Random Boolean Networks and Evolutionary Game Theory

J. McKenzie Alexander

Date: 10 December 2002

Abstract

Recent years have seen increased interest in the question of whether it is possible to provide an evolutionary game theoretic explanation for certain kinds of social norms. I sketch a proof of a general representation theorem for a large class of evolutionary game theoretic models played on a social network, in hope that this will contribute to a greater understanding of the long-term evolutionary dynamics of such models, and hence the evolution of social norms. More precisely, I show how many kinds of social networks can be translated into random boolean networks. The interesting and useful part of this result consists of the demonstration that, for many social networks, there is a bijection f between the state space of the social network and the state space of the random boolean network, such that the state S' follows the state S under the dynamical laws of the social network if and only if $f(S')$ follows the state $f(S)$ under the dynamics of the random boolean network. In some cases, it is not possible to find such a bijection; in these cases, one can find an injection f with the property that if S' follows S under the dynamics of the social network, then $f(S')$ follows $f(S)$ under the dynamics of the random boolean network. I then use this method to catalog all the basins of attraction for some simple two-strategy games (the prisoner's dilemma and the stag hunt) played on a ring, drawing on the work of Wuensche and Lesser (1992).

1. Introduction. Recent years have seen increased interest in the question of whether it is possible to provide an evolutionary game theoretic explanation for certain kinds of social norms [see, for example, Skyrms (1994), Skyrms (1996), D'Arms et!al. (1998), Alexander and Skyrms (1999), and Alexander (2000)]. Those who seek to provide such an account typically proceed as follows: first, one identifies a particular norm of interest -- the *explanandum* -- as well as a game representing the choice problem in which that norm is typically invoked. Second, interpreting the evolutionary game theoretic account as a model of cultural evolution,[1] one constructs a model of the process in which boundedly rational individuals interact, learn, and change their beliefs or behaviors. The goal is to demonstrate that the formal strategy or strategies corresponding to the social norm under study exist in some kind of equilibrium under the associated dynamics. If such an equilibrium exists, and the population converges to that equilibrium often enough, it is then argued that this provides, as Skyrms said regarding the concept of justice, "a beginning of an explanation of the origin" of the norm in question (Skyrms, 1994, pg.!320).

In this kind of explanatory account, it is important to note the respective roles played by (1) the fact that the strategies representing a social norm exist in some kind of equilibrium under the associated dynamics, and (2) the fact that the population converges *to that equilibrium* "often enough." The first point explains the stability and (perhaps) the normativity attached to that norm: the reason why people continue to follow it is that, once arrived at, that norm provides a reasonably successful solution to the interdependent decision problem at hand, such that no individual can expect to do better by deviating from it.[2] Under the

---

[1] This need not commit one to talk of "memes" or sociobiological explanations. See Bögers and Sarin (1993) for a model of imitative learning that leads to the replicator dynamics and Lachapelle (2000) for an argument on why cultural evolution may be viewed as a process independent of biological evolution.

[2] Notice that this explanation of why people continue to follow norms does not imply that it is impossible for norms to change. If the arena in which social interaction occurs changes sufficiently, the game used to

assumption that a person will not deliberately choose to follow a course of action that makes himself or herself worse off than before, the norm is "fixed" in the population since attempts to depart from it are not fruitful and consequently eliminated.[3] The second point explains why, in fact, we actually find ourselves following *that* norm rather then some other norm. In the ideal case[4], one finds that the population *always* arrives at the equilibrium corresponding to the norm we actually follow, no matter how it may have started. In this case it is clear why we find that norm present in society: the dynamics of boundedly rational interaction inevitably drive populations towards adopting that norm.[5]

In this paper I will not engage the question of whether this explanatory strategy succeeds or fails. My concern here lies with the question of how one establishes the second claim above, namely, that populations converge to certain equilibria "often enough." In general, this is the problem of calculating the basins of attraction for a given dynamical system and of determining their relative size.[6] This question points up an inherent difficulty lying at the heart of the evolutionary game theoretic approach. While one should take the results from a purported evolutionary game theoretic explanation of social norms seriously to the extent that the underlying model accurately represents the relevant aspects of social interaction among members of the population, the construction of a more accurate model often requires rejecting certain simplifying assumptions which facilitate calculating basins of attraction of a given equilibrium. That is, as the empirical adequacy of the evolutionary model increases (thus giving us more reason to take its results seriously), it becomes harder to establish the formal results that underwrite the explanation.

For example, Skyrms (1994) uses the replicator dynamics to develop an evolutionary game theoretic model of a population of boundedly rational individuals playing the game of "divide-the-dollar." In order to justify the use of the replicator dynamics, one must assume (in part) that the population is essentially infinite and perfectly mixed, such that all pairwise interactions are equally likely. These assumptions appear inaccurate for describing real human societies. As D'Arms et!al. (1998) note, one finds important differences in the frequencies with which certain equilibria arise when considering a finite population model. In a previous paper, I argued that considering finite populations in which interactions are constrained by an underlying social network also produces interesting differences in the kinds of equilibria which emerge. This illustrates the problem described above. Moving from Skyrms's replicator dynamic model to a finite-population, social network model involves moving from a continuous dynamical system described by a set of differential equations to a *discrete* dynamical system described by a set of transition rules, and, generally speaking, analyzing discrete systems is a more difficult problem than analyzing continuous systems. If we need to consider finite-population, social network models in order to construct an empirically adequate model of the evolution of social norms, we must improve our ability to determine the basins of attractions of such models. This paper seeks to establish one result which, it is hoped, will contribute to the general solution of this problem.

---

represent the original social context will no longer be an accurate representation. A new game representing the new social context would need to be introduced, and it would be an open question whether strategies in equilibrium in the original game would also be in equilibrium in the new game.

