

Foraging for Trust: Exploring Rationality and the Stag Hunt Game

Steven O. Kimbrough¹

University of Pennsylvania, Philadelphia, PA 19104, USA

kimbrough@wharton.upenn.edu,

WWW home page: <http://opim-sun.wharton.upenn.edu/~sok/>

Abstract. Trust presents a number of problems and paradoxes, because existing theory is not fully adequate for understanding why there is so much of it, why it occurs, and so forth. These problems and paradoxes of trust are vitally important, for trust is thought to be the essential glue that holds societies together. This paper explores the generation of trust with two simple, but very different models, focusing on repeated play of the Stag Hunt game. A gridscape model examines creation of trust among cognitively basic simple agents. A Markov model examines play between two somewhat more sophisticated agents. In both models, trust emerges robustly. Lessons are extracted from these findings which point to a new way of conceiving rationality, a way that is broadly applicable and can inform future investigations of trust.

1 Introduction

Can trust arise spontaneously—by an invisible hand as it were—among strategically interacting individuals? If so, under what conditions will it arise? When will it be stable and when will it not be stable? What interventions might be effective in promoting or undermining stability? If trust is established, under what conditions will it be destroyed? What are the rôles of social structure, game structure, and cognition in establishing or disestablishing trust? These questions belong to a much longer list of important and challenging issues that the problem of trust presents. Answering them fully and adequately constitutes a research program to challenge a community over a period of many years.

I aim in this paper to contribute in two ways to that program. In the end, although progress will be made, more work will have been added. First, I shall present findings, focusing on the Stag Hunt game, that bear more or less directly on at least some of these questions. I shall focus on the Stag Hunt game for several reasons. The game does capture well and succinctly certain aspects of the problem, the dilemma, of trust. It has the happy virtue of not being the (over-worked but still worthwhile) game of Prisoner's Dilemma. Also, it has fruitfully received new attention of late (e.g., [Sky04]), so that what I will add here will, I hope, enrich a topic that is very much in play.

The second way in which I aim to contribute to the trust research program is more indirect. Trust is a problem or a puzzle, even paradox, in part because there

seems to be more of it naturally occurring than can be explained by received theory (classical game theory). I shall say little by way of documenting this claim because space is limited and I take it that the claim is widely accepted. (Chapter 3, “Mutual Aid,” of [Sky96] is a discussion of how and why nature is *not* “red in tooth and claw.” Also behavioral game theory, reviewed in [Cam03] amply documents the imperfect fit between theory and observation in this domain. Those with a taste for blunt talk might consult [Gin00].) Instead, I hope to say something about how the puzzle of trust might be investigated. I will submit that the puzzle of trust arises, at least in part, because of a presupposed account of agent rationality. This account goes by different names, among them *expected utility theory* and *rational choice theory*. I want to propose, in outline form, a very different approach to conceiving of rationality. By way of articulating this general approach, which I shall call a *theory of exploring rationality*, I shall present a model of agent behavior which, in the context of a Stag Hunt game (as well as other games), explains and predicts the presence of trust.

2 Stag Hunt and a Framing of the Program

The Stag Hunt game (also known as the Assurance game [Gam05a]) gets its name from a passage in Jean Jacques Rousseau’s *A Discourse on the Origin of Inequality*, originally published in 1755.

Was a deer to be taken? Every one saw that to succeed he must faithfully stand to his post; but suppose a hare to have slipped by within reach of any one of them, it is not to be doubted but he pursued it without scruple, and when he had seized his prey never reproached himself with having made his companions miss theirs. [Rou04, Second Part]

Here is a representative summary of the Stag Hunt game.

The French philosopher, Jean Jacques Rousseau, presented the following situation. Two hunters can either jointly hunt a stag (an adult deer and rather large meal) or individually hunt a rabbit (tasty, but substantially less filling). Hunting stags is quite challenging and requires mutual cooperation. If either hunts a stag alone, the chance of success is minimal. Hunting stags is most beneficial for society but requires a lot of trust among its members. [Gam05b]

This account may be abstracted to a game in strategic form. Figure 1 on the left presents the Stag Hunt game with payoffs that are representative in the literature.¹ Let us call this our *reference game*. On the right of figure 1 we find the Stag Hunt game presented in a generic form. Authors differ in minor ways. Often, but not always, the game is assumed to be symmetric, in which

¹ For example, although it is not called the Stag Hunt, the game with these payoffs is discussed at length in [RGG76], where it is simply referred to as game #61.

case $R=R'$, $T=T'$, $P=P'$, and $S=S'$. I will assume symmetry. It is essential that $R>T>P\geq S$.²

	Hunt stag (S)	Chase hare (H)
Hunt stag (S)	4	3
Chase hare (H)	1	2

	Hunt stag (S)	Chase hare (H)
Hunt stag (S)	R'	T'
Chase hare (H)	S'	P'

Fig. 1. Stag Hunt (aka: Assurance game)

Thus formalized, the Stag Hunt game offers its players a difficult dilemma, in spite of the fact that their interests coincide. Each does best if both hunt stag (S,S). Assuming, however, that the game is played once and that the players lack any means of coming to or enforcing a bargain,³ each player will find it tempting to “play it safe” and hunt hare. If both do so, (H,H), the players get 2 each in our reference game, instead of 4 each by playing (S,S). Both of these outcomes—(S,S) and (H,H)—are Nash equilibria. Only (S,S), however, is Pareto optimal. There is a third Nash equilibrium for Stag Hunt: each player hunts stag with probability $\frac{P-S}{(R+P)-(T+S)}$. For our reference game, this amounts to a probability of $\frac{1}{2}$ for hunting stag (and $\frac{1}{2}$ for hunting hare). At the mixed equilibrium each player can expect a return of $2\frac{1}{2}$. Notice that if, for example, the row player hunts hare with probability 1, and the column player plays the mixed strategy, then the row player’s expected return is $2\frac{1}{2}$, but the column player’s expected return is $1\frac{1}{2}$. Uniquely, the safe thing to do is to hunt hare, since it guarantees at least 2. Hunting hare is thus said to be *risk dominant* and according to many game theorists (H,H) would be the predicted equilibrium outcome.⁴

