JOURNAL OF ECONOMIC THEORY **77**, 34–57 (1997)
ARTICLE NO. ET972315

Non-Exclusive Conventions and Social Coordination*

Sanjeev Goyal

Econometric Institute, Erasmus University, 3000 DR Rotterdam, The Netherlands

and

Maarten C. W. Janssen

Department of Economics, Erasmus University, 3000 DR Rotterdam, The Netherlands

Received April 11, 1996; revised April 5, 1997

We study the long run outcome when communities with different conventions interact. We introduce the notion of non-exclusive conventions to model the idea that, by incurring some additional costs, agents can remain flexible and hence coordinate their activities more successfully. We show that if these costs of flexibility are low (high) and interaction is local then the Pareto-efficient (risk-dominant) convention prevails in both communities. At intermediate cost levels, the conventions coexist. We also show that the importance of relative size of the two communities varies across interaction structures. *Journal of Economic Literature* Classification Numbers: C7, D6, F15, Z1. © 1997 Academic Press

1. INTRODUCTION

In a wide variety of economic and social situations conventions play a central role. Some well known examples of conventions are languages, currencies, product standards and units of measurement. In our view, conventions have two distinctive features: First, a convention is an arbitrary

* We gratefully acknowledge the helpful comments of an associate editor and a referee. We thank Hans Carlsson, David Easley, Douglas Gale, Bob Galesloot, Bruno Jullien, Patrick Rey, Aldo Rustichini, Pauline Rutsaert, Avner Shaked, Eric Van Damme, Fernando Vega-Redondo, and audiences at Alicante, Core (Louvain), Duke, Indian Statistical Institute (New Delhi), Laboratoire Economic Industriel (Crest-Ensae, Paris), Oxford, McGill, Tilburg, University College (London), UNC-Chappel Hill, VPI, The 1995 World Conference of the Econometric Society, and The Conference on Individual Rationality and Social Norms at Tinbergen Institute, Rotterdam for useful comments. We also acknowledge the research assistance of Gerrit Draisma.

0022-0531/97 $25.00

solution to a social problem. It is arbitrary in the sense that a convention is typically only one among the many ways to solve the problem at hand. Second, the only reason individuals in a society choose to conform to a particular convention is that others conform to the same convention. These features suggest that a convention can be considered as a solution to a coordination problem or more formally as an equilibrium of a coordination game.[1]

We are aware that some of the well known solutions to these coordination problems—e.g., U.S. dollars or British pounds, Spanish or Japanese, 110 or 220 voltage, cheques or bank-transfers—have emerged while the different societies remained in relative isolation from each other. We are now witnessing an increase in the interaction between members of these societies and this raises the issue: Can these different conventions coexist or will one convention drive out the others, in the long run? In this paper we examine this question by characterizing the long-run outcomes of a dynamic process in which, at the start, individuals belonging to different communities conform to different conventions.[2]

Our model has two distinguishing features: *non-exclusive conventions* and *local interaction among individual agents*. We first introduce these concepts and then illustrate our results with the help of a simple example. The introduction ends with a discussion of the related literature.

A non-exclusive convention is an equilibrium of a coordination game in which players have the option of choosing actions consistent with several conventions (thereby ensuring coordination), at some additional cost. The idea of non-exclusive conventions is motivated by the observation that people who travel frequently between different societies often speak two or more languages, hold different currencies and carry several credit cards. Such a person remains "flexible" by incurring some additional costs, e.g., by subscribing to a second credit card. When these costs are (relatively speaking) very high the option of remaining flexible is not worthwhile, and conventions become *exclusive*.

We refer to interaction between individuals as *local* if individuals are more likely to meet people in their "neighborhood" than those "far away" from them. In contrast, interaction is said to be *uniform* if an individual is equally likely to meet anybody in the population. The idea of non-uniform interaction is quite general and important in a variety of social and economic settings. Consider, as an example, an individual choosing her telephone network: in this decision, it is more important to check on which

[1] The study of conventions (and more generally, of coordination problems) was initiated by Lewis [12] and Schelling [15].

[2] Matsuyama *et al.* [14] study a model on the evolution of an international medium of exchange, which is similarly motivated.

network her friends and colleagues are, rather than to look at the general size of the different networks.

Our analysis focuses on the conflict between risk-dominance and Pareto-efficiency considerations and our results highlight the role of the cost of remaining flexible and the interaction structure in determining the long-run outcomes. For local interaction, our results may be summarized as follows: if the costs are low, the Pareto-efficient convention drives out the risk-dominant convention while the reverse is true when these costs are high. Moreover, if these costs are at an "intermediate" level then the two conventions can coexist in the long run. Our analysis of the uniform interaction case brings out two points. One, that the Pareto-efficient equilibrium is more likely for low cost levels as compared to high cost levels and two, that the initial configuration plays a relatively more important role under uniform interaction as compared to local interaction. We now illustrate the intuition behind these results in the context of a numerical example.

EXAMPLE. Consider a research field with 100 active researchers. Assume, to fix ideas, that these researchers are located on the integer points along a line. Each individual has a choice between two electronic networks: α and β. To be able to use a network, an individual must pay an annual subscription fee of c_1 in the course of a year. A researcher has the option of subscribing to both networks at a higher cost $c_2 > c_1$. For our analysis, it is only the difference in these costs that matters, and we consider $c = c_2 - c_1$.

We start the analysis in a situation where there are two communities, an α community (say, of game theorists) in which everyone subscribes to the α network and a β community (say, of evolutionary biologists) in which everyone subscribes to network β. There are 10 researchers in the first community and 90 in the second one. Researchers in the α group are located at points from 1 to 10 and those in the β group are located at points from 11 to 100. It is assumed that this group configuration evolved while the two groups were isolated, and we examine what happens when the two groups start interacting (due to the emergence of evolutionary game theory).

Suppose that network α is superior, i.e. everybody would (in principle) prefer to subscribe to α. However, network α performs badly in communicating with network β. The reverse holds true for network β: it is not as good as network α, but it communicates better with α than α does with it. Moreover, the quality of interaction is better when two researchers subscribe to the same network than if they subscribe to different networks. These considerations are summarized in the payoff matrix below.

$$\begin{array}{cc} & \begin{array}{cc} \alpha & \beta \end{array} \\ \begin{array}{c} \alpha \\ \beta \end{array} & \begin{pmatrix} 10,10 & 4,7 \\ 7,4 & 8,8 \end{pmatrix} \end{array}$$

It follows that network α Pareto-dominates network β. On the other hand, since $10-7<8-4$, network β risk-dominates network α in the sense of Harsanyi and Selten [10]. In addition, we assume that a researcher who subscribes to both networks receives a payoff of $10-c$ when she meets another researcher who subscribes to α or to both networks and $8-c$ when she meets a researcher subscribing to β only. The possibility of subscribing to both networks formalizes the idea of non-exclusive conventions.

We will look at two types of interaction among researchers: local and uniform. With local interaction, each individual has a probability of 1/4 that she will be matched with any of the 2 neighbors on either side of her.[3] In the uniform interaction model, the probability of being matched with anyone else in the population is 1/99.

We study the evolution of the individual choices over time. These decisions are based on simple rules which may be summarized by saying that individuals choose an action that yields them the highest expected payoff conditional on the belief that the current period distribution of choices will remain the same as in the last period. In our analysis, the social distribution of actions can be summarized by the location of the boundaries between the α and β regions. We study how the movement of these boundaries depends on the value of c.

We begin with an intermediate case in which $c=1$. In period 1, individuals expect the boundaries to remain the same as in the initial configuration. Given these expectations, simple calculations reveal that in period 1, the social configuration is that researchers 10 and 11 subscribe to both α and β, researchers at points 1–9 persist with subscription to α while 12–100 subscribe to β only. To see why this is the case consider researcher 10: Her expected payoff from subscribing to both α and β is 8 which is greater than 7 and 7.5, the expected payoff from subscribing to α or β alone. In period 2, researchers expect the boundary between the α region and the dual subscription region to be between individuals 9 and 10 and the boundary between the dual subscription region and the β region to be between individuals 11 and 12. Given these expectations, it is easily calculated that in period 2, it is optimal for researchers 1–9 to subscribe to α only, for 10 and 11 to subscribe to both networks, and 12–100 to subscribe to β only, i.e., nobody has an incentive to change their subscription. This implies that the configuration of period 2 can persist and the two conventions will coexist in the long run. This example illustrates the general point that Theorem 1 makes: if the value of c is "intermediate" and the initial size of the two communities is "large" relative to the domain of

[3] Researchers located at the end points are not important for our argument and we shall ignore potential end-point problems; the general model does not have this problem since we assume people are located around a circle.

interaction then Pareto-ranked and risk-ranked conventions can coexist in the long run.