[3] I do not intend the talk of "eliminating" strategies, attempts, or individuals from a population to be taken literally. As I am primarily interested in discussing cultural evolutionary game theoretic models, these phrases should be interpreted as saying only that no one in the population follows the strategy.

[4] Which may exist for certain cases of dividing resources between perfectly symmetric individuals. For further discussion, see Alexander (2000).

[5] This claim, of course, is subject to the qualification that the model of the interactive dynamics is reasonably accurate.

[6] In what follows, I shall assume that the dynamical systems under consideration are deterministic.

In what follows, I sketch a proof of a general representation theorem for a large class of evolutionary game theoretic models played on a social network, in hope that this will contribute to a greater understanding of the long-term evolutionary dynamics of such models. More precisely, I show how many kinds of social networks can be translated into random boolean networks. The interesting and useful part of the theorem consists of the demonstration that, for many social networks, there is a bijection f between the state space of the social network and the state space of the random boolean network, such that the state S' follows the state S under the dynamical laws of the social network if and only if $f(S')$ follows the state $f(S)$ under the dynamics of the random boolean network. In some cases, it is not possible to find such a bijection; in these cases, one can find an injection f with the property that if S' follows S under the dynamics of the social network, then $f(S')$ follows $f(S)$ under the dynamics of the random boolean network. I then use this method to catalog all the basins of attraction for some simple two-strategy games played on social networks using the algorithm of <u>Wuensche and Lesser (1992)</u>.

2. Random Boolean networks and social network models. Informally, a random boolean network (hereafter, RBN) consists of a number of nodes, each node having some number (possibly zero) of input and output "wires" connecting it to other nodes in the network. If a wire connects nodes u and!v, and the wire is an output of u, the wire is an input of v (and vice versa). The set of nodes, along with the input and output wires, are assumed to form a connected network. At any time t, each node in the network is either *on* or *off*, and whether a node!u is on or off at time $t+1$ is given by a boolean function of the state of the nodes on u's input wires. More precisely, a RBN consists of a set of nodes $N = \{1,\ldots,n\}$, a directed graph $G = (N,E)$, and a sequence of boolean functions $\langle f_1,\ldots,f_n \rangle$, such that the arity of f_i equals the number of edges entering node i, plus one (since the state of a node always influences its future state).[7]

Denote the state of the network at time t by $S_t \in \{0,1\}^n$, and let S_t^i be the state of node!i at time t, which is simply the ith component of the n-tuple S_t. Let $N_i = \{i_1,\ldots,i_k\}$ denote the subset of nodes of N connected by an edge pointing to i. (N_i is the *neighborhood* of i, or, alternatively, the set of *input nodes* for i.) If the initial state S_0 of the network is given, one can calculate any future state of the network according to the following rule:

$$S_t^i = \begin{cases} f_i(S_{t-1}^{i_1},\ldots,S_{t-1}^{i_k}) & \text{if } t > 0 \text{ and } i_1,\ldots,i_k \in N_i \\ S_0^i & \text{otherwise.} \end{cases}$$

The boolean functions are often referred to as *transition rules* because they fix the state transitions for each node in the network. A random boolean network is essentially a generalized cellular automata, obtained by lifting the assumption that each node has the exact same transition rule and neighborhood.

A social network model consists of a population of agents, each of whom has a particular strategy; a network of relations defined on the population, specifying the set of possible interactions among members; and a set of update rules for each agent, specifying how she switches strategies at the end of each round of play. More precisely, let $N = \{1,\ldots,n\}$ denote the population of agents, and let $G = (N,U)$ be an undirected graph defined on N. Let $v(a)$ denote the set of agents from the population immediately connected to a in G (this is also known as the *neighborhood* of a). In each round, every agent!a

---

[7] Strictly speaking, since the set of edges entering node i is unordered, one needs to specify which argument of f_i corresponds to each node incident on one of i's input wires. As including this level of detail would increase the complexity of exposition with no corresponding gain in perspicuity, I leave it to the reader to fill in these details if desired.

interacts with her neighbors, playing some game using strategy σ_a. At the end of each round of play, an agent's score equals the sum of the payoffs from each pairwise interaction, and each agent has the opportunity to change her strategy. It is assumed that an agent will only elect to change her strategy if she has an appropriate incentive--e.g., if at least one of her neighbors received a higher score than she did. The specific way in which the agent! a chooses a new strategy corresponds to a particular learning rule L_a, which may or may not be the same for all agents.

Since the general proof of the representation theorem consists of showing how, given an arbitrary social network model with a finite number of individual strategies and deterministic update rules, one would go about constructing a RBN having the desired properties,[8] I will first present two examples showing how to construct a RBN whose dynamics and state space are isomorphic to that of a given social network model. I will then consider a case where one cannot construct a RBN whose state space and dynamics are isomorphic to the social network, but where one *can* construct a RBN in which a representation of the state space and dynamics of the social network is embedded. I then generalize this general procedure (thus sketching the proof). In these examples, I assume agents in the social network employ the following learning rule:

Imitate best neighbor, conservatively.
> An agent a switches strategies if and only if at least one agent in $\nu(a)$ received a score higher than a. The strategy adopted by a is the strategy closest to her present one, according to some suitable metric.[9]

As two of the examples are simple two-strategy versions of the prisoner's dilemma and the stag hunt, in these cases this learning rule collapses to: switch strategies if and only if every person receiving the maximum highest score in your neighborhood followed the opposite strategy.[10] Note, though, that nothing regarding the statement or proof of the representation theorem depends upon this *particular* assumption as to the learning rule--what is crucial is only that there be a nondeterministic update rule specifying how future strategies are selected.

2.1 The Prisoner's Dilemma. Consider a simple network model of the prisoner's dilemma where seven agents play the prisoner's dilemma on a ring and all agents use the "imitate best neighbor, conservatively" update rule as described above.