We can use the Stag Hunt game as a model for investigation of trust. A player hunting stag trusts the counter-player to do likewise. Conversely, a player hunting hare lacks trust in the counter-player. Deciding not to risk the worst outcome (S) is to decide not to trust the other player. Conversely, if trust exists then risk can be taken. There is, of course, very much more to the subject of trust than can be captured in the Stag Hunt game. Still, something is captured. Let us see what we can learn about it.

Before going further it is worth asking whether Rousseau has anything else to say on the matter to hand. Typically in the game theory literature nothing else in this *Discourse* or even in other of Rousseau’s writings is quoted. As is

² Usually, and here, $P>S$. Some authors allow $T\geq P$ with $P>S$. None of this matters a great deal for the matters to hand.

³ In the jargon of game theory, this is a *noncooperative* game.

⁴ I’m using the term risk dominant in a general way, since adverting to its precise meaning would divert us. See [HS88] for the precise meaning.

well known, Rousseau wrote in praise of the state of nature, holding that people were free of war and other ills of society, and on the whole were happier. That needn't concern us here. What is worth noting is that Rousseau proceeds by conjecturing (his word, above) a series of steps through which man moved from a state of nature to the present state of society. Rousseau is vague on what drives the process. The view he seems to hold is that once the equilibrium state of nature was broken, one thing led to another until the present. Problems arose and were solved, one after the other, carrying humanity to its modern condition. He laments the outcome, but sees the process as more or less inevitable. With this context in mind, the passage immediately before the oft-quoted origin of the Stag Hunt game puts a new light on Rousseau's meaning. He is describing a stage in the passage from the state of nature to civil society.

Such was the manner in which men might have insensibly acquired some gross idea of their mutual engagements and the advantage of fulfilling them, but this only as far as their present and sensible interest required; for as to foresight they were utter strangers to it, and far from troubling their heads about a distant futurity, they scarce thought of the day following. Was a deer to be taken? ... [Rou04, Second Part]

Rousseau is describing behavior of people not far removed from the state of nature. Language, for example, comes much later in his account. These people end up hunting hare because “as to foresight they were utter strangers to it, and far from troubling their heads about a distant futurity, they scarce thought of the day following.” If we *define* the game to be one-shot, then there is no future to worry about. Rousseau is right: if there is no future or if the players cannot recognize a future, then stag will roam unmolested. Rousseau is also right in presuming that in the later development of civil society the future matters, agents can recognize this, and much coordination and hunting of stag occurs. Rousseau is *not* agreeing with contemporary game theorists in positing hunting of hare as the most rational thing to do in the Stag Hunt game.

More generally, trust happens. Our question is to understand how and why. We assume that there is a future and we model this (initially) by repeating play of a basic game, called the *stage game*, here the Stag Hunt. Given repeated play of a stage game, there are two kinds of conditions of play that call out for investigation. The first is condition is the *social aspect* of play. We investigate a simple model of this in §3. The second condition might be called the *cognitive aspect* of play. How do learning and memory affect game results? We discuss this second aspect in §4.

3 The Gridscape: A Simple Society

We shall work with a very simple model of social aspects of strategic interaction, called the *gridscape*. The gridscape is a regular lattice—think of a checkerboard—which we will assume is two-dimensional and wraps around on itself (is technically speaking a torus). Agents or players occupy cells on the gridscape and each

has 8 neighbors. Figure 2(a) illustrates. Cell (3,2) has neighbors (2,1), (2,2), (2,3), (3,1), (3,3), (4,1), (4,2), (4,3).⁵ Every cell has eight neighbors. Thus, the neighbors of (1,1) are (6,6), (6,1), (6,2) (1,6), (1,2) (2,6), (2,1), and (2,2). With the gridscape as a basis it is now possible to undertake a variety of experiments. We'll confine ourselves to a simple one.