We next consider the case where the cost of dual-subscription is low; suppose $c < 0.5$. In period 1, for researcher 9 the expected payoffs from the different actions are: 8.5 from α, 7.25 from β and $9.5 - c > 8.5$ from dual subscription. Thus, in addition to researcher 10 and 11, she too will switch to dual subscription in period 1. Analogous computations show that researcher 12 will also switch to dual subscription. Thus, in period 1, researchers 1-8 subscribe to α, 9–12 opt for dual subscription, and 13–100 subscribe to β only. Taking this configuration as the expectation for period 2, it is easy to see that it is optimal for 9 and 10 to switch back to subscribing to α only, while researchers 11–14 will opt for dual subscription. In period 3, the boundary between the conventions will shift further and 1–12 will subscribe only to α, 13–16 will subscribe to both networks, and 17–100 subscribe to β and so on. This process will continue and the β network will be eliminated. This example illustrates the first point point made by Theorem 2: if the costs of dual subscription are "small" then the risk-dominant convention will be eliminated by a Pareto-efficient convention, in the long run.

Let us finally consider the case when costs of dual subscription are large. For illustration consider $c > 2$. In this case it is easy to see that no one will subscribe to both networks, and we are back within the framework of exclusive conventions. It is now possible to show using calculations analogous to the previous cases that over time the boundaries will move towards 1 and the risk-dominant network will drive out the Pareto-efficient network in the long run. This example illustrates the second point made by Theorem 2.

One way to interpret the option of subscribing to both networks is as an insurance against being matched with a researcher who subscribes to the other network. The value of c can then be considered as an insurance premium. The idea behind the two Theorems on local interaction is then the following. If the insurance premium is relatively low, individuals at the boundary buy the insurance and risk-dominance considerations are not important. The reverse is true for a high insurance premium. At intermediate values of the insurance premium, risk-dominance and Pareto-dominance considerations are equally important and coexistence is dynamically stable.

To clarify the role of the interaction structure in determining long-run outcomes, we also consider the dynamic process under uniform interaction. Given the initial 10–90 split between the two communities, and some c, the expected payoffs of subscribing to α, β or both networks in period 1 is approximately equal to 4.6, 7.9 and $8.2 - c$, respectively. If $c > 0.3$ then the best response is for every one to subscribe to β only, while if $c < 0.3$ then the best response is for every one to subscribe to (α, β) only and thus eventually to α. In other words, the process converges, with everyone conforming to convention α or β. This example also suggests that the Pareto-efficient equilibrium is "more likely" when c is smaller, in the sense that the basin of attraction for the Pareto-efficient equilibrium is larger for smaller values of c.

A second point we wish to make concerns the relative importance of the initial size under local and uniform interaction, respectively. To illustrate this suppose that the initial configuration is different from before and that there are 90 researchers on the α network while there are 10 researchers on the β network. It is easily seen that under local interaction, the long-run outcomes would be identical to those obtained under the (alternative) initial configuration considered above. The outcomes under uniform interaction are greatly altered, however: Simple calculations about the expected payoffs reveal that the Pareto-efficient convention will obtain for all values of c. Thus the initial size of the different networks plays a more prominent role in determining long-run outcomes under uniform interaction as compared to local interaction.

The main purpose of the paper is to show that the arguments concerning the role of the cost, c, and the nature of the interaction structure are quite general and hold for all 2×2 games of interest.

Related literature. This paper is a contribution to the study of the evolution and stability of conventions. Some of the recent work in evolutionary game theory, see e.g., An and Kiefer [1], Anderlini and Ianni [2], Blume [4], Boyer and Orlean [5], Ellison [7], Kandori *et al.* [11], Sugden [16] and Young [17], has similar concerns.[4] The present paper is novel in the following respects. One, our model allows for non-exclusive conventions; by contrast, existing work has focussed on the case of exclusive conventions. Two, our model is deterministic; accordingly, the dynamics in our model are different from those studied in the recent evolutionary game theory literature. Three, the present paper focuses on the long-run implications of interaction between two communities with different conventions. This allows us to characterize the dynamic behaviour of the system in terms of a set of difference equations that describe the movement of the boundary between two conventions.[5]

[4] For a survey of this literature, see Young [18].

[5] Anderlini and Ianni [2] consider a finite agent model of local interaction and study the coexistence of exclusive conventions, among other things. The interaction model they consider is more general; however, to ensure convergence, they make use of "noise on the margin". By contrast, in our model, the dynamics are deterministic. The assumption of two spatially distinct regions conforming to different conventions allows us to obtain convergence without introducing noise in the process.

The plan of the rest of the paper is as follows. Section 2 presents the basic model. Section 3 contains the principal results pertaining to local interaction, while Section 4 discusses uniform interaction. Section 5 contains some concluding remarks. The proofs are given in an appendix.

2. THE MODEL

We consider an infinite period model where time is indexed by $t = 0, 1, 2, 3, \ldots$. In each period a unit measure of individuals, uniformly distributed over the circumference of a circle is (randomly) matched pairwise to play a 2 person normal form game.[6] We denote the population of individuals by N and an individual agent by x.

A central feature of our model is the idea that at some cost an individual can become adept at playing many actions. Players decide to invest in (or learn) an action $a_i \in A$ where $A = \{\alpha, \beta, (\alpha, \beta)\}$, prior to being matched with other players. Investing in action α or β is equally costly but investing in the dual action (α, β) costs more. Let c denote the difference between these two costs. We assume that $c > 0$. Once an individual makes the choice of which action to invest in, he or she is identified as being of that type. The set of types is also denoted by $A = \{\alpha, \beta, (\alpha, \beta)\}$.[7]

Once matched, players are only allowed to choose an action that they have already invested in. They coordinate on some action if it is feasible, and if they are both of type (α, β) then they choose to coordinate on the action combination with the highest payoff, i.e. they can solve the coordination problem. Implicit in this formulation is the idea that types can be identified prior to actual play of the game.[8] This assumption seems

[6] The assumption of a unit measure of agents is made to ensure that agents are small; the geometry of the circle is exploited in the definition of the neighborhoods of interaction (see A.2 below). Specifically, it allows us to formulate the neighbourhood of every individual in a natural "symmetric" way; this is not possible if players are located on a closed interval along a line.

[7] Allowing players to play mixed strategies does not change our results and adds to the complexity of the notation. This is because only the individuals on the boundaries between conventions are indifferent between different actions and with a continuum of agents the action of single isolated agents has no impact on the payoffs of the other agents. The evolution of conventions would be essentially the same as reported in Theorems 1–3.

[8] In some contexts other coordination assumptions may, however, be more natural. For example, when two individuals who know many languages and who are from the same country meet, then they typically converse in their native language. This practice can be modelled as follows: When two (α, β) types, belonging to the same region in the initial configuration, are matched then they coordinate on the action that they were coordinating on at the start (i.e., at $t = 0$). The dynamic process under this alternative assumption can be analyzed with the help of difference equations analogous to those used in Theorems 1–3 below.

reasonable in some settings where communication concerning types is natural, as is the case, for example, in the use of word processing packages, a language or a credit card.