---

[8] The "desired properties" being that either (1) one can find a bijection between the state spaces of the two networks which preserves the dynamics, or that (2) one can find an injection from the state space of the social network into the state space of the RBN which preserves the dynamical relations of the social network.

[9] This update rule differs slightly from more commonly used ones to insure that the update rule be deterministic.

[10] This can intuitively be thought of as a highly risk-adverse learning rule where one adopts a new strategy only in the face of strong evidence that one's current strategy is inadequate.
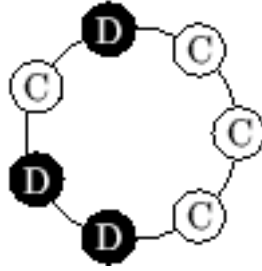
Figure 1: Prisoner's dilemma played on a seven-person ring.

Take the payoffs for this prisoner's dilemma to be those in table![1].

Table 1: Payoff matrix for the prisoner's dilemma.

!	Cooperate	Defect
Cooperate	3	1
Defect	4	2

This social network, then, has the population $N = \{1,\ldots,7\}$, the graph $G = (N,U)$ as illustrated in figure![1], and a common update rule L for all individuals in this population. Constructing a random boolean network equivalent to this social network requires specifying a set of nodes N', a graph $G' = (N',E)$, and a sequence of transition rules f_i for each node in N'.

To begin, since there are only two possible strategies in the original social network, Cooperate and Defect, we may simply let N' be a set of seven nodes, where each node represents one agent in the original social network, and the strategies Cooperate and Defect are represented by having the node be *on* or *off*, respectively. (In the third example I will present a case where more complicated representations are required, but such is unnecessary here.)

We now need to construct a graph G' specifying the inputs for each node in N'. Since these inputs are the only source of new ``information'' that a node may refer to when changing state, the graph G' needs to represent all of the relevant connections between agents in the social network. If whether an agent a adopts a strategy in the next round of play depends on an agent b, G' must include a connection between the node representing a and the node representing b.

One minor complication occurs because the edges in a social network are undirected whereas the edges in a RBN are directed. The reason for this is simply that in a social network the relations of interaction and updating are symmetric: if A interacts with B, then B interacts with A; and if A inspects B's strategy when updating, then B inspects A's strategy when updating. Though RBNs do not require such symmetry, it is easily represented. For each undirected edge $\{A,B\} \in G$, we need to introduce a pair of directed edges in G' between the node representing A and the node representing B.

It is important to realize that the dependency relations between agents in a social network extend beyond the immediate connections specified in G. Given a ring configuration as in figure![1], consider the immediate and surrounding neighbors of a typical agent! A, as illustrated in the ring segment of figure![2].


Figure 2: Segment of the seven-person prisoner's dilemma ring.

Although the score of A depends only on the *strategies* held by ② and ③, whether A changes strategies according to the rule L depends not just on A's score, but on the score of players ② and ③ as well. In turn, the scores of ② and ③ depend on the strategies held by ① and ④. That is, the strategy A adopts at the end of the round depends not only on the strategies held by immediate neighbors ② and ③, but *implicitly* on the strategies held by the "neighbors once removed," ① and ④. Social networks with imitative learning rules generally have an implicit radius of influence for each agent reaching one step beyond the immediate connections in the interaction graph. Consequently, additional edges representing these dependencies must be included in G'. Figure 3 illustrates the additional edges which must be added to the network in order to take the implicit radius of influence into consideration. On account of the symmetry of social networks, each undirected edge will need to be replaced by a pair of directed edges, which are omitted in figure 3 for reasons of clarity.
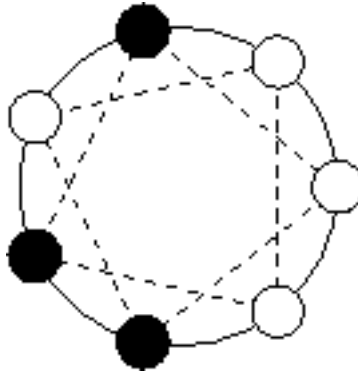
Figure 3: The graph G' for the constructed RBN. Solid edges correspond to edges present in the original social network. Dashed edges represent dependencies between agents in the original social network who did not have explicit edges connecting them. All edges are shown as undirected for clarity, but will need to be replaced by a pair of directed edges to obtain the real graph G'.
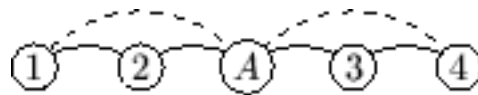
Figure 4: Dependencies for node A in the constructed RBN.

At this point, all of the relevant dependencies have been incorporated in G'. The remaining task is to construct the transition rules for the random boolean network. The symmetries present in the original social network mean that each node in the RBN will have the same transition rule. Before we can construct this transition rule, we must first impose an ordering on the input wires for each node in the network. Using the illustration of figure 4 as a reference, let us represent the state of a node A and her neighbors ①, ②, ③, and ④ by a five-bit string $B_1 B_2 B_A B_3 B_4$ where bit B_x is 0 if x is off and 1 if x is on. Thus the state where ① is on, ② is off, A is off, ③ is on, and ④ is on will be represented by 10011. In this state, since A defects on one cooperator and one defector (③ and ②, respectively), A's score will be six. Since ② defects on one cooperator and one defector (① and A, respectively), ②'s score is six as well. ③'s score is only four since ③ cooperates with one defector and one cooperator (A and ④, respectively). Thus A will continue to defect in the next generation. We may represent this transition rule for A as $10011 \rightarrow 0$, indicating that, when given inputs 10011, the future state of node A will be 0.

Figure 2 lists the remainder of the transition rules for the node A of figure 4; these transition rules are used by all of the other nodes in the RBN as well. At this point we are finished: we have a RBN equivalent to the original social network, and we know that the RBN is equivalent by the process of construction used to obtain it. If the future strategy of A depends, in some way, upon the strategy held by B, there is an input wire connecting the node representing B to the node representing A in the RBN. Additionally, the transition rule assigned to the node representing A has taken the precise nature of these dependencies into account.