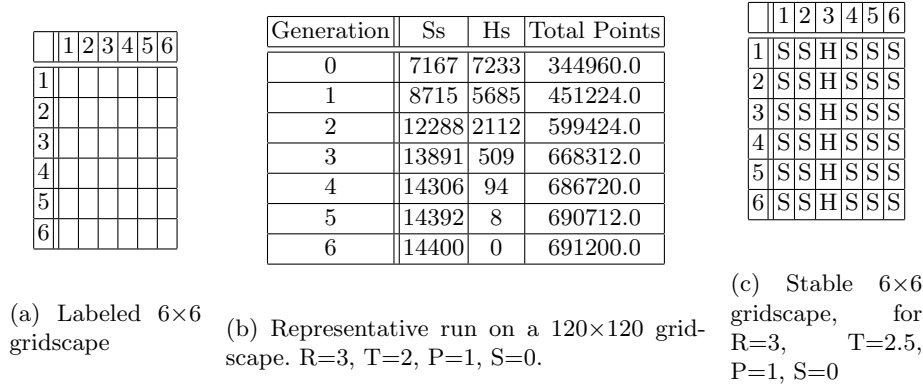


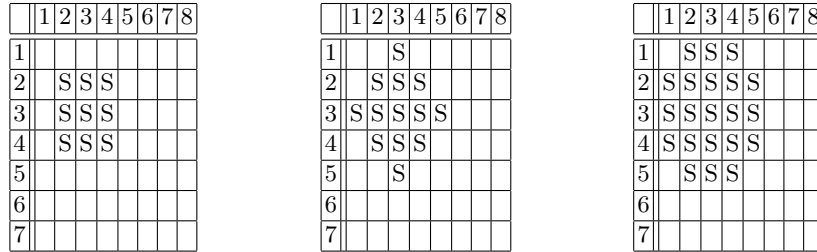
Fig. 2. Labeled 6×6 gridscape

Protocol 1 (Basic Gridscape Stag Hunt Protocol) *Each cell in the gridscape is initialized by placing on it either a hare hunter, H, or a stag hunter, S. H and S are the consideration set of policies of play for the experiment. Players using the H policy always hunt hare; and players using the S policy always hunt stag. Initialization is random in the sense that every cell has the same probability of being initialized with an H (or an S). After initialization, play proceeds by discrete generations. Initialization concludes generation 0. At the start of each subsequent generation each player (at a cell) plays each of its 8 neighbors using its policy in use from the consideration set. The player records the total return it gets from playing the 8 neighbors. After all players have played their neighbors, each player updates its policy in use. A player changes its policy in use if and only if one of its neighbors is using the counter-policy and has achieved a strictly higher total return in the current generation than has the player or any of its neighbors with the same policy in use achieved. (This is called the IMITATE THE BEST NEIGHBOR policy.) Policy updating completes the generation. Play continues until a maximum number of generations is completed.*

The table in figure 2(b) shows representative data when the gridscape is seeded randomly (50:50) with stag hunters (Ss) and hare hunters (Hs) and protocol 1 is

⁵ This is called the *Moore neighborhood*. The *von Neumann neighborhood*, consisting of the four neighbors directly above, below, to the left, and to the right, is also widely studied. The results we report here are not sensitive to which of the two neighborhood definitions is in force.

executed. Here, after 5 generations the Hs go extinct. The stag hunters conquer the board. This is the usual case, but it is not inevitable. To see why, consider the examples in figure 3, which shows a 3×3 block of stag hungers (Ss). The remaining cells are blank and for the purposes of the discussion may be filled in as needed.



(a) Generation x (b) Generation x+1 (b) Generation x+2

Fig. 3. Growth of a block of stag hunters when $R=4$, $T=3$, $P=2$, and $S=1$

The first thing to notice is that in figure 3(a) at (3,3) we have an S that is completely surrounded by Ss. This cell obtains a total reward of $8 \times R = 8 \times 4 = 32$ in our reference example. It is impossible to do equally well or better, given the setup. In consequence, given the protocol, once a 3×3 block of Ss is created none of its members will ever change to H. This is true for all versions of the Stag Hunt game (under protocol 1). We say that a 3×3 block of Ss cannot be invaded. More generally, it is easy to see that no rectangular block of Ss larger than 3×3 can be invaded either. (Note further that blocks of hare hunters are not so advantaged. An internal hare hunter gets $8T < 8R$ in all Stag Hunt games.)

Can a block of stag hunters grow? Assume that in figure 3(a) the blank cells are all hare hunters. In general, the stag hunter at (2,3) will get a return of $5R + 3S$ which is $5 \times 4 + 3 \times 1 = 23$ in our reference game. The hare hunter at (1,3) will get $5P + 3T$ in general and $5 \times 2 + 3 \times 3 = 19$ in the reference game. And in general, so long as $5R + 3S > 5P + 3T$ a hare hunter in this position will convert to stag hunting. Note that not all Stag Hunt games will support this conversion. For example, $R=101$, $T=100$, $P=99$, and $S=0$ will not. Figures 3(b) and (c) show the next two generations and the pattern is clear: the stag hunters will drive the hare hunters to extinction.

Is conquest by stag hunters inevitable if a 3×3 block is created and the game rewards are sufficient for it to grow in a field consisting entirely of hare hunters? Equivalently, if $5(R - P) > 3(T - S)$ and a 3×3 block (or larger) of stag hunters forms, is it inevitable that hare hunters are driven to extinction? No it is not. For example, the configuration in figure 2(c) is stable for the given payoff values.

	$1 - \varepsilon$: TFT	ε : ALLD	Total
$1 - \varepsilon$: TFT	$(1 - \varepsilon)(l_e R)$ $(1 - \varepsilon)40$	$\varepsilon((l_e - 1)P + S)$ 9ε	$40 - 31\varepsilon$
ε : ALLD	$(1 - \varepsilon)(T + (l_e - 1)P)$ $12(1 - \varepsilon)$	$\varepsilon(l_e P)$ 10ε	$12 - 2\varepsilon$

Table 1. State 1: Payoffs to Row in Stag Hunt example when the system is in state 1 and $l_e = 10$.

There are many nice questions to ask and many interesting variations on the basic gridscape model for the Stag Hunt game. For present purposes, however, the following points are most on topic.

