrium.

the case if agents are indifferent about the signal they send, as long as they are going to play  $B$ . This is the dreaded “babbling equilibrium” of cheap talk games, and any efficiency result must have some way of excluding this equilibrium.

Suppose instead that drift introduces a tendency to play strategy  $BBB$  rather than  $ABB$ ,<sup>26</sup> so that drift pushes the system toward a state in which most agents play  $BBB$ . Suppose also that drift tends to introduce the strategy  $AAB$  in relatively large proportions. Because  $AAB$  is the strategy by which agents will be led from the Stag-Hunt equilibrium of  $(B, B)$  to  $(A, A)$ , these are the conditions most favorable to eliminating stationary points near  $C_3$ . If the composition of drift is such that these forces are sufficiently strong, then there may be no stationary point yielding payoff  $(3, 3)$ , no matter how small the level of the drift (i.e., no matter how small  $\lambda$ ). How strong must the bias toward  $BBB$  and  $AAB$  be in order to eliminate stationary states near  $C_3$ ? We have computed numerical solutions for the replicator dynamics with drift which suggest that the bias toward  $BBB$  and  $AAB$  must be very strong before a stationary point with payoffs near  $(3, 3)$  ceases to exist. In particular, a specification of drift that attaches probability  $\frac{2}{5}$  to strategies  $BBB$  and  $AAB$  and  $\frac{1}{30}$  to each of the remaining strategies does not suffice.

In light of this, we consider it premature to conclude that evolutionary processes select efficient outcomes in cheap talk games. This contrasts with much of the literature on evolutionary processes in cheap talk games, which has concentrated on efficiency results. Sobel [49], for example, derives conditions under which efficient equilibria are selected. To see where the differences arise, notice that Sobel rejects a component if there is *any* realization of the underlying stochastic drift process  $G$  that leads away from the component. This may be an appropriate notion for an ultralong-run analysis, since the ultralong-run is a period of time long enough that any realization of the process that *can* happen *will* happen. For a long-run analysis, however, there is no alternative to modelling the process of drift.

## 6 Comparative Statics and Testability

Our interest in drift was motivated by discrepancies between theoretical predictions and observed play. How does Proposition 2 help in analyzing how people play games? In particular, what are the potentially testable

---

<sup>26</sup>We could just as well assume that drift pushes agents toward  $ABB$ .

predictions that we could extract from Proposition 2?

**Comparative Statics.** We study predictions based on comparative statics exercises in which one or more observable parameters are varied while the remaining observable parameters are held fixed. If these predictions are violated in the laboratory, the theory on which they are based is refuted, though only conditionally so, since the predictions inescapably depend on maintained hypotheses about unobservables. It is therefore important that such maintained hypotheses not be overly strong.

We take the payoff function  $\pi$  of a game as our observable. This is determined by the parameters  $a, b, c, d, e,$  and  $k$  of Figure 1. These parameters will be subjected to the constraints

$$a > e > c \quad b > d. \quad (7)$$

The process  $\dot{z} = F(z, \pi) + \lambda G(z, \pi)$  will be treated as unobservable.<sup>27</sup> A maintained hypothesis about the learning process  $f$  derived as an approximation of  $F(z, \pi)$  will be that it is not only regular and monotonic, but also *comparatively monotonic*.<sup>28</sup> By the latter, we mean the following. Denote the  $i$ th strategy of player  $\ell$  by  $s_{\ell i}$  and let  $f_{\ell i}(z, \pi)$  be the coordinate of  $f$  corresponding to strategy  $i$  for player  $\ell$ . Let  $\bar{\pi}_{\ell i}(z)$  be the average payoff to strategy  $i$  for player  $\ell$  in state  $z$ . Now consider two payoff functions  $\pi : S \rightarrow \mathbb{R}^n$  and  $\pi' : S \rightarrow \mathbb{R}^n$  and fix a state  $z$ . Suppose there exists a strategy  $i \in S_h$  for player  $h$  such that  $\pi(s_{\ell j}, s_{-\ell}) = \pi'(s_{\ell j}, s_{-\ell})$  if  $j \neq i$  or  $\ell \neq h$  and such that  $\bar{\pi}_{hi} > \bar{\pi}'_{hi}$ . If  $z_{hi} > 0$ , then  $f_{hi}(z, \pi) \geq f_{hi}(z, \pi')$  and if  $z_{hj} > 0$  for  $j \neq i$ , then  $f_{hj}(z, \pi) \leq f_{hj}(z, \pi')$ , while  $f_{\ell i}$  for  $\ell \neq h$  is unaffected. This assumption ensures that if we fix a state  $z$  and then consider a change in the payoffs to player  $\ell$  of strategy  $i$  that increases the average payoff of strategy  $i$  in state  $z$ , then the rate at which strategy  $i$  grows increases, and the rate at which other strategies grow for player  $i$  decreases.