Table 2: Transition rule representing the *imitate best neighbor, conservatively* update rule for an agent playing the prisoner's dilemma on a ring.

!11111 →1	10111 →0	01111 →1	00111 →1
!11110 →1	10110 →0	01110 →1	00110 →0
!11101 →0	10101 →0	01101 →0	00101 →0
!11100 →1	10100 →0	01100 →0	00100 →0
!11011 →0	10011 →0	01011 →0	00011 →0
!11010 →0	10010 →0	01010 →0	00010 →0
!11001 →0	10001 →0	01001 →0	00001 →0
!11000 →0	10000 →0	01000 →0	00000 →0

2.2 The Stag Hunt. Consider a social network model in which seven people situated on a ring play the stag hunt with their neighbors, and assume that the specific form of the stag hunt be that of table!3.

Table 3: Payoff matrix for the stag hunt

!	Hunt stag	Hunt rabbit
Hunt stag	3	0
Hunt rabbit	1	2

Constructing a RBN equivalent to this social network would proceed exactly as above, with only two exceptions. First, the *on* and *off* states of nodes in the RBN represent the strategies "Hunt stag" and "Hunt rabbit" instead of "Cooperate" and "Defect." Second, the particular transition rules assigned to each node will differ from those for the prisoner's dilemma. Table!4 lists the transitions rule for each node in the resulting RBN. Interestingly, for all of the strategic differences between the prisoner's dilemma and the stag hunt, note the relatively little difference between the two transition rules.

Table 4: Transition rule representing the *imitate best neighbor, conservatively* update rule for an agent playing the stag hunt on a ring.

|			
!11111 → 1	10111 → 1	01111 → 1	00111 → 1
!11110 → 1	10110 → 1	01110 → 1	00110 → 1
!11101 → 1	10101 → 0	01101 → 0	00101 → 0
!11100 → 1	10100 → 0	01100 → 0	00100 → 0
!11011 → 0	10011 → 0	01011 → 0	00011 → 0
!11010 → 0	10010 → 0	01010 → 0	00010 → 0
!11001 → 0	10001 → 0	01001 → 0	00001 → 0
!11000 → 0	10000 → 0	01000 → 0	00000 → 0

2.3 The Nash bargaining game. Consider, as before, the Nash bargaining game played on a seven-person ring, with similar assumptions as above.[11] Suppose that the good is divided into seven equal slices and that each player's strategy is restricted to an integral number of slices. This particular social network poses a greater challenge in constructing an equivalent RBN than the previously considered cases because there are more than two possible strategies in the game.

The solution employs the fact that we can use *several* nodes in the RBN to represent a *single* player, using the aggregate state of those nodes to represent the player's strategy. Figure!5 illustrates one way this may be done.

---

[11] In the Nash bargaining game, players have strategies indicating how much of a divisible good they want. Without being able to communicate with each other, both players submit their strategy to a referee. The referee compares the total amount of the good available with the sum of the players' requests. If the sum of their requests does not exceed the total amount of the good available, each person receives what he asked for; otherwise, neither player receives anything. See Skyrms (1996) for a discussion of this game.
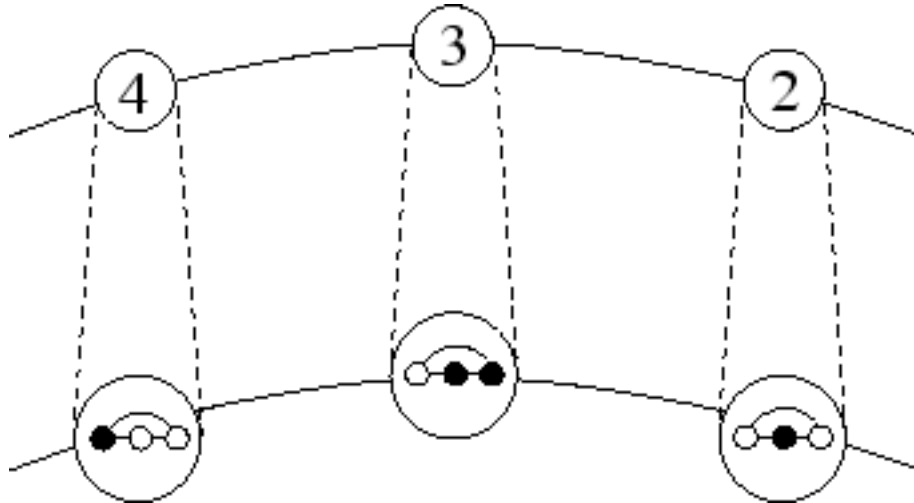
Figure 5: Encoding complex strategies using several nodes in a RBN.

Since there are eight possible strategies[12], we may represent a player's strategy by a block of three nodes, using the "on" and "off" states as bits to denote the strategy in binary.