1. The gridscape Stag Hunt results described above are robust. What happens—whether Hs come to dominate or not, whether a stable mixture results and so on—depends on the game payoffs (What is it worth if both players hunt stag? etc.) and the initial configuration. For a broad range of cases, however, hunting stag will eventually dominate the society. Trust—in the form of hunting stag predominately—can arise spontaneously among strategically interacting individuals. In fact, this is far from an implausible outcome.
2. Trust in Stag Hunt on the gridscape is also robust in a second sense: once it is established in large part, it is not easily dislodged. Mutations occurring at a small rate in a field of stag hunters will create mostly isolated hare hunters who will convert to S in the next generation. If stag hunters do well without noise they will do reasonably well with it.
3. The gridscape model under protocol 1 evidences a clear social effect. What we may call the *shadow of society* appears and affects the outcome. The policies played by one’s neighbors may be, and usually are, influenced by policies played by players who are not one’s neighbors. Recalling figure 3(a), what happens to a hare hunter at (1,3) depends very much on the fact that the neighboring stag hunters are themselves adjacent to other stag hunters. Thus, while the hare hunter at (1,3) beats the stag hunter at (2,3), in the sense that it gets more points in the one-on-one play, the stag hunter at (2,3) in aggregate does better and it is the hare hunter who is converted.
4. The Prisoner’s Dilemma game arguably presents a trust dilemma in more extreme form than does the Stag Hunt. Can cooperators largely take over the gridscape? Yes, under certain, more restricted conditions. In Prisoner’s Dilemma, we require $T > R > P > S$ and $2R > T + S$. Further, we relabel the policies. Hunt Stag becomes Cooperate and Hunt Hare becomes Defect. On the gridscape, the 3×3 (and larger) block is key in the analysis. If, for example, we set $T = R + 1$, $P = 1$ and $S = 0$ in Prisoner’s Dilemma, then so long as $R > 4$, a 3×3 (and larger) block of cooperators will be able to expand in a field of defectors. Defectors may not be eliminated, but they may become very much minority constituents of the gridscape. Note further that if Prisoner’s Dilemma games are repeated (either infinitely with discounting or finitely), and the TIT FOR TAT policy replaces ALWAYS COOPERATE,

then the payoff structure will, under broad conditions, become a Stag Hunt (cf. [Cam03, chapter 7], [Sky04, chapter 1]).

5. The agents on the gridscape have an update policy, IMITATE THE BEST NEIGHBOR, which they use to choose policies for play from their consideration sets. Under this update policy, from protocol 1, agents change their policies of play if, after a round of play, one of their neighbors has used the alternative policy and gotten more points than either the player or one its neighbors playing with the player's policy of play. This is a reasonable update policy, but there are reasonable alternatives. Is there a sense in which it is optimal? Is some other update policy optimal? Is there a sense in which it is an ESS (evolutionarily stable strategy) [May82]?

These are all interesting questions, well worth investigation. Space is limited, however, and doing so would divert us from the main theme. A larger issue raised is this. Any reasonable update policy, including ours, may be interpreted as taking a stand on the uncertainty faced by the agent.⁶ Agents in games may be interpreted as seeking maximum return. That indeed is a presumption underlying the strategic framework. It is not a presumption that the agents are conscious or have intentions.

Regarding the stand on uncertainty, agents following our update policy are engaging in risky behavior whenever they opt for hunting stag. Yet when all agents so behave collective stag hunting robustly follows. Note the "Total Points" column in figure 2(b). In the first generation the agents collectively garnered 344960 points from the gridscape. If the agents had not updated their policies of play this is where it would stay. As we see, with the update policy used (IMITATE THE BEST NEIGHBOR) the agents collectively more than doubled their take from the gridscape. IMITATE THE BEST NEIGHBOR has this to commend itself: it does well when playing against itself. In Prisoner's Dilemma the same can be said for TIT FOR TAT. Notice as well that NEVER UPDATE, ALWAYS HUNT STAG does well against itself in Stag Hunt and ALWAYS COOPERATE does well against itself in Prisoner's Dilemma. Further, IMITATE THE BEST NEIGHBOR does well against NEVER UPDATE, ALWAYS HUNT STAG, as TIT FOR TAT does well against ALWAYS COOPERATE. Before pursuing these comments further, indeed as a means of doing so, let us turn to a more sophisticated model of learning by agents in games.

4 A Model for Exploring Rationality

We turn now to a more sophisticated model for learning by two agents engaged in repeated play of a stage game.⁷ The model—MLPS: Markov Learning in Policy Space—is highly stylized and has quite unrealistic assumptions. It is, however, valid as an approximation of realistic conditions; I ask for the reader's indulgence.

⁶ I am using *uncertainty* here in its technical sense, which contrasts with risk [LR57].

In a decision under risk we have an objectively supported probability distribution (or density) on the outcomes. Not so in a decision under uncertainty.

⁷ More extensive treatment of this model may be found in [Kim04b].

	ε : TFT	$1 - \varepsilon$: ALLD	Total
$1 - \varepsilon$: TFT	$\varepsilon(l_e R)$ $\varepsilon 40$	$(1 - \varepsilon)((l_e - 1)P + S)$ $9(1 - \varepsilon)$	$9 + 31\varepsilon$
ε : ALLD	$\varepsilon(T + (l_e - 1)P)$ 12ε	$(1 - \varepsilon)(l_e P)$ $10(1 - \varepsilon)$	$10 + 2\varepsilon$

Table 2. State 2: Payoffs to Row in Stag Hunt example when the system is in state 2 and $l_e = 10$.

	$1 - \varepsilon$: TFT	ε : ALLD	Total
ε : TFT	$(1 - \varepsilon)(l_e R)$ $(1 - \varepsilon)40$	$\varepsilon((l_e - 1)P + S)$ 9ε	$40 - 31\varepsilon$
$1 - \varepsilon$: ALLD	$(1 - \varepsilon)(T + (l_e - 1)P)$ $12(1 - \varepsilon)$	$\varepsilon(l_e P)$ 10ε	$12 - 2\varepsilon$

Table 3. State 3: Payoffs to Row in Stag Hunt example when the system is in state 3 and $l_e = 10$.