These considerations concerning actions and payoffs are summarized in the following matrix representation of the 2-person normal-form game (where $a \geqslant b$):

$$\begin{array}{cc} & \begin{array}{ccc} \alpha & \beta & (\alpha, \beta) \end{array} \\ \begin{array}{c} \alpha \\ \beta \\ (\alpha, \beta) \end{array} & \begin{pmatrix} a, a & e, d & a, a-c \\ d, e & b, b & b, b-c \\ a-c, a & b-c, b & a-c, a-c \end{pmatrix} \end{array}$$

We are interested in problems of coordination and more specifically in the interaction between Pareto-efficiency and risk-dominance. This motivates the following restrictions on the payoffs:

$$a > d,\ b > e;\ a \geqslant b,\ b > d,\ b-e \geqslant a-d, \qquad \text{and} \qquad c > 0. \tag{A.1}$$

Since $a > d$ and $b > e$, there are two pure strategy equilibria of this game, (α, α) and (β, β). We shall refer to them as the α and β convention, respectively. Convention α is said to Pareto-dominate convention β if $a > b$ and convention β is said to risk-dominate convention α if $b-e > a-d$.[9]

We shall also discuss games in which risk dominance and Pareto-efficiency considerations are not in conflict. For this class of games we have the following restrictions on the payoffs:

$$a > d,\ b > e;\ a \geqslant b,\ b > d,\ a-d > b-e, \qquad \text{and} \qquad c > 0. \tag{A.1$'$}$$

The second important feature of our model is the structure of interaction among individual agents. We assume that an individual x only meets individuals drawn randomly from some neighbourhood $N(x) \subseteq N$, of where she is located. We define this neighbourhood in terms of the arc length $\delta \in (0, 1/2]$ on either side of the point x. We shall use the notation $|x, y|$ to denote the arc length (measured clockwise) between any two points x and y on the circle. This formulation is adapted from the literature on discrete choice models (Anderson *et al.*, [3]). Formally, for any $x \in N$, the probability of being matched with some $y \in N$ is given by a density function $f_x(y)$, where

$$f_x(y) = \begin{cases} 1/(2\delta) & \text{if} \quad y \in N(x) \\ 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

[9] We assume that $b > d$ to be able to concentrate on the cases in which coordination on some action is always better than failing to coordinate.

For $\delta = 1/2$, the above assumption on interaction yields uniform interaction, i.e., $N(x) = N$ and the probability of being matched with an agent is independent of her location. For $\delta < 1/2$, on the other hand, the probability of interaction depends on the location, and interaction is "local". This is a simple and rather special model of local interaction; we expect that the proofs of the main results, which exploit the symmetry of f, can be generalized relatively easily.[10]

Denote the social configuration of individual choices by $s\colon N \to A$. Given a configuration, s, the expected payoff for individual x from action a_i is denoted by $\pi_x(a_i \mid s)$. For the case of uniform interaction, it depends solely on the aggregate proportion of agents choosing the different actions. So, for example, the expected payoff from action α is expressed as: $\pi_x(\alpha \mid s) = p^\alpha a + p^\beta e + p^{(\alpha,\beta)} a$, where p^α, p^β and $p^{(\alpha,\beta)}$ denote the aggregate proportion of the population choosing α, β and (α, β), respectively. We also define $p = (p^\alpha, p^\beta, p^{(\alpha,\beta)}) \in \Delta$, where Δ is the unit simplex in R^3.

In case of local interaction, the expected payoff from an action also depends on the precise spatial distribution of these proportions. For the cases of interest, the regions where different actions are played are measurable, and, therefore, the expected payoffs are well-defined. As an illustration, consider the expected payoff to an agent x located on the boundary between two large α and β regions. The expected payoffs for action α are given by, $\pi_x(\alpha \mid s) = (\delta a)/(2\delta) + (\delta e)/(2\delta) = (a+e)/2$. We now define a social equilibrium of this game.

Social equilibrium. A social equilibrium is a configuration of individual choices $\hat{s}$ such that

$$\pi_x(\hat{s}(x) \mid \hat{s}) \geqslant \pi_x(a_i \mid \hat{s}), \ \forall a_i \in A, \ \forall x \in N.$$

For a fixed parameter vector $r = (a, b, c, d, e, \delta)$, let the set of social equilibria be denoted by Q_r. Social equilibria in which both α and β are chosen are referred to as coexistence social equilibria and the set of such equilibria is denoted by Q_r^{cx}.

Dynamics. We are interested in studying the interaction of societies which conform to different conventions. This motivates our focus on initial configurations, s_0, in which there exists one α-region and one β-region, denoted by Γ_0^α and Γ_0^β, respectively, such that $\Gamma_0^\alpha \cup \Gamma_0^\beta = N$.[11] The class of

[10] For example, our results extend easily to the case where f is a triangular function. See Goyal [9] for a discussion on other interaction models.

[11] Our main results continue to hold if we allow for regions of (α, β) between (relatively) large regions of α and β.

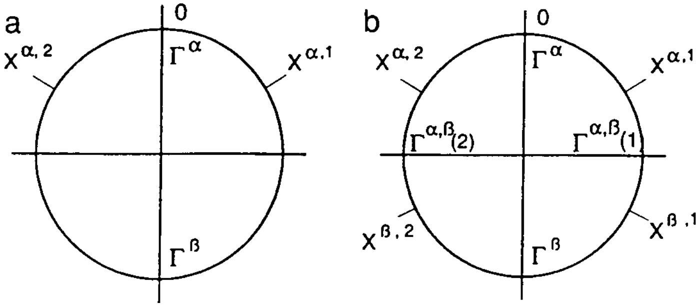

FIGURE 1

such s_0 can be parameterized by the size of the α region, $|\Gamma^{\alpha}|$. As $\Gamma_0^{\alpha} \cup \Gamma_0^{\beta} = N$, we have $x_0^{\alpha,1} = x_0^{\beta,1}$ and $x_0^{\alpha,2} = x_0^{\beta,2}$; thus we can, with some abuse in notation, say that $s_0 \in \mathrm{I} = [0, 1]$. The boundaries of the α and β regions in period t are denoted by $x_t^{\alpha,1}$ and $x_t^{\alpha,2}$, and $x_t^{\beta,1}$ and $x_t^{\beta,2}$, respectively. We shall, without loss of generality, assume that, at $t = 0$, the α region is centered around 0 which is the twelve o'clock point, i.e., $|x_0^{\alpha,2}, 0| = |0, x_0^{\alpha,1}|$.[12] An example of such a configuration is given in Fig. 1(a).

Our analysis will, on occasion, deal with configurations depicted in Fig. 1(b) and we refer to the two (α, β) regions as $\Gamma^{(\alpha,\beta)}(1) = [x^{\alpha,1}, x^{\beta,1}]$ and $\Gamma^{(\alpha,\beta)}(2) = [x^{\beta,2}, x^{\alpha,2}]$. The vector of boundary points is denoted by x^B.

In every period, individuals choose an action that is a (myopic) best response to their expectations concerning the choices of other individuals.[13] There are different ways in which these expectations may be formed. Individuals may, for instance, access market data from the previous year and use this to forecast the current period's pattern of subscription. In our analysis, we use a simple formulation of expectations formation which captures this idea: we posit that individuals expect the current period's

[12] In what follows, we shall use a plus sign (+) to denote a clockwise movement and a minus sign (−) to denote a anti-clockwise movement. So for instance, $x = y - \delta$ means that x is located at an arc distance of δ from the point y, where the distance is calculated anti-clockwise.

[13] One motivation for these decision rules derives from the following type of economic situations: In the course of a year, individual agents have to decide which credit cards or telephone company they will subscribe to. The attractiveness of, say, the different credit cards depends on the number of other individual agents who use a similar credit card. The statistic is, however, not yet available (since it is an outcome of decisions in the current period) and so individuals form some expectations about it.

configuration to be the same as in the previous period, i.e.,[14]

$$E(s_t \mid s_{t-1}) = s_{t-1}, \quad \forall t = 1, 2, 3, ... \tag{E}$$

Given these expectations, individuals choose an action $a_i \in A$, that gives them the highest expected payoff. Formally, for any t,

$$s_t(x) \in \arg\max_{a_i \in A} \pi_x(a_i \mid E(s_t)) \tag{B}$$

Thus given s_{t-1}, (E) and (B) define s_t. We study the long-run behaviour of s_t and our interest is in the dynamic stability of different types of social equilibria.[15]

A set of social equilibria is said to be dynamically stable if there exists a non-trivial basin of attraction for it. So, for instance, we say that Q_r^{cx} is *dynamically stable* if there exists an interval $I_r \subseteq I$ such that if $s_0 \in I_r$ then $s_t \to \hat{s} \in Q_r^{cx}$.[16]

3. LOCAL INTERACTION

Our main results, Theorems 1–3, consider the nature of long-run outcomes under local interaction. We show that if risk-dominance and Pareto-efficiency considerations conflict then the value of c is crucial in determining the long-run convention. By contrast, if the Pareto-efficient convention is also risk-dominant then, for all values of c, only the Pareto-efficient convention survives.