The one trade-off with representing complex strategies in this way appears when we proceed to construct the network G' of inputs to each node. In the previous examples, the network of the RBN simply needed to mirror the network of the original model, with some extra edges added to capture some subtle dependencies created by the nature of the learning rule. Here, though, the wiring of the RBN will be considerably more complex. Each node used to represent a player A (and there will be more than one) will have to be connected to *every* node used to represent those players that A's future strategy depends on. Figure!6 illustrates all of the input wires which must be introduced in order to capture the dependency of the middle player's strategy on his *immediate* neighbors. Additional input wires would have to be included to capture the implicit dependency of the middle player's strategy on his neighbor's neighbors, as discussed above.
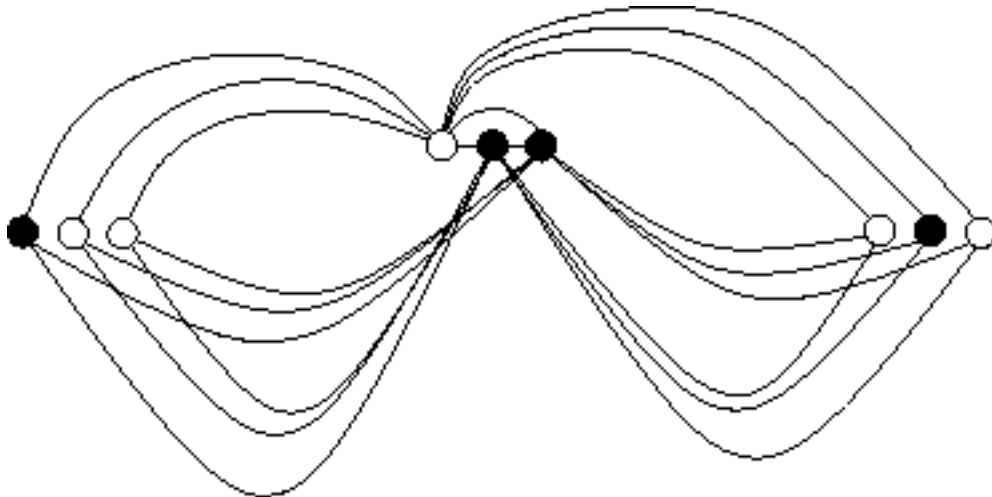
Figure 6: Complete input wiring diagram for the center block of three nodes.

---

[12] Ask for no cake, ask for one slice of the cake,..., ask for seven slices of the cake.

Using multiple nodes to represent a single player allows arbitrarily complex strategies in a social network to be represented in a RBN. The primary drawback, aside from the rapid increase in the complexity of the wiring scheme, is that in some cases it is not possible to construct a RBN whose state space is isomorphic to the original social network. For example, consider the Nash bargaining game where the good to be divided is sliced into 10 equal pieces. If we were to use the method of representation described above, we would have to use at least four nodes to represent each player, since fewer than four nodes would be unable to represent all of the possible strategies. Yet using four nodes to represent each player introduces spurious states into the RBN that do not naturally correspond to any possible state in the original social network. Although inconvenient, the solution is to assign an arbitrary interpretation to these spurious states, using this interpretation consistently when constructing the transition rules. In this case, the state space and dynamics of the RBN will contain an embedded representation of the state space and dynamics of the social network.

3. The Construction Method. Here is a summary of the construction method employed in the previous section.

Given: A social network consisting of a population N, a graph $G = (N, U)$, a game Γ, and for each agent $a \in N$, a learning rule L_a and a set of strategies S_a.

1. Let $|S_i|$ be the number of strategies available to player $i \in N$. One must use at least $\left\lceil \log_2 \left(|S_i| \right) \right\rceil$ nodes in the RBN to represent player! i. If the update rule L_i is not memoryless[13], one additional node will need to be used for each bit of memory required by the update rule. Let $\boxed{i}$ denote the set of all nodes used in the RBN to represent player i.
2. If j is a neighbor of i, add input wires from each node in $\boxed{i}$ to each node in $\boxed{j}$, and vice versa.
3. If the future strategy of player i implicitly depends on another player k who is not a neighbor of i (as discussed in section 2.1), add input wires from each node in $\boxed{k}$ to each node in $\boxed{i}$.
4. Note that, by construction, each node in $\boxed{i}$ is connected by input wires to every node used to represent a player in the original social network that i's future strategy may depend on. Therefore, it is possible to define the transition rules for each node in $\boxed{i}$ such that the next state of $\boxed{i}$ represents the next strategy adopted by i.

4. Conclusion. By using the method described above, one may construct a random boolean network representing a social network. In many cases, the state space and dynamics of the resulting RBN will be isomorphic to those of the original social network; in the remainder of the cases, there is an embedding of the state space of the original social network into the RBN which preserves the dynamical relations between states. It is hoped that this result will contribute towards our understanding of RBNs and social network models of cultural evolution.

This result may do so in three different ways. First, since random boolean networks have a simpler structure than social networks, it should be easier to prove theorems regarding the long-term limit behavior of random boolean networks than social networks. Since social networks can be viewed as a subclass of RBNs, any theorems proved for all random boolean networks automatically apply to social networks.

Second, Wuensche (1994) showed how to generate computationally all the basins of attraction for RBNs of reasonably small sizes.[14] The ability to catalog all the basins of attraction for social network models should

---

[13] For example, in the prisoner's dilemma the well-known strategy tit-for-tat is not memoryless.

[14] Wuensche and Lesser (1992) describes an algorithm capable of calculating the basins of attraction for cellular automata, which suffices for all of the figures contained in appendix A. In general, though, one will

aid in formulating theorems about the basins of attraction for particular social networks. Given that one needs a reasonable amount of data in order to form a conjecture, the ability to catalog basins of attraction for particular social networks allows one to generate, relatively rapidly, data regarding basins of attraction for social networks which was previously difficult to obtain. The appendix to this paper illustrates the complete basins of attraction (up to rotational equivalence) for three different social network models.

Finally, the technique developed by Wuensche and Lesser for displaying the basins of attraction for RBNs raises several interesting questions for future research. For example, how does adding new edges to a social network affect the basin of attraction? Does gradually increasing the number of edges between agents in the population gradually change the shape of the basin of attraction? Is there a critical point at which something akin to a "phase transition" occurs, where a slight change in the number of edges in the network leads to a dramatic change in the basin of attraction? What difference does it make if all agents in the population employ the same learning rule? Since it is possible to automatically map the basins of attraction for RBNs, we can now automatically map the basins of attraction for social networks. Since changes in the basin of attraction correspond to immediately recognizable *visible* differences in the map (i.e., compare tables!7 and!8, showing the difference in the basin of attraction between the stag hunt and prisoner's dilemma played on a 8-person ring), we can, quite literally, *see* how changing a social network model changes its basins of attraction. It is hoped that this method of analyzing social networks will lead to the identification of regularities between the nature of social network models and their basins of attraction, and perhaps more.