The key idea is that agents have a *consideration set of policies for play*, \mathcal{S} . The *supergame* consists of an indefinitely long sequence of *games*, each of which is a finite sequence of *rounds of play* of a stage game (e.g., Stag Hunt). Agents draw elements from their \mathcal{S} s and use them as *focal policies* for a period of time, or number of rounds of play, called a *game*. Each game is divided into n_e epochs of length l_e . Thus, the number of rounds of play in a game is $n_e l_e$. During an epoch an agent plays its current focal policy with probability $(1 - \varepsilon)$, and other policies from its consideration set the rest of the time, with probability ε .

At the end of each game, g_{t-1} , a player, p , picks a focal policy, f_p^t , from its consideration set, \mathcal{S} , for play in game g_t . The players use the *fitness-proportional* choice rule. Let $\widehat{V}(p, i, j, k)$ be the average value per round of play returned to player p for policy i , when p has focal policy j and $-p$ (the counter-player) has focal policy k . (Similarly, $V(p, i, j, k)$ is the value realized in a particular round of play.) Then

$$\Pr(f_p^{t+1} = i | f_p^t = j, f_{-p}^t = k) = \widehat{V}(p, i, j, k) / \sum_i \widehat{V}(p, i, j, k) \quad (1)$$

That is, the probability that a player chooses a policy for focus in the next game is the proportion of value it returned per round of play, compared to all the player's policies, during the previous game.

There is nothing egregiously unrealistic about these assumptions. The MLPS model strengthens them for the sake of mathematical tractability. Specifically, it is assumed that a mechanism is in place so that the two players are exactly coordinated. Each has its games begin and end at the same time (round of play in the sequence). Further, each game is neatly divided into epochs and the random choices are arranged so that each player's \widehat{V} values exactly realize their expected values. The upshot of this is that the \widehat{V} values seen by the players are constant, as are the underlying expected values. The resulting system is a

	ε : TFT	$1 - \varepsilon$: ALLD	Total
ε : TFT	$\varepsilon(l_e R)$ $\varepsilon 40$	$(1 - \varepsilon)((l_e - 1)P + S)$ $9(1 - \varepsilon)$	$9 + 31\varepsilon$
$1 - \varepsilon$: ALLD	$\varepsilon(T + (l_e - 1)P)$ 12ε	$(1 - \varepsilon)(l_e P)$ $10(1 - \varepsilon)$	$10 + 2\varepsilon$

Table 4. State 4: Payoffs to Row in Stag Hunt example when the system is in state 4 and $l_e = 10$.

stationary Markov process with states $\mathcal{S}_p \times \mathcal{S}_{-p}$ and the equilibrium distribution of states can be analytically determined.

To illustrate, assume the stage game is Stag Hunt with $R = 4, T = 3, P = 1$, and $S = 0$. Assume that each player has a consideration set of two policies of play: (1) TIT FOR TAT (TFT) in which the player begins (in the epoch) by hunting stag and subsequently mimics the behavior of the counter-player on the previous round of play, and (2) ALWAYS DEFECT (ALLD) in which the player always hunts hare. This system has four possible states: (1) both players in the game have TFT as their focal policy, (TFT, TFT), (2) player 1 (Row) has TFT as its focal policy and player 2 (Column) has ALLD as its focal policy, (TFT, ALLD), (3) (ALLD, TFT), and (4) (ALLD, ALLD). With $l_e = 10$ we get the payoffs for the various states as shown in tables 1–4.

Letting $\varepsilon = 0.1$, routine calculation leads to the transition matrix indicated in table 5.

	s(1)=(1,1)	s(2)=(1,2)	s(3)=(2,1)	s(4)=(2,2)
s(1)	$0.7577 \cdot 0.7577$ $= 0.5741$	$0.7577 \cdot 0.2423$ $= 0.1836$	$0.2423 \cdot 0.7577$ $= 0.1836$	$0.2423 \cdot 0.2423$ $= 0.0587$
s(2)	$0.5426 \cdot 0.7577$ $= 0.4111$	$0.5426 \cdot 0.2423$ $= 0.1315$	$0.4574 \cdot 0.7577$ $= 0.3466$	$0.4574 \cdot 0.2423$ $= 0.1108$
s(3)	$0.7577 \cdot 0.5426$ $= 0.4111$	$0.7577 \cdot 0.4574$ $= 0.3466$	$0.2423 \cdot 0.5426$ $= 0.1315$	$0.2423 \cdot 0.4574$ $= 0.1108$
s(4)	$0.5426 \cdot 0.5426$ $= 0.2944$	$0.5426 \cdot 0.4574$ $= 0.2482$	$0.4574 \cdot 0.5426$ $= 0.2482$	$0.4574 \cdot 0.4574$ $= 0.2092$

Table 5. Stag Hunt transition matrix data assuming fitness proportional policy selection by both players, based on previous Tables 1–4. Numeric example for $\varepsilon = 0.1 = \varepsilon_1 = \varepsilon_2$.