We know even less about the drift process  $g$ , derived as an approximation of  $G(z, \pi)$ , than about  $f$ . We have suggested that we expect drift to depend upon a host of factors in addition to the payoffs in the game, and will often expect drift to have very little to do with payoffs. In most of what follows,

<sup>27</sup>This is not to say that data cannot be gathered that is relevant to how people learn. The problem is that we do not know how to incorporate this data into the theory in a reliable manner.

<sup>28</sup>We now replace  $F(z)$  by  $F(z, \pi)$  to capture the dependence of the selection process on payoffs.

our maintained hypothesis is that  $g$  is independent of the payoffs in the game. However, we comment on cases where the predictions are less sensitive to this hypothesis than others. We refer to this as the case of *exogenous* drift.

It remains to discuss the initial condition  $z(0) = z(0, \pi)$ . Our comparative statics predictions are predicated on the assumption that  $z(0)$  is independent of payoffs but that we do not know in which basin of attraction of  $f + \lambda g$  the initial condition  $z(0)$  lies. Our predictions will therefore not apply games for which  $z(0)$  turns out to vary significantly with  $\pi$ .

**The Chain-Store Game.** These comparative-statics considerations suggest that it would be useful to run experiments that compare versions of the Chain-Store Game with varying payoffs. The first task is to determine which payoff configurations are more likely to give stable stationary points near the component  $\mathcal{N}$ .

**Proposition 7** *Fix payoffs  $a, b, c, d, e$  and  $k$  satisfying (7) for the Chain-Store Game of Figure 1. Let the selection dynamic  $f$  be monotonic, regular and comparatively monotonic and let the drift  $g$  be exogenous. If there exists a subset of the component  $\mathcal{N}$  of Nash equilibria satisfying conditions (a)–(b) of Proposition 2, then such a subset also exists for any larger values of  $e$  and  $d$  or smaller values  $a, b$  and  $c$ , that preserve (7). The converse can fail in each case.*

**Proof** First, we calculate  $\psi$  as the solution to  $a(1 - \psi) + c\psi = e$ , yielding  $\psi = (a - c)/(e - c)$ . Then  $\mathcal{N} = \{(r, 1) : \phi \leq r \leq 1\}$ . Furthermore, we can calculate  $d\psi/da > 0$ ,  $d\psi/dc > 0$  and  $d\psi/de > 0$ . Next, we notice that for a fixed state  $(h, n)$ , the average payoffs to strategies  $H, L, Y$ , and  $N$  are given by

$$\begin{aligned}\bar{\pi}_H &= e \\ \bar{\pi}_L &= a(1 - r) + rc \\ \bar{\pi}_Y &= fn + (1 - n)b \\ \bar{\pi}_N &= fn + (1 - n)d\end{aligned}$$

Now let  $a, b$ , or  $c$  decrease or  $d$  or  $e$  increase. Then  $\bar{\pi}_H$  and  $\bar{\pi}_N$  increase while  $\bar{\pi}_L$  and  $\bar{\pi}_Y$  decrease. By comparative monotonicity,  $f_n(z)/f_r(z)$  decreases for each  $z \in \mathcal{N}$ . In addition, our analysis of  $\psi$  shows that  $\mathcal{N}$  expands. For fixed drift, this ensures that if (5) held originally, then it continues to hold.  $\square$

Chain-Store experiments can thus be conducted with varying values of the payoffs  $a$  and  $b$ . An outcome consistent with the theory would be the observation of the subgame-perfect equilibrium for large values of  $a$  and  $b$  and the Nash, nonsubgame-perfect equilibrium for small values of  $a$  and  $b$ . Violations of this relationship would challenge the theory. Similar experiments can be done with other combinations of payoffs.

We have held drift to be exogenously fixed throughout this exercise. In many cases, this may be a suitable first approximation, as in biological examples. Binmore, Gale and Samuelson [4] argue that, in the Ultimatum Game, the drift process may be related to payoffs in the sense that a population may have a higher drift rate as the payoff differences between its strategies are smaller. The particular drift process examined in [4] reinforces the effects of movements in payoffs and Proposition 7 continues to hold for this drift process. An analysis like that leading to Proposition 7 can be performed for any process of drift, though the analysis will be more complicated, and the results will be clearer in some cases than in others, depending upon what is assumed about drift.

**The Dalek Game.** Similar comparative-statics considerations arise in connection with the Dalek Game. We could construct a general form of the Dalek Game, the first two rows of which would match the Chain Store Game of Figure 1 and the final row of which would correspond to a strictly dominated strategy for Player  $I$ . This suggests experiments in which the payoffs of the Dalek Game are manipulated. From Propositions 7 and 5, we again have that the experimental results are consistent with our model if we observe outcomes near the Nash equilibrium that is not subgame perfect for small values of  $a$ ,  $b$ , and  $c$  and large values of  $d$  and  $e$ . Violations of this pattern would again challenge the theory. One implication is that we should observe Nash equilibria that are not subgame perfect when the value to player  $II$  of taking the outside option is relatively large.

In light of this, it is interesting to note that Balkenborg [1], when conducting experiments with the Dalek Game, finds the outside option is virtually always chosen. This provides the beginnings of a comparative statics exercise, but additional insight into the model requires additional experiments with different payoffs. Toward this end, Binmore *et al* [5] examine a related game in which player  $II$  can either take an outside option or play the Nash Demand Game with player  $I$ . In the latter, the two players simultaneously make demands, splitting the difference if the demands are compatible and

receiving nothing otherwise. The implication of learning with drift is again that the outside option should be chosen when player  $II$ 's outside option is relatively lucrative. The experimental results of [5] show that the outside option is often taken, and is taken more often when it is more lucrative.

**The Burning-Money Game.** Now consider the Burning-Money Game. Once again, the theory suggests conducting experiments with varying pay-offs. In this case, the obvious payoff to vary is the amount of money to be burned. The forward induction argument leading to  $(Not, T; LL)$  holds as long as the amount to be burned lies in the interval  $(1, 3)$ . If stationary states near the Nash (but not subgame-perfect) equilibrium outcome  $(Burn, T; RL)$  exist, then they do so when the amount of money to be burned is relatively low. Alternatively, stationary states near the Nash equilibrium outcome  $(Not, B; RR)$  exist when the amount of money that is potentially burned is relatively high. Our suspicion is that the latter is the most likely possibility, a suspicion reinforced by the observation that this is the largest component of Nash equilibria that are not subgame perfect. Experimental outcomes will then be consistent with the theory if the money is not burned and player  $I$  has to settle for the outcome  $(B, R)$  in the original game whenever the amount of money to be burned is relatively high. Violations of this pattern once again pose a challenge to the theory.

## 7 Conclusion

The ideas behind this paper are simple: The criterion for a model to be successful is that it include important factors and exclude unimportant ones. But how do we know what is important and what is not? In the case of evolutionary games, the model itself provides the answers. If the model produces stationary states that are not hyperbolic and do not occur in components that satisfy some variation of asymptotic stability, then important factors have been excluded from the model and the latter should be expanded.

The factors to be added to the model are important, in the sense that they can have a significant impact on the behavior of the dynamic system, but they also may be arbitrarily small in magnitude. It is presumably because they are small that they are excluded from the model in the first analysis. How can a model whose behavior is shaped by arbitrarily small factors be of any use in applications? One conclusion of this paper is that, while the factors themselves may be small, their existence can nevertheless



be used to derive comparative-static results that do not depend upon observing arbitrarily small magnitudes. We are hopeful that these comparative static results can form the basis of an empirical analysis.

## References

- [1] Dieter Balkenborg. Tests of the forward induction hypothesis. Mimeo, University of Bonn, 1994.
- [2] E. Ben-Porath and E. Dekel. Coordination and the potential for self-sacrifice. Mimeo, Northwestern University, 1988.
- [3] N. P. Bhatia and G. P. Szegö. *Stability Theory of Dynamical Systems*. Springer-Verlag, New York, 1970.
- [4] Ken Binmore, John Gale, and Larry Samuelson. Learning to be imperfect: The ultimatum game. *Games and Economic Behavior*, 8:56–90, 1995.
- [5] Ken Binmore, Chrix. Proulx, Larry Samuelson, and Joe Swierzbsinski. Hard bargains and lost opportunities. SSRI Working Paper 9517, University of Wisconsin, 1995.
- [6] Ken Binmore and Larry Samuelson. An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics*, 150:45–63, 1993.
- [7] Ken Binmore and Larry Samuelson. Muddling through: Noisy equilibrium selection. SSRI working paper 9410, University of Wisconsin, 1993.
- [8] Ken Binmore and Larry Samuelson. Drift. *European Economic Review*, 38:859–867, 1994.
- [9] Ken Binmore, Larry Samuelson, and Richard Vaughan. Musical chairs: Modelling noisy evolution. *Games and Economic Behavior*, 11:1–35, 1995.
- [10] Andreas Blume, Yong-Gwan Kim, and Joel Sobel. Evolutionary stability in games of communication. *Games and Economic Behavior*, 5:547–575, 1993.

- [11] Gary E. Bolton. A comparative model of bargaining: Theory and evidence. *American Economic Review*, 81:1096–1136, 1991.
- [12] T. Börgers and R. Sarin. Learning through reinforcement and replicator dynamics. Mimeo, University College London, 1993.
- [13] Richard T. Boylan. Laws of large numbers for dynamical systems with randomly matched individuals. *Journal of Economic Theory*, 57:473–504, 1991.
- [14] Richard T. Boylan. Continuous approximation of dynamical systems with randomly matched individuals. *Journal of Economic Theory*, 66:615–625, 1995.
- [15] Antonio Cabrales. Stochastic replicator dynamics. Mimeo, University of California, San Diego, 1993.
- [16] Eddie Dekel and Drew Fudenberg. Rational behavior with payoff uncertainty. *Journal of Economic Theory*, 52:243–267, 1990.
- [17] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61:1047–1072, 1992.
- [18] Dean Foster and Peyton Young. Stochastic evolutionary game dynamics. *Journal of Theoretical Biology*, 38:219–232, 1990.
- [19] Drew Fudenberg and Chris Harris. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57:420–441, 1992.
- [20] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer Verlag, New York, 1983.
- [21] W. Güth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Behavior and Organization*, 3:367–388, 1982.
- [22] Werner Güth and Reinhard Tietz. Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, 11:417–49, 1990.
- [23] J. Hale. *Ordinary Differential Equations*. John Wiley and Sons, New York, 1969.

- [24] Morris W. Hirsch and Stephen Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, New York, 1974.
- [25] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge, 1988.
- [26] Michihiro Kandori and George J. Mailath and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56, 1993.
- [27] Yong-Gwan Kim and Joel Sobel. An evolutionary approach to pre-play communication. Mimeo, University of Iowa and University of California, San Diego, 1992.
- [28] Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, 1983.
- [29] Elon Kohlberg and Jean-Francois Mertens. On the strategic stability of equilibria. *Econometrica*, 54:1003–1038, 1986.
- [30] David M. Kreps and Robert J. Wilson. Sequential equilibrium. *Econometrica*, 50:863–894, 1982.
- [31] George Mailath, Larry Samuelson, and Jeroen Swinkels. Extensive form reasoning in normal games. *Econometrica*, 61:273–302, 1993.
- [32] Akihiko Matsui. Cheap-talk and cooperation in society. *Journal of Economic Theory*, 54:245–258, 1991.
- [33] Richard D. McKelvey and Thomas R. Palfrey. An experimental study of the centipede game. *Econometrica*, 60:803–836, 1992.
- [34] Paul Milgrom and John Roberts. Predation, reputation and entry deterrence. *Journal of Economic Theory*, 27:280–312, 1982.
- [35] Roger B. Myerson. Proper equilibria. *International Journal of Game Theory*, 7, 1978.
- [36] John H. Nachbar. ‘Evolutionary’ selection dynamics in games: convergence and limit properties. *International Journal of Game Theory*, 19:59–89, 1990.
- [37] Barry Nalebuff and Avinash Dixit. *Thinking Strategically*. W. W. Norton and Company, New York, 1991.

- [38] Jack Ochs and Alvin E. Roth. An experimental study of sequential bargaining. *American Economic Review*, 79:355–384, 1989.
- [39] Klaus Ritzberger. The theory of normal form games from the differentiable viewpoint. *International Journal of Game Theory*, 23:207–236, 1993.
- [40] Klaus Ritzberger and Jörgen Weibull. Evolutionary selection in normal form games. Stockholm University Working Paper 5, Institute for Advanced Studies, Vienna and Stockholm University, Sweden, 1993.
- [41] Alvin E. Roth. Bargaining experiments. In John Kagel and Alvin E. Roth, editors, *Handbook of Experimental Economics*, pages 253–348. Princeton University Press, 1995.
- [42] Alvin E. Roth and Ido Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212, 1995.
- [43] Larry Samuelson. Evolutionary foundations of solution concepts for finite, two-player, normal-form games. In Moshe Y. Vardi, editor, *Theoretical Aspects of Reasoning About Knowledge*. Morgan Kaufmann Publishers, Inc., 1988.
- [44] Larry Samuelson and Jianbo Zhang. Evolutionary stability in asymmetric games. *Journal of Economic Theory*, 57:363–391, 1992.
- [45] David A. Sánchez. *Ordinary Differential Equations and Stability Theory: An Introduction*. W. H. Freeman and Company, San Francisco, 1968.
- [46] Karl H. Schlag. Dynamic stability in the repeated prisoners' dilemma played by finite automata. Mimeo, University of Bonn, 1993.
- [47] R. Selten. The chain-store paradox. *Theory and Decision*, 9:127–159, 1978.
- [48] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive-form games. *International Journal of Game Theory*, 4:25–55, 1975.
- [49] Joel Sobel. Evolutionary stability in communication games. *Economics Letters*, 1993. Forthcoming.

- [50] Jeroen Swinkels. Adjustment dynamics and rational play in games. *Games and Economic Behavior*, 5:455-484, 1993.
- [51] Richard H. Thaler. *The Winner's Curse*. Princeton University Press, Princeton, 1992.
- [52] Eric van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory*, 13:1-13, 1984.
- [53] Eric van Damme. Stable equilibria and forward induction. SFB Discussion Paper A-128 128, University of Bonn, 1987.
- [54] Eric van Damme. Stable equilibria and forward induction. *Journal of Economic Theory*, 48:476-509, 1989.
- [55] Karl Wärneryd. Cheap talk, coordination, and evolutionary stability. Mimeo, Stockholm School of Economics, 1990.
- [56] Peyton Young. The evolution of conventions. *Econometrica*, 61:57-84, 1993.
- [57] Peyton Young and Dean Foster. Cooperation in the short and in the long run. *Games and Economic Behavior*, 3:145-156, 1990.



**Institut für Höhere Studien**  
**Institute for Advanced Studies**

Stumpergasse 56

A-1060 Vienna

Austria

Phone: +43-1-599 91-145

Fax: +43-1-599 91-163