Appendix A. Basins of attraction. The following figures illustrate complete basins of attraction for random boolean networks equivalent to a social network in which agents play the prisoner's dilemma or the stag hunt. In these figures, a row of squares represents a state of the RBN. Darkly shaded squares represent nodes in the "off" state, lightly shaded squares represent nodes in the "on" state. The interpretation of these nodes depends on the game: for the prisoner's dilemma, "on" and "off" should be read as "cooperate" and "defect," respectively; for the stag hunt, "on" and "off" should be read as "hunt stag" and "hunt rabbit," respectively. In all cases, the topology of the original social network was a ring. The line of squares rearranges the nodes from a ring-shape into a line shape, so the leftmost and rightmost squares are, strictly speaking, adjacent.

The center of each diagram represents a fixed point of the dynamics (for these RBNs, there are no cycles). Given a particular state S, its successor state S' is located closer towards the center of the diagram; a single evolutionary trajectory thus begins from the initial state and moves towards the center of the figure, continuing to loop once it has reached the fixed point.

In listing the basins of attraction below, I have suppressed basins of attraction rotationally equivalent to those displayed. Notice that in the prisoner's dilemma played on a five-person ring, the state with two adjacent defectors is stable. Because there are five such attracting states[15] equivalent under rotation, I omit these since they do not count as a *qualitatively* different basin of attraction. Hence the number of states included in each figure does not equal the total number of possible states for the RBN.

---

need to use the algorithm described in Wuensche (1994), which is capable of handling arbitrary random boolean networks.