At convergence of the Markov process:

Pr(s(1))	Pr(s(2))	Pr(s(3))	Pr(s(4))
0.4779	0.2134	0.2134	0.0953

So 90%+ of the time at least one agent is playing TFT. Note the expected take for Row per epoch by state:

1. $(1 - \varepsilon)(40 - 31\varepsilon) + \varepsilon(12 - 2\varepsilon) = 34.39$
2. $(1 - \varepsilon)(9 + 31\varepsilon) + \varepsilon(10 + 2\varepsilon) = 11.91$
3. $\varepsilon(40 - 31\varepsilon) + (1 - \varepsilon)(12 - 2\varepsilon) = 14.31$
4. $\varepsilon(9 + 31\varepsilon) + (1 - \varepsilon)(10 + 2\varepsilon) = 10.39$

Further, in expectation, Row (and Column) gets $(0.4779 \ 0.2134 \ 0.2134 \ 0.0953) \cdot (34.39 \ 11.91 \ 14.31 \ 10.39)' = 23.02$ (per epoch of length $l_e = 10$, or 2.302 per round of play), much better than the 10.39 both would get if they played ALLD with ε -greedy exploration. Note that even the latter is larger than the return, 10 per epoch or 1 per round, of settling on the risk-dominant outcome of mutually hunting hare. There is a third, mixed, equilibrium of the one-shot Stag Hunt game. For this example it occurs at $((\frac{1}{2}S, \frac{1}{2}H), (\frac{1}{2}S, \frac{1}{2}H))$. At this equilibrium each player can expect a return of 2 from a round of play. Players playing under the MLPS regime learn that trust pays. A few points briefly before we turn to the larger lessons to be extracted from these examples.

1. Markov models converge rapidly and are quite robust. The results on display here hold up well across different parameter values (e.g., for ε). Further, relaxation of the mechanism of play so that agents get imperfect, but broadly accurate, estimates of the expected values of the V quantities will not produce grossly different results. We get a nonstationary Markov process, but in expectation it behaves as seen here.
2. The MLPS model also has attractive behavior for different kinds of games. Players in Prisoner's Dilemma games will learn a degree of cooperation and do much better than constant mutual defection. In games of pure conflict (constant sum games) the outcomes are close to those predicted by classical game theory. And in coordination games players go far by way of learning to coordinate. See [Kim04b] for details.
3. If we retain the core ideas of the MLPS model, but entirely relax the synchronization conditions imposed by the game mechanism, simulation studies produce results that qualitatively track the analytic results: the players learn to trust and more generally the players learn to approach Pareto optimal outcomes of the stage game [CLK04].

5 Discussion

Neither the gridscape model nor the MLPS model with protocol 1 nor the two together are in any way definitive on the emergence of trust in repeated play of Stag Hunt games. They tell us something: that trust can arise spontaneously among strategically interacting agents, that this can happen under a broad range of conditions, that it can be stable, and so on. The models and their discussion here leave many questions to be investigated and they raise for consideration many new questions. Much remains to be done, which I think is a positive result of presenting these models. I want now to make some remarks in outline by way of abstracting the results so far, with the aim of usefully framing the subject for further investigation.

LPS models: learning in policy space. Both the gridscape model and the MLPS model with protocol 1 are instances of a more general type of model, which I call an LPS (learning in policy space) model. In an LPS model an agent has a consideration set of policies or actions it can take, \mathcal{S} , and a learning or update, L/U , policy it employs in selecting which policies to play, or actions to take, at a given time. In the gridscape model, $\mathcal{S} = \{H, S\}$ for every player. In the MLPS model with protocol 1, $\mathcal{S} = \{\text{TFT}, \text{ALLD}\}$ for both players. In the gridscape model the L/U policy employed by all players was IMITATE THE BEST NEIGHBOR. In the MLPS model, the players used the fitness-proportional update rule, in the context of the mechanism described in the previous section.

LPS models categorize strategies. In classical game theory the players are conceived as having *strategies*, complete instructions for play, which they can be thought of as choosing before the (super)game starts. The possible strategy choices constitute what we call the consideration set, \mathcal{S} . Because strategies are picked *ex ante* there is no learning, although the strategies can be conditioned on play and can mimic any learning process. The agents employ what we might call the *null learning/update rule*, L/U_\emptyset . In an LPS model with a non-null L/U policy, the consideration set of policies of play does not include all possible strategies in the game. Policies in \mathcal{S} are tried sequentially and played for a limited amount of time, then evaluated and put into competition with other members of \mathcal{S} . The L/U policy constitutes the rules for comparison, competition and choice. The total number of possible strategies is not affected by imposition of the LPS framework, but the strategies are implicitly categorized and the agents choose among them during the course of play (instead of *ex ante*). The consideration set of *strategies* used by an agent is implicit in its consideration set of policies, its L/U policy, the structure of the game, and the play by the counter-players. Thus, LPS models subsume standard game-theoretic models. A *proper* LPS model, however, has a non-null L/U policy. Normally, when I speak of an LPS model I shall be referring to a proper LPS model.

Folk Theorem undercuts. According to the Folk Theorem,⁸ nearly any set of outcomes in an indefinitely repeated game can be supported by some Nash equilibrium. In consequence, the Nash equilibrium becomes essentially worthless as a predictive or even explanatory tool, in these contexts. The problems of trust arise against this backdrop and against the following point.

Refinements unsatisfying. Refinements to the classical theory, aimed at selecting a subset of the Nash equilibria in predicting outcomes, have been less than fully satisfying. This is a large subject and it takes us well beyond the scope of the present paper. However, the favored refinement for Stag Hunt would be universal hunting of hare, because it is the risk dominant equilibrium. (For a general discussion see [VBB91,VBB90].) Agents playing this way might well be viewed as “rational fools” [Sen77] by LPS agents.