[14] Several other recent papers have used adaptive expectations to study long-run social learning; for a discussion on some of the issues involved see Mailath [13], Kandori *et al.* [11]. Our main results, Theorems 1–3, are robust to some alternative specifications of the dynamics which involve n-period adaptive expections and infrequent individual choices. Under local interaction, players only need to have expectations concerning choices of agents in their neighborhoods. Thus, local information is sufficient to define the expectations in (E). The present formulation has been adopted due to its expositional simplicity.

[15] Thus, in principle, different individuals could be making their decisions at different points within an interval of time, e.g. renewing their subscription for a credit card at different dates in the same calender year. The crucial assumption is that all individuals must use the expectations as defined in (E). This is reasonable since market statistics concerning performance of different products are only released at regular intervals which may be interpreted as coinciding with our time periods.

[16] Recall that Q_r^{cx} refers to the set of coexistence equilibria. This requirement of dynamic stability is weaker than the more familiar requirement that every $\hat{s} \in Q_r^{cx}$ has a basin of attraction. In our setting, this weaker requirement is suitable, since there are a continuum of coexistence equilibria and we are primarily interested in the issue of whether coexistence is stable.

We start by characterizing a class of coexistence social equilibria, in which interaction between regions conforming to different conventions is mediated by a set of individuals who invest in both actions (as in Fig. 1(b)). The payoffs from different actions are continuous in x. Hence, in equilibrium, for agent $x^{\alpha,1}$, the costs of investing in both actions must be exactly offset by the benefits of doing so. In other words,

$$c = \frac{1}{2\delta}(x^{\beta,1} - x^{\alpha,1})(b-e).$$

This implies that $|x^{\alpha,1} - x^{\beta,1}| = \delta(1 - 2c/(b-e))$. Likewise, for agent $x^{\beta,1}$ it must be true that

$$c = \frac{1}{2\delta}[(x^{\beta,1} - x^{\alpha,1})(a-b) + (x^{\alpha,1} - x^{\beta,1} + \delta)(a-d)].$$

This implies that $|x^{\alpha,1} - x^{\beta,1}| = \delta(a-d-2c)/(b-d)$. Putting these restrictions together yields the following condition for coexistence: $2c(d-e) = (a-b)(b-e)$. We can now state:

Proposition 1.[17] *Suppose* A.1–A.2 *hold,* $\delta < 1/2$ *and* $2c(d-e) = (a-b)(b-e)$. *Any configuration* $\hat{s}$ *in which* x^B *satisfy* (i)–(ii) *is a coexistence social equilibrium*:

(i) $|x^{\alpha,1}, x^{\beta,1}| = |x^{\beta,2}, x^{\alpha,2}| = \delta(1 - (2c/(b-e)))$

(ii) $|x^{\alpha,2}, x^{\alpha,1}| \geqslant \delta$; $|x^{\beta,1}, x^{\beta,2}| \geqslant \delta$.

Some remarks on the nature of coexistence follow. First, we observe that coexistence is only possible for non-generic values of c. It is worth emphasizing that this is an artifact of the assumption of a continuum of agents. The robustness of coexistence is related to the degree of overlap of neighborhoods of boundary individuals: specifically, the more similar these neighborhoods, the smaller the potential range of values of c for which coexistence is feasible. In the continuum model, the neighborhoods of boundary individuals are (in a measure-theoretic sense) identical and this generates the non-genericity observed in Proposition 1. In finite agent models neighborhoods will typically be distinct and coexistence more robust.[18] Second, note that $b-e > a-d$ is consistent with $2c(d-e) = (a-b)(b-e)$, implying that risk-ranked conventions can coexist in equilibrium. Finally, note that since the boundaries of the regions are such that they equate the benefits from dual subscription with the costs of doing so,

[17] A proof of this proposition is available on request from the authors.

[18] Appendix B discusses this issue in greater detail.

an increase in c must be compensated by a greater reward from dual subscription, i.e., a greater likelihood of meeting individuals who can do only β or only α. Thus an increase in c implies a decrease in the sizes of $\Gamma^{(\alpha,\beta)}(1)$ and $\Gamma^{(\alpha,\beta)}(2)$.

Theorem 1 shows that if $2c(d-e)=(a-b)(b-e)$ then, under local interaction, the set of coexistence social equilibria Q_r^{cx} is dynamically stable.[19]

THEOREM 1. *Suppose that* A.1–A.2 *hold*, $2c(d-e)=(a-b)(b-e)$ *and* $a>b$.

(i) *If* $\delta \leqslant 1/6$ *then* Q_r^{cx} *is dynamically stable.*

(ii) *If* $\delta \in (1/6, 1/2)$ *and* $c \geqslant (b-e)(3/4-1/(8\delta))$ *then* Q_r^{cx} *is dynamically stable.*

The proofs of Theorems 1–3 exploit the fact that the dynamic process $\{s_t\}$ can be studied in terms of the movement of boundaries $\{x_t^B\}$. This is useful since the movement of boundaries is captured by a set of simple difference equations so long as the α and β regions are "large".[20] The proof of Theorem 1 then uses two arguments, both of which exploit the hypothesis that $2c(d-e)=(a-b)(b-e)$. The first argument shows that if the regions are large initially then they will remain large over time, while the second argument shows that the solution to the difference equations is a converging sequence of boundaries and the limit is a coexistence equilibrium as characterized in Proposition 1.

We now turn to an examination of the case where $2c(d-e)\neq(a-b)(b-e)$. The first interesting result in this case is that if the cost c is "low" then only the Pareto-efficient convention survives, in the long run. By contrast, for high values of c, the risk-dominant convention prevails.[21] Let I_r denote the interval of initial configurations in which $|\Gamma_0^{\alpha}| > \delta + 2\delta(1-(2c)/(b-e))$ and $|\Gamma_0^{\alpha}| > \delta + 2\delta(1-(2c)/(b-e))$.

THEOREM 2. *Suppose that* A.1–A.2. *hold*, $\delta < 1/2$ *and* $s_0 \in \mathrm{I}_r$.

(i) *If* $2c(d-e)<(a-b)(b-e)$ *then all players conform to the Pareto-efficient convention after a finite period.*[22]

[19] An analysis of the pure coordination case in which $a=b$ and $d=e$ is available from the authors upon request.

[20] In particular, the difference equations can be used if $|\Gamma_0^{\alpha}|$ and $|\Gamma_0^{\beta}|$ are both bigger than $\delta+2\delta(1-(2c)/(b-e))$. This is feasible if $2[\delta+2\delta(1-(2c)/(b-e))]\leqslant 1$, a requirement that is satisfied for all $c>0$, if $\delta\leqslant 1/6$. If $\delta>1/6$, then it is satisfied so long as $c>(b-e)(3/4-1/(8\delta))$. These considerations explain the restrictions on the value of c.

[21] This conclusion is similar to Theorem 1 in Ellison [7].

[22] Note that given (A.1), $2c(d-e)<(a-b)(b-e)$ is satisfied only if $a>b$. Thus α is the Pareto-efficient convention in this case.

(ii) *If* $2c(d-e)>(a-b)(b-e)$ *then all players conform to the risk-dominant convention after a finite period.*

The proof of Theorem 2 exploits the same set of difference equations for the boundaries between regions as Theorem 1. In part (i), the boundaries $x_t^{\beta,1}$ and $x_t^{\beta,2}$ are, roughly speaking, strictly increasing and strictly decreasing, respectively. Thus the size of the β region shrinks while the size of the α region expands over time. In part (ii), the opposite trend is observed. In both cases, the process leads to the emergence of a single convention for both communities.[23]

Theorems 1 and 2 show that when a risk-dominant convention interacts with a Pareto-efficient convention, the cost of remaining flexible, c, is crucial in determining which convention will survive in the long run. For completeness, we also analyze the case where risk-dominance and Pareto-efficiency considerations are not in conflict. In this case we obtain the following result:

THEOREM 3. *Suppose* A.1′–A.2. *hold and* $\delta<1/2$. *If* $s_0\in \mathrm{I}_r$ *then all players conform to the* α *convention after a finite period.*

The proof of this result follows from two observations: one, that so long as $2c<b-e$, the hypothesis (A.1′) implies the inequality $2c(d-e)<(a-b)(b-e)$ and two, that if $2c>b-e$ then no individual optimally chooses the action (α,β) and from standard arguments, the process converges to the risk-dominant convention. This implies that under (A.1′), only the α convention survives in the long run. Thus when there is no conflict between risk-dominance and Pareto-efficiency considerations the value of c is not important.[24]

4. UNIFORM INTERACTION

In this section we will analyze the dynamic process under uniform interaction. This analysis is based on the computation of the basins of attraction

[23] At some finite point in time when the shrinking region becomes sufficiently small, the difference equations are no longer adequate to describe the movement of boundaries. This is because the difference equations are derived under the assumption that the α and β regions are sufficiently large so that the analysis of the movement of one boundary can be carried out independently of the movement of the other. The proof of Theorem 2 provides additional arguments showing that when a region becomes very small, it will eventually disappear.

[24] It is worth adding that Theorems 2 and 3 do not require the full force of the hypothesis, $s_0\in \mathrm{I}_r$. for instance, the conclusion of Theorem 2(i) obtains under the weaker condition, $|\Gamma_0^\alpha|>\delta+2\delta(1-(2c)/(b-e))$. We retain the stronger hypothesis only to maintain consistency with Theorem 1.

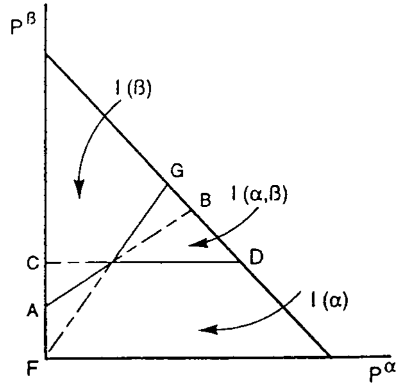

FIGURE 2

of the different social equilibria. In recent years, uniform interaction has been studied extensively and these computations are relatively well-known. The presentation is accordingly brief and informal. The basins of attraction depend upon the values of the parameters, *a*, *b*, *c*, *d*, *e*. An illustrative example of the basins of attraction is presented in Fig. 2 for the case in which a coexistence equilibrium exists.

This figure is obtained by deriving the locus of points at which individual agents are indifferent between different actions. Thus the segment AB is the locus of proportions at which $\pi_x(\alpha \mid s) = \pi_x(\beta \mid s)$ and is defined by $p^\beta = (a-b)/(a-e) + p^\alpha(b-d)/(a-e)$. The line segment CD is the locus of points at which $\pi_x(\alpha \mid s) = \pi_x((\alpha, \beta) \mid s)$ and is defined by $p^\beta = c/(b-e)$. Finally, line segment FG, which is the locus of points at which actions β and (α, β) are equally attractive, is defined by $p^\beta = 1 - c/(a-b) + p^\alpha(b-d)/(a-b)$. In the figure, $I(\alpha)$, $I(\beta)$ and $I(\alpha, \beta)$ denote the basins of attraction of configurations $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, respectively. It is important to note that the social configuration in which everyone chooses (α, β), $(0, 0, 1)$, lies in the basin of attraction of $(1, 0, 0)$. Thus $I(\alpha, \beta)$ is actually a subset of the basin of attraction of the Pareto-efficient equilibrium. Given A.1, it is immediate that the slope of FG is larger than the slope of AB and this implies that a coexistence equilibrium is only possible if these line segments have the configuration depicted in Fig. 2.

It is now easy to see that starting from any $s_0 \in \mathrm{I}(\alpha)$ under (E) and (B), $p_t = (1, 0, 0)$ for all $t \geqslant 1$. Similar arguments establish that starting from $s_0 \in \mathrm{I}(\beta)$, $p_t = (0, 1, 0)$ for all $t \geqslant 1$. Finally, if $s_0 \in \mathrm{I}(\alpha, \beta)$ then under (E) and (B) every player chooses (α, β) in period $t = 1$ implying that $p_1^{(\alpha, \beta)} = 1$. Under (E), $E(s_2 \mid s_1) = s_1$ and it follows from (B) that $p_2 = (1, 0, 0)$. From this it is immediate that the coexistence equilibrium is dynamically unstable under uniform interaction.

A second observation pertains to the role of cost the c in determining long-run outcomes. From the expressions for AB, CD and FG and Fig. 2, it is easily seen that the size of $I(\alpha)$, $I(\beta)$ and $I(\alpha, \beta)$ are related to the level of c. Thus if c is lowered then $I(\beta)$ shrinks while the effective basin of attraction for α which is the complement of the set $I(\beta)$ expands. In this sense the Pareto-efficient equilibrium can be said to become more likely as the level of c declines.[25] Note that under local interaction low values of c also favour convergence to the Pareto-efficient equilibrium.

Our third observation pertains to the relative importance of the initial size of different regions in determining long-run outcomes. To illustrate this point consider high and low values of c. When the cost is large, say $c > (b-e)/2$ then the dual-action is never optimal and yet for initial values of $s_0 \in I(\alpha)$ the dynamic process $s_t \to (1, 0, 0)$, the Pareto-efficient equilibrium. In contrast, there are low values of this cost, say $c < (a-b)/2$, where $I(\beta)$ is non-empty and for any $s_0 \in I(\beta)$ the dynamic process $s_t \to (0, 1, 0)$, the risk-dominant equilibrium. Comparing these results to Theorems 2 and 3 of the previous section leads us to the conclusion that the initial size of the α and β regions plays a more important role under uniform interaction than under local interaction.

5. CONCLUDING REMARKS

We have argued that when a community with a Pareto-efficient convention interacts with a community with a risk-dominant convention, the costs of becoming adept in both conventions play an important role in determining if one convention prevails over the other or if the conventions coexist. Our analysis shows that when these costs are low then the Pareto-efficient convention prevails in both communities eventually, while the risk-dominant convention prevails for high cost levels. For intermediate cost levels the two conventions coexist.

Our argument was developed with the help of a simple dynamic model. Two features of this model—the interaction structure and the dynamic process of individual choice—deserve special mention. We assumed that individuals are located around a circle and interact only with some interval of agents around them. We also supposed that individual choice was defined by a form of myopic best response. We have considered alternative formulations to check for the robustness of our findings. In particular, we find that a similar result obtains if interaction is two-dimensional and

[25] This suggests that there may be a relationship between c and the stochastic stability of different conventions. In a subsequent paper, Galesloot and Goyal [8] study this issue.

individuals are located on a torus. We also examined a variety of extensions with other versions of the best response dynamic, which involved n-period adaptive expectations and infrequent strategy revision rules.[26] Our analyses suggest that the basic intuition concerning the role of these costs is sound.

APPENDIX A: PROOFS

The proof of Theorem 1 rests upon two Lemmas, 1.1 and 1.2; we state and prove these results. Let $D_\alpha = \delta(1-(2c)/(b-e))$ and $D_\beta = \delta(1-(2c)/(a-d))$. Also let I_r denote the interval of initial configurations in which $|\Gamma_0^\alpha| > \delta + 2D_\alpha$ and $|\Gamma_0^\beta| > \delta + 2D_\alpha$.

LEMMA 1.1. *Let* $s_0 \in \mathrm{I}_r$. *Then* $\{x_i^B\}$ *is well defined and for* $t \geqslant 1$, *it is true that*:

(a) $x_t^{\alpha,1} \in [x_0^{\alpha,1} - D_\alpha, x_0^{\alpha,1}]$; $x_t^{\alpha,2} \in [x_0^{\alpha,2}, x_0^{\alpha,2} + D_\alpha]$; $x_t^{\beta,1} \in [x_0^{\beta,1}, x_0^{\beta,1} + D_\alpha]$; $x_t^{\beta,2} \in [x_0^{\beta,2} - D_\alpha, x_0^{\beta,2}]$

(b) $|\Gamma_t^\alpha| \geqslant \delta$; $|\Gamma_t^\beta| \geqslant \delta$.

Proof of Lemma 1.1. First note that by definition of I_r, x_0^B is well defined, and naturally satisfies requirements (a) and (b).

Recall that in period 1,

$$E(x_1^{\alpha,1} \mid x_0^B) = x_1^{\alpha,1};\ E(x_1^{\alpha,2} \mid x_0^B) = x_0^{\alpha,2} \tag{1}$$

$$E(x_1^{\beta,1} \mid x_0^B) = x_1^{\beta,1};\ E(x_1^{\beta,2} \mid x_0^B) = x_0^{\beta,2} \tag{2}$$

The payoffs for any $x \in [x_0^{\alpha,1} - D_\alpha, x_0^{\alpha,1}]$ may be expressed as

$$\pi_x(\alpha \mid x_0^B) = \frac{1}{2\delta}\delta a + \frac{1}{2\delta}|x, x_0^{\alpha,1}|\, a + \frac{1}{2\delta}|x_0^{\alpha,1}, x+\delta|\, e \tag{3}$$

$$\pi_x(\beta \mid x_0^B) = \frac{1}{2\delta}|x_0^{\alpha,1}, x+\delta|\, b + \frac{1}{2\delta}|x-\delta, x_0^{\alpha,1}|\, d \tag{4}$$

$$\pi_x(\alpha, \beta \mid x_0^B) = \frac{1}{2\delta}|x-\delta, x_0^{\alpha,1}|\, a + \frac{1}{2\delta}|x_0^{\alpha,1}, x+\delta|\, b - c. \tag{5}$$

Since $2c < (a-d) \leqslant (b-e)$, Eqs. (3)–(5) show that for $x \in (x_0^{\alpha,1} - D_\alpha, x_0^{\alpha,1}]$, both α and β are payoff dominated by (α, β). Moreover, by the

[26] Details of these extensions are available from the authors upon request.

continuity of payoffs with respect to x (which is easily checked), it follows that at $x' = x_0^{\alpha,1} - D_\alpha$,

$$\pi_{x'}((\alpha, \beta) \mid x_0^B)) = \pi_{x'}(\alpha \mid x_0^B)) > \pi_{x'}(\beta \mid E(x_1^B)), \tag{6}$$

and for $x \in [0, x')$, since $|\Gamma_0^\alpha| > \delta + D_\alpha$, α yields a higher payoff than (α, β) and β. These arguments enable us to define

$$x_1^{\alpha,1} = x_0^{\alpha,1} - D_\alpha. \tag{7}$$

Similar arguments show that the other boundary points may be defined as

$$x_1^{\alpha,2} = x_0^{\alpha,2} + D_\alpha \tag{8}$$

$$x_1^{\beta,1} = x_0^{\beta,1} + D_\beta \tag{9}$$

$$x_1^{\beta,2} = x_0^{\beta,2} + D_\beta. \tag{10}$$

From Eqs. (7)–(10) it is immediate that x_1^B satisfies requirement (a); moreover, since $|\Gamma_0^\alpha| > \delta + 2D_\alpha$, it follows that $|\Gamma_1^\alpha| = |\Gamma_0^\alpha| - 2D_\alpha > \delta$. Finally, it can be shown that $|\Gamma_1^\beta| > \delta$ and requirement (b) is also satisfied in period 1.

We next consider period 2. Recall that $(x_2^{\alpha,1}, x_2^{\beta,1})$ depend on $x_1^{\alpha,1}$ and $x_1^{\beta,1}$. There are two cases to consider here: (i) $|x_1^{\alpha,1}, x_2^{\beta,1}| \leqslant D_\alpha$, and (ii) $|x_2^{\alpha,1}, x_2^{\beta,1}| > D_\alpha$. We shall suppose that $x_2^{\beta,2} \in [x_0^{\beta,1}, x_0^{\beta,1} + D_\alpha]$ and work out the nature of payoffs and hence boundaries on this basis. Once these boundaries are found, we shall show that they lie in the interval initially assumed.

Case (i). In this case, the payoff computations can be made using the same general principles as in stage 1. In particular, at $x_2^{\beta,1}$ the payoffs from action β and action (α, β) must be equal and this implies that

$$x_2^{\beta,1} = \frac{(a-b)\, x_1^{\beta,1} + (b-d)\, x_1^{\alpha,1}}{a-d} + \delta\left(1 - \frac{2c}{a-d}\right). \tag{11}$$

Case (ii). Similar calculations as in case (i) show that the boundary $x_2^{\beta,1}$ is given by Eq. (11) if $x_1^{\alpha,1} \geqslant x_2^{\beta,1} - \delta$. Simple calculations show that $x_1^{\alpha,1} \geqslant x_2^{\beta,1} - \delta$ if

$$\frac{c}{b-e} + \frac{c}{a-d} + \frac{c}{a-b} \geqslant 1. \tag{C}$$

We proceed assuming that (C) holds; the effects of relaxing (C) are considered later.

It is easily seen that in both case (i) and (ii)

$$x_2^{\alpha,1} = x_1^{\beta,1} - D_\alpha. \tag{12}$$

Hence, the boundaries $(x_2^{\alpha,1}, x_2^{\beta,1})$ are well defined at stage 2. Moreover, from (9)–(12), it follows that $x_2^{\alpha,1}$ satisfies (a). Next we consider $x_2^{\beta,1}$. Simple calculations suggest that (11) can be rewritten as

$$x_2^{\beta,1} = \frac{(a-b)\,x_1^{\beta,1} + (b-d)\,x_0^{\beta,1}}{a-d} - \frac{b-d}{a-d} D_\alpha + D_\beta. \tag{13}$$

Since $2c(d-e) = (a-b)(b-e)$, it can be shown that $x_2^{\beta,1} \in [x_0^{\alpha,1}, x_0^{\alpha,1} + D_\alpha]$.

Analogous arguments may be made for the pair $x_2^{\alpha,2}$ and $x_2^{\beta,2}$. This shows that (a) is satisfied at stage 2. Finally, since $|\Gamma_0^\alpha| > \delta + 2D_\alpha$ and $|\Gamma_0^\beta| > \delta + 2D_\alpha$, and given that x_2^B satisfies (a), it follows that x_2^B also satisfies (b). The proof now follows by induction on t.

We now consider the case when (C) is violated; payoffs calculations reveal that period 2 boundaries are given by: $x_2^{\beta,1} = x_1^{\beta,1} + \delta(1-(2c)/(a-b))$ and $x_2^{\alpha,1} = x_1^{\beta,1} - \delta(1-2c)/(b-e))$. First, note that the violation of (C) implies that $x_2^{\beta,1} > x_2^{\alpha,1}$. Second, given (A.1) and $2c(d-e) = (a-b)(b-e)$ it follows that $2c > (a-b)$ and it is easily verified that $x_0^{\alpha,1} < x_1^{\beta,1} + \delta(1-(2c/a-b)) < x_0^{\alpha,1} + D_\alpha$. So, (a) and (b) are satisfied in period 2. Finally, one can show that in period 3 boundaries are given by expressions similar to (10) and (11). This is because $x_2^{\beta,1} - x_2^{\alpha,1} = 2\delta(1-(c/(a-b) - (c/(b-e)))$ which is less than $(2c\delta)/(a-b)$, and thus, $x_2^{\alpha,1} \geqslant x_3^{\beta,1} - \delta$. The rest of this case then follows from earlier arguments. Q.E.D.

LEMMA 1.2. *$x_t^B \to \hat{x}^B$ where $\hat{x}^{\alpha,1} = \hat{x}^{\beta,1} - D_\alpha$; $\hat{x}^{\alpha,2} = \hat{x}^{\beta,2} + D_\alpha$.*

Proof of Lemma 1.2. Consider any $s_0 \in \mathrm{I}_r$. From Lemma 1.1 we know that the sequence $\{x_t^B\}$ is well defined and that its long-run properties may be studied in terms of two different sub-processes, pertaining to the boundaries of $\Gamma^{\alpha,\beta}(1)$ and $\Gamma^{\alpha,\beta}(2)$, respectively. These boundaries are defined by the following difference equations.

$$x_t^{\alpha,1} = x_{t-1}^{\beta,1} - D_\alpha \tag{14}$$

$$x_t^{\alpha,2} = x_{t-1}^{\beta,2} - D_\alpha \tag{15}$$

$$x_t^{\beta,1} = \frac{(a-b)\,x_{t-1}^{\beta,1} + (b-d)\,x_{t-2}^{\beta,1}}{a-d} - \frac{b-d}{a-d} D_\alpha + D_\beta \tag{16}$$

$$x_t^{\beta,2} = \frac{(a-b)\,x_{t-1}^{\beta,2} + (b-d)\,x_{t-2}^{\beta,2}}{a-d} - \frac{b-d}{a-d} D_\alpha + D_\beta. \tag{17}$$

We focus on the sub-process $\{x_t^{\alpha,1}, x_t^{\beta,1}\}$ and begin by noting that the behaviour of the sequence is captured by (14) and (16). We can show that if $2c(d-e)=(a-b)(b-e)$ then $(b-d)/(a-d)\, D_\alpha - D_\beta = 0$. Hence, we can write (16) as

$$x_t^{\beta,1} = \frac{(a-b)\, x_{t-1}^{\beta,1} + (b-d)\, x_{t-2}^{\beta,1}}{a-d}. \tag{18}$$

Equation (18) is a homogeneous difference equation with the coefficients adding up to 1. The characteristic equation for this homogeneous difference equation may be factorized as follows: $(z-1)((a-d)\, z+(b-d))=0$. It is easily seen that for this characteristic equation, one root is equal to 1 and under $a>b$ the other has an absolute value of strictly less than 1. Standard considerations allow us to say then that the sequence $\{x_t^{\beta,1}\}$ converges to a limit $\hat{x}^{\beta,1}$. Equation (14) then tells us that $x_t^{\alpha,1} \to \hat{x}^{\alpha,1}$, where $\hat{x}^{\alpha,1} = \hat{x}^{\beta,1} - D_\alpha$. Q.E.D.

Proof of Theorem 2(i). We study the sequence $\{x_t^B\}$, under the condition $2c(d-e)<(a-b)(b-e)$. From computations in Lemma 1.1, it follows that x_1^B is defined as in Eqs. (7)–(10). Likewise, $(x_2^{\alpha,1}, x_2^{\beta,1})$ is defined as in Eqs. (12)–(13). These computations along with arguments in Lemma 1.1 show that: $x_2^{\beta,1} \geqslant x_0^{\beta,1}$, $x_2^{\beta,2} \leqslant x_0^{\beta,2}$, $x_2^{\alpha,1} \geqslant x_0^{\alpha,1} - D_\alpha$, $x_2^{\alpha,2} \leqslant x_0^{\alpha,2} + D_\alpha$.

It is possible to show that the Eqs. (14)–(17), may be used to define x_t^B so long as the following relations are satisfied:

$$\frac{(a-b)\, x_{t-1}^{\beta,1} + (b-d)\, x_{t-1}^{\alpha,1}}{a-d} + \delta\left(1-\frac{2c}{a-d}\right) \leqslant x_{t-1}^{\beta,2} - \delta \tag{19}$$

$$\frac{(a-b)\, x_{t-1}^{\beta,2} + (b-d)\, x_{t-1}^{\alpha,2}}{a-d} - \delta\left(1-\frac{2c}{a-d}\right) \leqslant x_{t-1}^{\beta,1} + \delta \tag{20}$$

$$x_{t-1}^{\beta,1} - D_\alpha \geqslant x_{t-1}^{\beta,2} + \delta \tag{21}$$

$$x_{t-1}^{\beta,2} + D_\alpha \leqslant x_{t-1}^{\beta,1} - \delta \tag{22}$$

(Note that when $2c \geqslant (a-b)$ these equations also define the boundary process from period 3 onwards when (C) is violated (see the argument in the proof of Lemma 1.1). When $2c<(a-b)$ convergence follows from the monotonicity of the $\{x_t^{\beta,1}\}$ and $\{x_t^{\beta,2}\}$ sequences; details of the argument are omitted.)

Observe that since $x_k^{\beta,1} \geqslant x_0^{\beta,1}$, $x_k^{\beta,2} \leqslant x_0^{\beta,2}$, for $k=1, 2$ Eqs. (21)–(22) are satisfied for $t=1, 2, 3$. Furthermore, it can be shown that if inequalities in Eqs. (19)–(20) are satisfied up to period $t-1$, then $x_k^{\beta,1} \geqslant x_0^{\beta,1}$, $x_k^{\beta,2} \leqslant x_0^{\beta,2}$, for $k=1, 2, ..., t-1$, thus implying that inequalities in Eqs. (21)–(22) will be satisfied for period t. The process $\{x_t^B\}$ is thus well defined so long as inequalities (19)–(20) are satisfied.

The next step shows that if (19)–(20) are not satisfied then convergence to the Pareto-dominant convention obtains. There are two cases to be considered here:

Case (i). Equations (19)–(20) are not satisfied and in addition

$$\frac{(a-b)\, x_{t-1}^{\beta,1}+(b-d)\, x_{t-1}^{\alpha,1}}{a-d}+\delta\left(1-\frac{2c}{a-d}\right)<x_{t-1}^{\alpha,2}-\delta \tag{23}$$

$$\frac{(a-b)\, x_{t-1}^{\beta,2}+(b-d)\, x_{t-1}^{\alpha,2}}{a-d}-\delta\left(1-\frac{2c}{a-d}\right)>x_{t-1}^{\alpha,1}+\delta \tag{24}$$

are also violated. In this case, at $\bar{x}$ equal to the L.H.S. of the inequality, $\pi_{\bar{x}}(\beta \mid x_{t-1}^B)<\pi_{\bar{x}}((\alpha,\beta) \mid x_{t-1}^B)$. From Lemma 1.1 we know that for all $x<\bar{x}$, $\pi_x(\beta \mid x_{t-1}^B)<\pi_x((\alpha,\beta) \mid x_{t-1}^B)$. It is easily seen that $\pi_x(\beta \mid x_{t-1}^B)$ and $\pi_x((\alpha,\beta) \mid x_{t-1}^B)$ are constant in the interval $(x_{t-1}^{\alpha,2}-\delta, x_{t-1}^{\alpha,1}+\delta)$, and so $\pi_x(\beta \mid x_{t-1}^B)<\pi_x((\alpha,\beta) \mid x_{t-1}^B)$ for all $x\in\Gamma_{t-1}^{\beta}$. This means that $|\Gamma_t^{\beta}|=0$ and under (B) that completes the proof.

Case (ii). This is the more general case where (19)–(20) are not satisfied but (23)–(24) are satisfied. In this case, the argument used for case 1 does not apply and convergence to the α convention may not be immediate. We use an indirect argument. This argument is quite simple but tedious, so we provide a sketch only: We begin by noting the main difference between the two cases: in case (ii), $|\Gamma_t^{\beta}|\neq 0$. If we denote the boundary calculated under (19) by $\bar{x}$ then it is easily seen that in the present case, $\pi_{\bar{x}}(\beta \mid x_{t-1}^B)<\pi_{\bar{x}}((\alpha,\beta) \mid x_{t-1}^B)$ and thus the true boundary $x_t^{\beta,1}>\bar{x}$. Let $\bar{t}$ be the first period in which case (ii) obtains. Then the above observation yields the following general insight: for all $t\geqslant\bar{t}$, $x_t^{\beta,1}\geqslant\bar{x}_t^{\beta,1}$ and $x_t^{\beta,2}\leqslant\bar{x}_t^{\beta,2}$, where, for example, $\bar{x}_t^{\beta,1}$ is the hypothetical boundary calculated by using (19)–(22) over time. Convergence now obtains as a corollary of the convergence of the $\bar{x}_t^B$ process, which is established below.

It then only remains to show that at some finite point the inequalities in (19)–(20) will be violated. The proof proceeds by way of contradiction. Suppose not. Focus, without loss of generality, on the inequality in expression (19). Note that so long it is not violated, the process $\{x_t^{\beta,1}\}$ is defined using the formula given by Eq. (19).

Note that since $2c(d-e)<(a-b)(b-e)$, it follows that $D_{\beta}-(b-d)/(a-d)\, D_{\alpha}>0$. We know from the proof of Lemma 1.2 that the homogeneous component converges. Moreover, the particular solution of the difference equation is of the form Kt, with $K>0$, and so the particular solution will dominate and determine the behaviour of the sequence $\{x_t^{\beta,1}\}$, in the long run. Thus, the sequences $x_t^{\beta,1}$ and $x_t^{\beta,2}$ are, respectively, increasing and decreasing without bound, in the long run. This implies that there is some finite point t at which (19) is violated, which contradicts our supposition. Q.E.D.

Proof of Theorem 2(ii). By hypothesis $2c(d-e)>(a-b)(b-e)$. It is easily calculated that this implies that $D_\beta-(b-d)/(a-d)\,D_\alpha<0$. This observation along with arguments analogous to Theorem 2(i) show that at some finite point $\bar{t}$, Eqs. (21)–(22) will be violated and thus $|\Gamma_{\bar{t}}^{\alpha}|=0$.[27]

This does not, however, complete the proof since the β convention may be strictly Pareto-dominated by convention α and if the (α,β) region is very large then it may become optimal for some individuals to switch back to α. It is possible to show that, in the long run, this cannot happen. We provide a sketch of the proof for this claim.

The first step is to show, using arguments analogous to Theorem 2(i), that the actual sequence $\{x_t^B\}$ is bounded uniformly by an artificial process $\{\bar{x}^B\}$, i.e., $x_t^{\alpha,1}\leqslant\bar{x}_t^{\alpha,1}$, $x_t^{\alpha,2}\geqslant\bar{x}_t^{\alpha,2}$, $x_t^{\beta,1}\leqslant\bar{x}_t^{\beta,1}$ and $x_t^{\beta,2}\geqslant\bar{x}_t^{\beta,2}$. The artificial process $\{\bar{x}_t^B\}$ is defined as in (14)–(17) except that $\bar{x}_t^{\alpha,1}=\max\{x_{t-1}^{\beta,1}-D_\alpha,0\}$ and $\bar{x}_t^{\alpha,2}=\min\{x_{t-1}^{\beta,2}+D_\alpha,0\}$. The second step is to show that the artificial process converges with the limit given by: $\bar{x}^{\alpha,1}=\bar{x}^{\alpha,2}=0$, $\bar{x}^{\beta,1}=D_\beta(a-d)/(b-d)$ and $\bar{x}^{\beta,2}=-D_\beta(a-d)/(b-d)$. The third step uses this limiting result to argue that after some finite point $\bar{x}_t^{\beta,1}<D_\alpha$, and $\bar{x}_t^{\beta,2}>-D_\alpha$. This observation implies that (19)–(20) are violated and that $|\Gamma_t^{\alpha}|=0$ after some finite t. The final step uses this fact along with the hypothesis that $2c(d-e)>(a-b)(b-e)$ (implying that $2c>(a-b)$) to show that $\{\bar{x}^{\beta,1}\}$ and $\{\bar{x}_t^{\beta,2}\}$ are, respectively, monotonically strictly decreasing and increasing after some finite t. That completes the proof. Q.E.D.

APPENDIX B: COEXISTENCE OF CONVENTIONS

In this appendix we examine the robustness of coexistence in terms of the values of c. We also ask if this robustness of coexistence is related to the dimensionality of interaction. We are mainly concerned with coexistence of the following form: there is a region of agents conforming to the α convention, a region conforming to the β convention and an intermediate

[27] The hypotheses here allow for $a=b$; in this case the homogeneous part of the difference equation is given by, $x_t^{\beta,j}=x_{t-2}^{\beta,j}$ for $j=1,2$. The same arguments as in Theorem 2(i) can now be applied.

boundary region, which comprises of agents adept in both conventions. Such a pattern is depicted below.

$$-------\overbrace{|-|-|-|-|-1}^{\alpha}-\overbrace{2-|-|-3}^{(\alpha, \beta)}-\overbrace{4-|-|-|-|-|}^{\beta}-------$$

In this figure, agents 1 and 2 represent the boundary between the α and (α, β) regions while agents 3 and 4 represent the boundary between the (α, β) and β regions. This configuration allows for the possibility that there are other agents between 2 and 3 as well as the possibility that agents 2 and 3 are the same person.

The pattern given above is sustainable only if the boundary agents are optimally choosing the actions. Let η^i_k refer to the likelihood of agent i interacting with an agent choosing an action k. Then for agent 1, for action α to be preferred to (α, β) it must be true that $c \geqslant \eta^1_\beta(b-e)$. Likewise, for agent 2 to prefer (α, β) over α it must be true that $c \leqslant \eta^2_\beta(b-e)$. This implies that for the above mentioned configuration to be sustainable c must satisfy the requirement: $\eta^1_\beta(b-e) \leqslant c \leqslant \eta^2_\beta(b-e)$. Likewise, for agent 3 to prefer (α, β) over β and for agent 4 to prefer β over (α, β) it must be true that c satisfies the requirement: $\eta^4_{(\alpha, \beta)}(a-b) + \eta^4_\alpha(a-d) \leqslant c \leqslant \eta^3_{(\alpha, \beta)}(a-b) + \eta^3_\alpha(a-d)$. It is then immediate that if the neighborhoods of boundary individuals are similar then coexistence is problematic. If on the other hand the neighborhoods of boundary agents are significantly different then coexistence is likely to be more robust. This observation explains why coexistence is non-generic in the continuum agent model presented in the paper: in that model the neighborhoods of boundary agents are (in a measure-theoretic sense) identical. Exploiting this insight, it is quite easy to see that coexistence is robust when a finite number of agents located on a line interact with their four immediate neighbors. Straightforward calculations reveal that coexistence of two conventions, with one mediating agent choosing (α, β), is sustainable so long as the following conditions are met: $\max\{\frac{1}{4}(b-e), \frac{1}{4}(a-d) + \frac{1}{4}(a-b)\} < c < \frac{1}{2}(a-d)$. It is easily checked that these conditions are satisfied by a wide range of parameter values.

The above discussion shows that coexistence in a one-dimensional interaction model is robust with a finite number of agents. This suggests that it is not the dimensionality of interaction *per se* that determines the robustness of coexistence.[28]

[28] It also gives some intuition as to why coexistence is likely to be more robust in a model with a finite number of agents who are located on a torus and interact with their (say) 8 immediate neighbors. In such a setting, the neighborhoods for two agents in adjacent rows or columns differ by three agents, which is a significant difference given that an agent's neighborhood comprises of nine agents (including herself). Anderlini and Ianni [2] show that coexistence is robust in a model of local interaction with agents located on a torus. The above discussion also helps us place their results on coexistence in perspective.

REFERENCES

1. M. An and N. Kiefer, A dynamic model of local and global interaction among economic agents, mimeo, Cornell University, Ithaca, 1993.
2. L. Anderlini and A. Ianni, Path dependence and learning form neighbour, *Games Econ. Behavior* **13** (1996), 141–178.
3. S. Anderson, A. de Palma, and J. F. Thisse, "Discrete Choice Theory of Product Differentiation," MIT Press, Cambridge, MA, 1991.
4. L. Blume, The statistical mechanics of strategic interaction, *Games Econ. Behavior* **5** (1993), 387–424.
5. R. Boyer and A. Orlean, How do conventions evolve?, *J. Evol. Econ.* **2** (1992), 165–177.
6. V. Crawford, Adaptive dynamics in coordination games, *Econometrica* **63** (1995), 103–144.
7. G. Ellison, Learning, local interaction and coordination, *Econometrica* **61** (1993), 1047–1073.
8. B. Galesloot and S. Goyal, Costs of flexibility and equilibrium selection, *J. Math. Econ.*, in press.
9. S. Goyal, Interaction structure and social change, *J. Inst. Theoretical Econ.* **152** (1996), 472–495.
10. J. Harsanyi and R. Selten, "A General Theory of Equilibrium Selection in Games," MIT Press, Cambridge, MA, 1988.
11. M. Kandori, G. Mailath, and R. Rob, Learning mutation and long run equilibrium in games, *Econometrica* **61** (1993), 29–56.
12. D. K. Lewis, "Conventions: A Philosophical Study," Harvard Univ. Press, Cambridge, MA, 1969.
13. G. Mailath, Introduction: Symposium on evolutionary game theory, *J. Econ. Theory* **57** (1992), 269–277.
14. K. Matsuyama, N. Kiyotaki, and A. Matsui, Towards a theory of international currency, *Rev. Econ. Studies* **60** (1993), 283–309.
15. T. Schelling, "The Strategy of Conflict," Harvard Univ. Press, Cambridge, MA, 1960.
16. R. Sugden, The coexistence of conventions, *J. Econ. Behavior Organization* **28** (1995), 241–256.
17. P. Young, The evolution of conventions, *Econometrica* **61** (1993), 57–84.
18. P. Young, The economics of conventions, *J. Econ. Perspectives* **10** (1996), 105–22.