[15] 1 and 2 defect, all others cooperate; 2 and 3 defect, all others cooperate; 3 and 4 defect, all others cooperate; 4 and 5 defect, all others cooperate; and 5 and 1 defect, all others cooperate.
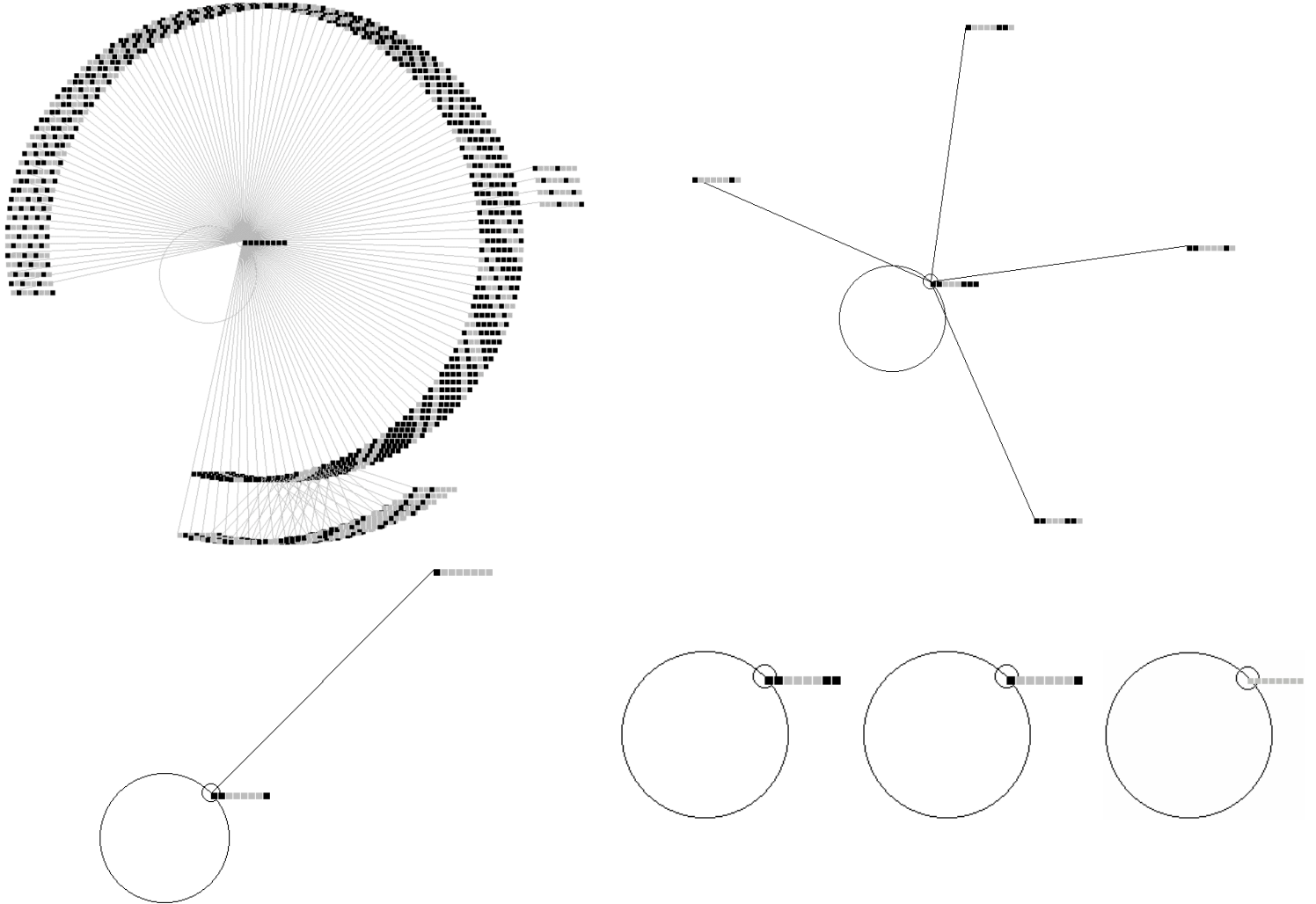
Table 5: Basins of attraction for 5-person prisoner's dilemma played on a ring.
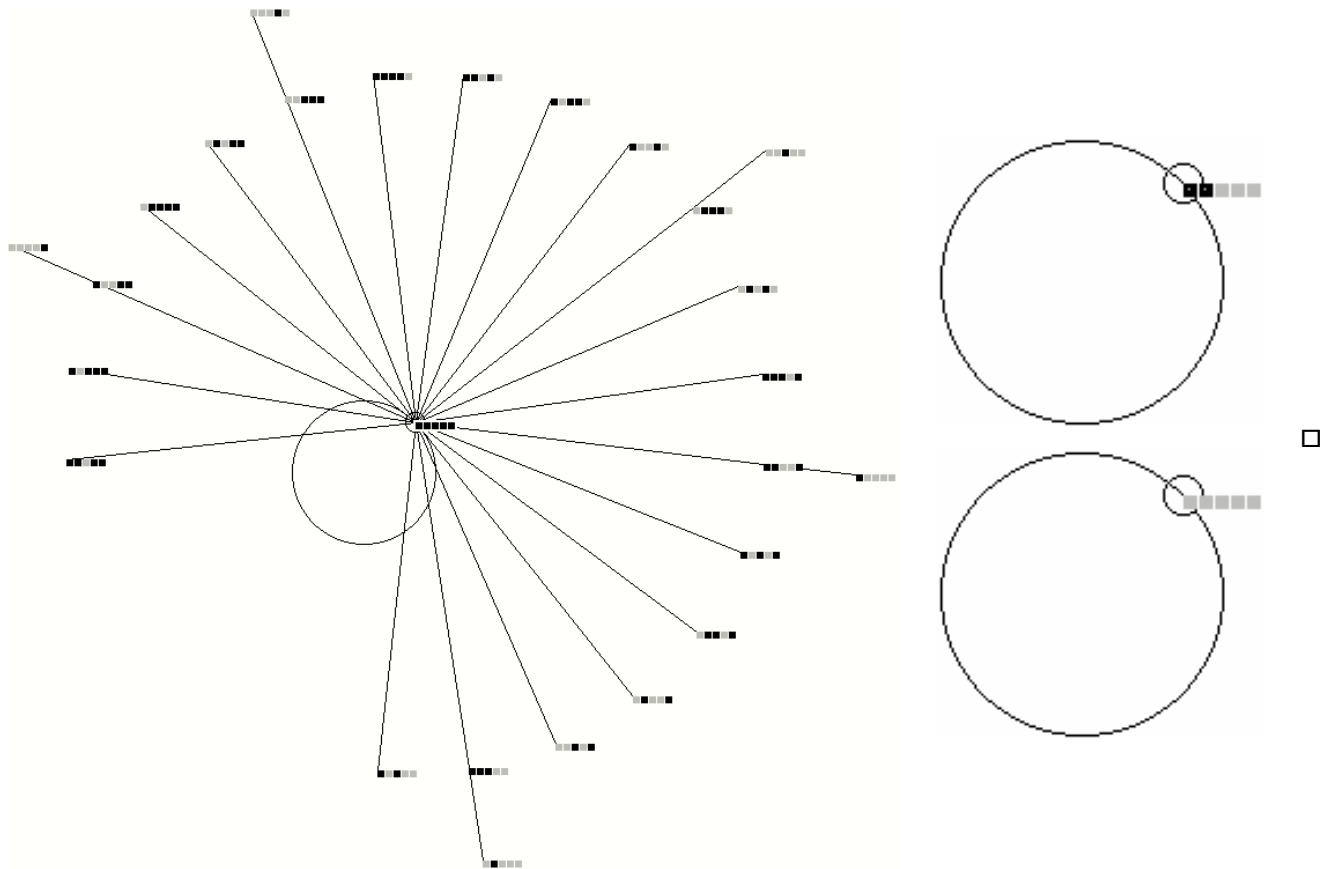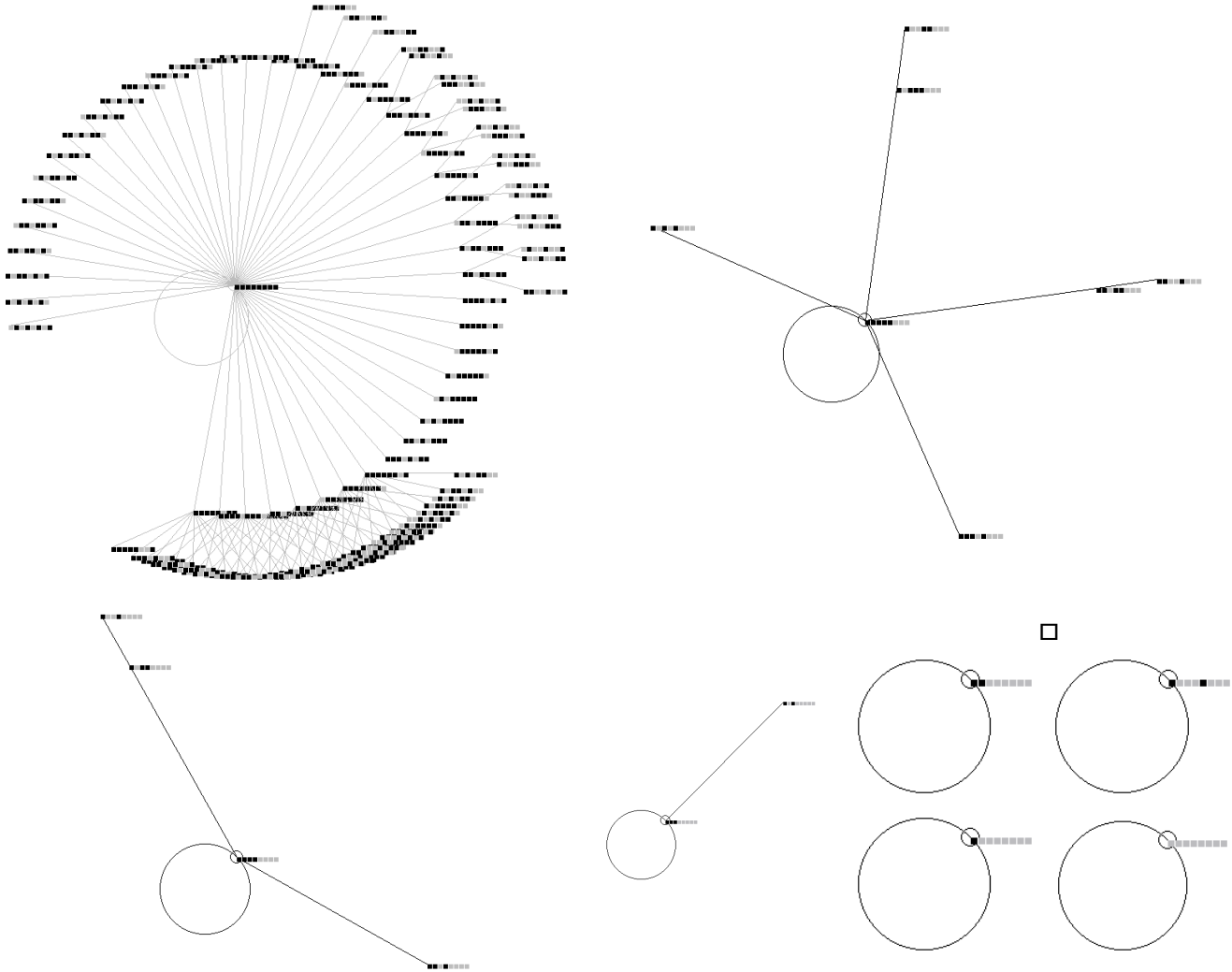
Table 6: Basin of attraction for 6-person prisoner's dilemma played on a ring
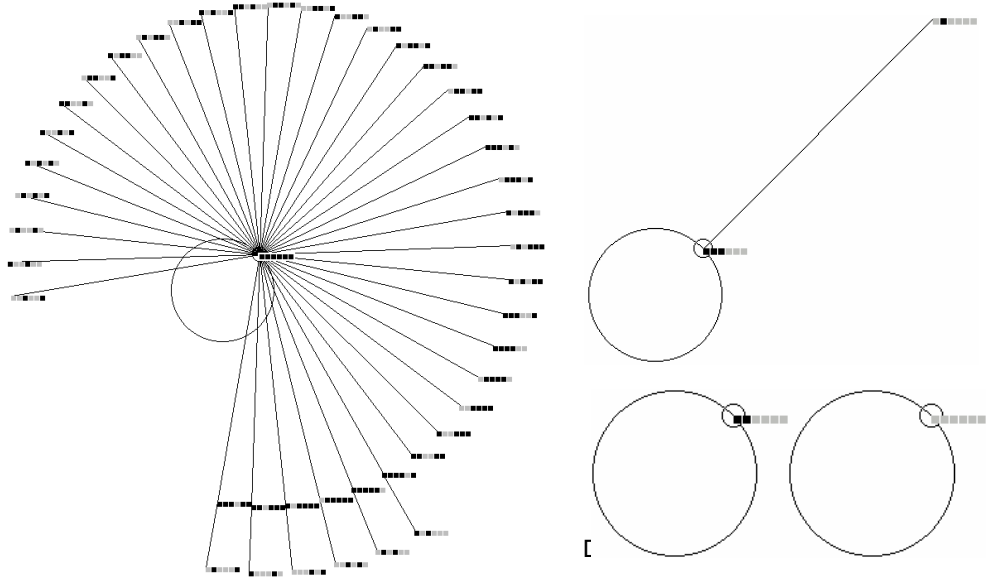
Table 7: Basin of attraction for 8-person prisoner's dilemma played on a ring.