LPS agents may be rational. At least naïvely, the L/U regimes employed by our gridscape and MLPS agents are sensible, and may be judged rational, or at least not irrational. Exploring the environment, as our LPS agents do, probing

⁸ A genuine theorem, described in standard texts, e.g., [Bin92].

it with play of different policies, informed by recent experience, is on the face it entirely reasonable. Why not try learning by experience if it is not obvious what to do in the absence of experience? I shall now try to articulate a sense in which LPS agents may be judged rational, even though they violate the rationality assumptions of classical game theory and rational choice theory.

Contexts of maximum taking (MT). Given a set of outcomes whose values are known, perhaps under risk (i.e., up to a probability distribution), given a consistent, well-formed preference structure valuing the outcomes, and given a set of actions leading (either with certainty or with risk) to the outcomes, rational choice theory (or utility theory) instructs us to choose an action that results in our taking the maximum expected value on the outcomes. Presented with valued choices under certainty or risk, we are counseled to take the maximum value in expectation. Although the theory is foundational for classical game theory and economics, it has also been widely challenged both from a normative perspective and for its empirical adequacy.⁹

Contexts of maximum seeking (MS). In an MS context an agent can discriminate among outcomes based on their values to the agent, but the connection between the agent's possible actions and the resulting outcomes is uncertain in the technical sense: the agent does not have an objectively well grounded probability distribution for associating outcomes with actions. In seeking the maximum return for its actions, the agent has little alternative but to explore, to try different actions and to attempt to learn how best to take them.¹⁰

Exploring rationality is appropriate for MS contexts. The claim I wish to put on the table is that in MS as distinct from MT contexts, rationality is best thought of as an appropriate learning process. An agent is rational in an MS context to the extent that it engages effectively in learning to obtain a good return. In doing so, it will be inevitable that that agent engages in some form of trial and error process of exploring its environment. Rationality of this kind may be called an *exploring rationality* to distinguish it from what is often called *ideal rationality*, the kind described by rational choice theory and which is, I submit, typically not appropriate in MS contexts. See [Kim04a] for further discussion of the concept of an exploring rationality.

Evaluate exploring rationalities analytically by performance. LPS models with their articulated L/U regimes afford an excellent framework for evaluating forms of exploring rationality. Such evaluation will turn largely on performance under a given L/U regime. For starters and for now informally, an L/U regime may be assessed with regard to whether it is generally a strong performer. Rational admissibility is a useful concept in this regard.

⁹ Good, wide-ranging discussion can be found in [Fri96,GS94]. A classic paper [KMRW82] develops a model in which for the finitely repeated Prisoner's Dilemma game it is sometimes rational for a player to cooperate, *provided the player believes the counter-player is irrational*. Since both players would benefit by mutual cooperation it seems a stretch to call all attempts to find it irrational.

¹⁰ Classical game theory seeks to finesse this situation by assuming classical rationality and common knowledge. The present essay may be seen as an exploration of principled alternatives to making these very strong assumptions.

General Definition 1 (Rational Admissibility) *A learning (update) regime for policies of play in an indefinitely repeated game is rationally admissible if*

1. *It performs well if played against itself (more generally: it performs well if universally adopted).*
2. *It performs well if played against other learning regimes that perform well when played against themselves (more generally: the other learning regimes perform well if universally adopted).*
3. *It is not vulnerable to catastrophic exploitation.*

To illustrate, in the gridscape model IMITATE THE BEST NEIGHBOR performs well against itself in that when everyone uses it, as we have seen, trust breaks out and stag hunting prevails robustly. The null L/U policy of ALWAYS HUNT STAG also does well against itself, and both IMITATE THE BEST NEIGHBOR and ALWAYS HUNT STAG will do well against each other. ALWAYS HUNT STAG, however, is catastrophically vulnerable to ALWAYS HUNT HARE. IMITATE THE BEST NEIGHBOR on the other hand will do better, although how much better depends on the payoff structure of the stage game. Some stag hunters may open themselves up to exploitation because they have one neighbor who hunts stag and is surrounded by stag hunters. In sum, with reference to the set of these three L/U policies, IMITATE THE BEST NEIGHBOR is uniquely rationally admissible (robustly, across a wide range of stag game payoff structures). A similar point holds for the MLPS model discussed above.

Two additional comments. First, “not vulnerable to catastrophic exploitation” is admittedly vague. It is not to my purpose to provide a formal specification here. I believe that more than one may be possible and in any event the topic is a large one. The motivating intuition is that a learning regime is vulnerable to exploitation if it learns to forego improving moves for which the counter-players have no effective means of denial. Thus, an agent that has learned to hunt stag in the face of the counter-player hunting hare is being exploited because it is foregoing the option of hunting hare, the benefits of which cannot be denied by the counter-player. Similarly, agents cooperating in Prisoner’s Dilemma are not being exploited. Even though each is foregoing the temptation to defect, the benefits of defecting can easily be denied by the counter-player following suit and also defecting. Second, the similarity between the definition, albeit informal, of rational admissibility and the concept of an ESS (evolutionarily stable strategy, [May82]) is intended. In a nutshell, a main message of this paper is that for repeated games it is learning regimes and consideration sets of policies, rather than strategies alone, that are key to explanation. (And dare one suggest that rational play in one-shot games may sometimes draw on experience in repeated games?)

Evaluate exploring rationalities empirically, for descriptive adequacy. As noted, it is well established that rational choice theory (ideal rationality) is not descriptively accurate at the individual level. In light of the results and observations given here, one has to ask to what degree subjects at variance from the received theory are perceiving and responding to contexts of maximum seeking

(MS), rather than the postulated MT contexts. In any event, it is worth noting that foraging by animals—for food, for mates, for shelter or other resources—is a ubiquitous natural form of behavior in an MS context [GC00,SK86], for which models under the LPS framework would seem a good fit. Experimental investigation is only beginning. I think it shows much promise.

* * *

In conclusion, the problems and paradoxes of trust are vitally important on their own. Trust is the “cement of society”.¹¹ Understanding it is crucial to maintenance and design of any social order, including and especially the new social orders engendered by modern communications technologies, globalization, global warming, and all that comes with them. I have tried to contribute in a small way to understanding how and when trust can emerge or be destroyed. The gridscape and MLPS models are helpful, but they can be only a small part of the story and even so their depths have barely been plumbed. But it’s a start; it’s something. The more significant point, I think, is that the problems of trust lead us, via these very different models, to a common pattern that abstracts them: LPS, learning in policy space, and contexts of maximum seeking (MS), as distinguished from contexts in which maximum taking (MT) is appropriate. The fact, demonstrated here and elsewhere, that agents adopting this stance generate more trust and improve their take from the environment, is encouraging. So is the observation that such behavior is analogous to, if not related to or even a kind of, foraging behavior.

Acknowledgements. Many thanks to Alex Chavez and James D. Laing for comments on an earlier version of this paper.

References

- [Bin92] Ken Binmore, *Fun and games: A text on game theory*, D.H. Heath and Company, Lexington, MA, 1992.
- [Cam03] Colin F. Camerer, *Behavioral game theory: Experiments in strategic interaction*, Russell Sage Foundation and Princeton University Press, New York, NY and Princeton, NJ, 2003.
- [Els89] Jon Elster, *The cement of society: A study of social order*, Studies in rationality and social change, Cambridge University Press, Cambridge, UK, 1989.
- [Fri96] Jeffrey Friedman (ed.), *The rational choice controversy*, Yale University Press, New Haven, CY, 1996, Originally published as *Critical Review*, vol. 9, nos. 1–2, 1995.
- [Gam05a] GameTheory.net, *Assurance game*, <http://www.gametheory.net/Dictionary/Games/AssuranceGame.html>, Accessed 8 February 2005.
- [Gam05b] ———, *Stag hunt*, <http://www.gametheory.net/Dictionary/Games/StagHunt.html>, Accessed 8 February 2005.
- [GC00] Luc-Alain Giraldeau and Thomas Caraco, *Social foraging theory*, Princeton University Press, Princeton, NJ, 2000.

¹¹ Elster’s term [Els89], after Hume who called causation the cement of the universe.

- [Gin00] Herbert Gintis, *Game theory evolving: A problem-centered introduction to modeling strategic interaction*, Princeton University Press, Princeton, NJ, 2000.
- [GS94] Donald P. Green and Ian Shapiro, *Pathologies of rational choice theory: A critique of applications in political science*, Yale University Press, New Haven, CT, 1994.
- [HS88] John C. Harsanyi and Reinhard Selten, *A general theory of equilibrium selection in games*, MIT Press, Cambridge, MA, 1988.
- [Kim04a] Steven O. Kimbrough, *A note on exploring rationality in games*, Working paper, University of Pennsylvania, Philadelphia, PA, March 2004, Presented at SEP (Society for Exact Philosophy), spring 2004. <http://opim-sun.wharton.upenn.edu/~sok/comprats/2005/exploring-rationality-note-sep2004.pdf>.
- [Kim04b] ———, *Notes on MLPS: A model for learning in policy space for agents in repeated games*, working paper, University of Pennsylvania, Department of Operations and Information Management, December 2004, <http://opim-sun.wharton.upenn.edu/~sok/sokpapers/2005/markov-policy.pdf>.
- [CLK04] Steven O. Kimbrough, Ming Lu, and Ann Kuo, *A note on strategic learning in policy space*, Formal Modelling in Electronic Commerce: Representation, Inference, and Strategic Interaction (Steven O. Kimbrough and D. J. Wu, eds.), Springer, Berlin, Germany, 2004, pp. 463–475.
- [KMRW82] David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson, *Rational cooperation in the finitely repeated prisoners' dilemma*, *Journal of Economic Theory* **27** (1982), 245–252.
- [LR57] R. Duncan Luce and Howard Raiffa, *Games and decisions*, John Wiley, New York, NY, 1957, Reprinted by Dover Books, 1989.
- [May82] John Maynard Smith, *Evolution and the theory of games*, Cambridge University Press, New York, NY, 1982.
- [RGG76] Anatol Rapoport, Melvin J. Guyer, and David G. Gordon, *The 2×2 game*, The University of Michigan Press, Ann Arbor, MI, 1976.
- [Rou04] Jean Jacques Rousseau, *A discourse upon the origin and the foundation of the inequality among mankind*, <http://www.gutenberg.org/etext/11136>, 17 February 2004, Originally published, in French, in 1755.
- [Sen77] Amartya K. Sen, *Rational fools: A critique of the behavioural foundations of economic theory*, *Philosophy and Public Affairs* **6** (1977), 317–344.
- [SK86] David W. Stephens and Jorn R. Krebs, *Foraging theory*, Princeton University Press, Princeton, NJ, 1986.
- [Sky96] Brian Skyrms, *Evolution of the social contract*, Cambridge University Press, Cambridge, UK, 1996.
- [Sky04] ———, *The stag hunt and the evolution of social structure*, Cambridge University Press, Cambridge, UK, 2004.
- [VBB90] John B. Van Huyck, Raymond C. Battalio, and Richard O. Beil, *Tacit coordination games, strategic uncertainty, and coordination failure*, *The American Economic Review* **80** (1990), no. 1, 234–248.
- [VBB91] ———, *Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games*, *The Quarterly Journal of Economics* **106** (1991), no. 3, 885–910.