Table 8: Basin of attraction for 8-person stag hunt played on a ring

Bibliography

Alexander, J.!McKenzie (2000). "Evolutionary Explanations of Distributive Justice." *Philosophy of Science*.

Alexander, Jason and Brian Skyrms (1999). "Bargaining with Neighbors: Is Justice Contagious?" *Journal of Philosophy* 96, 11: 588-598.

Bögers, Tilman and Rajiv Sarin (1993). "Learning Through Reinforcement and Replicator Dynamics." Unpublished manuscript (revised November 1994).

D'Arms, Justin, Robert Batterman, and Krzyzstof Górny (1998). "Game Theoretic Explanations and the Evolution of Justice." *Philosophy of Science* 65: 76-102.

Lachapelle, Jean (2000). "Cultural Evolution, Reductionism in the Social Sciences, and Explanatory Pluralism." *Philosophy of the Social Sciences* 30, 3: 331-361.

Nowak, Martin A. and Robert M. May (1992). "Evolutionary games and spatial chaos." *Nature* 359: 826-829.

– (1993). "The Spatial Dilemmas of Evolution." *International Journal of Bifurcation and Chaos* 3, 1: 35-78.

Skyrms, Brian (1994). "Sex and Justice." *Journal of Philosophy* 91: 305-320.

– (1996). *Evolution of the Social Contract*. Cambridge University Press.

Wuensche, A. and M. J. Lesser (1992). *The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata*, volume 1 of *Santa Fe Institute Studies in the Sciences of Complexity*. Addison-Wesley.

Wuensche, A. (1994). "The Ghost in the Machine: Basins of Attraction of Random Boolean Networks." In *Artifical Life III*, ed. C. G. Langton, Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA.