Satisficing Learning Dynamics in the Stag Hunt

Matthew S. Nokleby
Elect & Comp Engineering Dept.
Brigham Young University, Provo, Utah
e-mail: msn@ee.byu.edu
Phone: 1-801-422-5349

Wynn C. Stirling
Elect & Comp Engineering Dept.
Brigham Young University, Provo, Utah
e-mail: wynn@ee.byu.edu
Phone: 1-801-422-7669

A. Lee Swindlehurst
Elect & Comp Engineering Dept.
Brigham Young University, Provo, Utah
e-mail: swindle@ee.byu.edu
Phone: 1-801-422-4343

***Abstract*— Satisficing game theory is an alternative to traditional game theory that allows players' utilities to be conditioned upon the preferences of others. This construction, however, requires that players have accurate information about each other's preferences. We discuss a learning mechanism for satisficing players to estimate other's preferences through repeated interactions. We apply this mechanism to the Stag Hunt game, and show that under most circumstances, players can indeed learn players' preferences sufficiently to make correct decisions.**

I. Introduction

In the design of systems of autonomous agents, decision mechanisms play an important role. Much emphasis has been placed upon game theory as a useful decision theory upon which to base agent's choices. Game theory, as developed by von Neumann and Morgenstern [1], is based on the concept of *individual rationality*, which asserts that players seek solely to maximize their benefit without regard for the preferences of others. The Nash equilibrium [2] of a game defines players' strategies such that no player can improve his payoff by unilaterally modifying his strategy, giving players an optimal strategy if all other players similarly attempt to maximize.

However, in this narrow maximization, players can miss opportunities for social behavior that could benefit them collectively as well as individually. Stirling [3] has recently put forth *social rationality* as an alternative form of rationality which may be more useful in the synthesis of potentially cooperative artificial systems. This definition of rationality, as formalized in satisficing game theory [4], rests upon two major assertions: (a) players may consider other players' preferences when constructing their utility functions, and (b) rather than strictly maximizing, players are willing to take actions for which the relative benefits outweigh the costs. We call such actions *satisficing*.

In [5], Nokleby and Stirling consider evolutionary methods by which socially rational players and communities may adapt their preferences to the game structure and the preferences of other players. Here, we consider "learning" from a different perspective. Since players in satisficing theory can condition their utilities on the preferences of others, players need to know each other's preference values if they are to make socially rational decisions. While in many cases it may be reasonable to assume that players have this information *a priori*, such will not always be the case. Thus, we focus on a passive learning mechanism by which players, through repeated interactions, can learn each other's preferences.

1-4244-0166-6/06/$20.00 @2006 IEEE

To explicate this model, we focus on the Stag Hunt discussed in [5, 6]. The Stag Hunt is a game which juxtaposes the potential benefits of cooperation with the risks for doing so. As usually formalized, it involves two hunters. They can catch a stag only if they hunt stag together. Each hunter, however, can catch a hare acting alone. Thus, each may go for the higher, riskier payoff of the stag, or content himself with the lower—but more secure—payoff of hare-hunting. In the satisficing model, players consider both their own risk-aversion with that of the other player in determining which action to take. They therefore must have correct preference information in order to take coherent actions.

First, we present a brief introduction to satisficing game theory. We then describe the Stag Hunt in both classical and satisficing frameworks. We next formalize the satisficing learning dynamics. After presenting simulation results, we give our conclusions.

II. Satisficing Game Theory

Conventional von Neumann-Morgenstern game theory assumes that each player forms its preferences by taking into consideration the possible actions, *but not the preferences for action* of the other players. These preferences are encoded into von Neumann-Morgenstern utility functions, which provide a quantitative expression of the qualitative preferences. By contrast, socially rational decision makers form their preferences via *social utilities* [3] that encode individual preferences as functions of the *preferences* of others as well as of themselves. Social utilities possess a structure, or syntax, that is mathematically identical to the structure of probability mass functions, albeit with completely different semantics. Just as probability mass functions account for the *epistemic* properties of random variables, social utilities account for the *praxeic*, or action-taking properties of decision-making agents.

To develop the mathematical structure of satisficing game theory, let us consider a community of two players, X_1 and X_2, with countable action spaces U_1 and U_2. For each player, we define two social utilities which describe the preferences from two different perspectives. The *selectability function*, denoted p_S, considers actions in terms of their benefits. The *rejectability function*, denoted p_R, considers actions in terms of their costs. Because these utilities are mass functions, both p_S

and p_R are normalized across their action spaces and are non-negative functions. These utilities provide players with a formal mechanism for defining what is good enough: the actions for which the degree of success is at least as great as is the degree of resource consumption. The *individually satisficing set* for X_i is the set

$$\Sigma_i = \{u \in U\colon p_{S_i}(u) \geq q p_{R_i}(u)\}, \tag{1}$$

where q is called the *index of caution*. Nominally, $q = 1$, but a player may alter its definition of "good enough" by increasing or decreasing q. Setting $q \leq 1$ ensures that $\Sigma \neq \emptyset$. We may combine these two individual sets of decisions by forming the *satisficing rectangle*, which is the Cartesian product

$$\Re_{12} = \Sigma_1 \times \Sigma_2. \tag{2}$$

Any action vector $(u_1, u_2) \in \Re_{12}$ is simultaneously satisficing to each player in terms of its individual preferences.

By analogy with conventional probability, when constructing statistical models of multiple random phenomena, local relationships are often characterized by Bayesian networks [7, 8]. Similarly, for societies whose relationships can be characterized by marginal and conditional social utilities defined over small clusters of individuals, the individual mass functions can also be constructed by the chain rule. The result is a *praxeic network*, consisting of a directed acyclic graph (DAG) whose vertices are the selecting and rejecting selves, and whose edges are the conditional selectability and rejectability functions. Once the interdependence relationships are defined, we may apply well-established techniques for solving Bayesian networks such as Pearl's Belief Propagation Algorithm [7].

III. The Stag Hunt

A. Convential Framework

We first review the conventional two-player approach to the Stag Hunt. Let s and h denote the decision to hunt stag and hare, respectively. We assume that the payoff for hunting hare is independent of the other player. That is, both players must hunt stag to get the higher payoff, but each player can independently derive positive utility from hunting hare. A payoff matrix for this game is shown in Table I.

TABLE I
PAYOFF MATRIX FOR A TWO-PLAYER STAG HUNT.

		Player 1	
Player 2		s	h
s		(5, 5)	(1, 2)
h		(2, 1)	(2, 2)

There are two pure-strategy Nash equilibria for this game: (s, s) and (h, h). Both are also evolutionarily stable equilibria, meaning that a population of hare-hunters cannot be "invaded" by a group of stag-hunters, and vice versa [9]. The (s, s) equilibrium, however, may be regarded as the less stable equilibrium when we consider variations in the payoffs [10, 11]. Consider the effects of momentary fluctuations: if, while both players are hunting stag, one player's hare-hunting utility temporarily moves above 5, he will begin hunting hare, with the other player following to avoid completely wasting his effort. Thus, one player's choice can move the group from the (s, s) equilibrium towards (h, h). However, no variations in stag-hunting utility can move the group towards (s, s). A player who begins to hunt stag from the (h, h) equilibrium will have minimal payoff unless the other player quickly shifts to hunting stag. The need here for simultaneity significantly weakens the possibility for cooperation.

B. Satisficing Framework

The selecting and rejecting nodes for the two-player Stag Hunt are denoted S_i and R_i, $i = 1, 2$, with identical action spaces $U_i = \{s, h\}$ for $i = 1, 2$. In the framework presented in [5], rejectability is associated with the raw opportunity cost of an action. The cost of hunting hare is the payoff derived from catching stag, and the cost of hunting stag is the payoff for catching hare. Since stag hunting yields a greater resource, the opportunity cost for hunting hare is greater than the opportunity cost for hunting stag. We associate selectability with successful cooperation. That is, if a successful stag hunt is available to the selecting self, it will prefer stag hunting, otherwise, it will prefer hare hunting.

Next, we define the interconnections between the four selves and form the praxeic network, as shown in Figure 1. In addition to the vertices corresponding to the four selves, this network contains three epistemic nodes that are included to account for uncertainty. To account for the possibility of failure in both hare-hunting and stag-hunting, we introduce three binary random variables: θ_s, θ_{h_1}, and θ_{h_2}. $\theta_{h_i} = 1$ means that a successful hare hunt is available to the players *if they choose to pursue*. That is, a player who chooses to hunt hare will succeed if $\theta_{h_i} = 1$ and fail if $\theta_{h_i} = 0$. Similarly, $\theta_s = 1$ signifies that a successful stag hunt is available to the selecting selves S_i, and $\theta_s = 0$ means that stag hunting will result in failure.

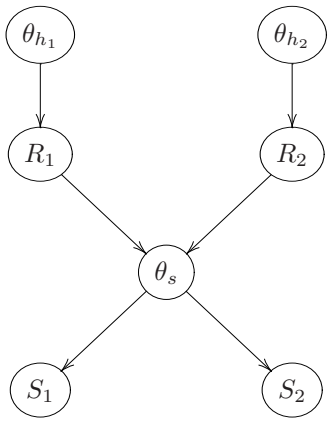

Fig. 1. The epistemi-praxeic DAG for the Stag Hunt.

Let ϕ_{s_i} and ϕ_{h_i}, denote the utility of stag and hare, respectively, expressed in arbitrary units. Normalizing, the utility of

hare-hunting becomes

$$\mu_i = \frac{\phi_{h_i}}{\phi_{h_i} + \phi_{s_i}} = \frac{1}{1 + \frac{\phi_{s_i}}{\phi_{h_i}}} \tag{3}$$

for $i = 1, 2$, and we see that this utility depends only on the ratio $\frac{\phi_{s_i}}{\phi_{h_i}}$. The utility of stag-hunting is then $1 - \mu_i$.

From Table I, we might initially expect that $\phi_{s_i} = 5$ and $\phi_{h_i} = 2$, resulting in $\mu_i = \frac{2}{7}$. However, we also choose to consider the risk-aversion of the players. A risk-averse player might increase ϕ_{h_i}, while a payoff-seeking player, ignoring risk, might do the opposite. A maximally risk-averse player will hunt stag only if its success is certain, while a fully payoff-seeking player will hunt stag regardless of the odds. To ensure a meaningful game, we assume that both players will never prefer hare to stag, or $\mu_i < \frac{1}{2}$ for $i = 1, 2$.

These utilities assume that it is possible for X_i to hunt hare successfully. If, however, X_i cannot hunt hare successfully, there is no opportunity cost for stag hunting. We therefore define the conditional rejectability functions as

$$p_{R_i|\theta_{h_i}}(u_i|1) = \begin{cases} \mu_i, & \text{for } u_i = s \\ 1 - \mu_i, & \text{for } u_i = h \end{cases} \tag{4}$$

and

$$p_{R_i|\theta_{h_i}}(u_i|0) = \begin{cases} 0, & \text{for } u_i = s \\ 1, & \text{for } u_i = h \end{cases} \tag{5}$$

for $i = 1, 2$. These conditional rejectabilities characterize the utility structures for hunting hare given that a hunt could or could not succeed. To compute the unconditional rejectability, we must take into consideration the probabilities of successful hunts and compute the expected utility of a successful hunt. These probabilities are characterized by the probability mass functions for θ_{h_i}:

$$p_{\theta_{h_i}}(\vartheta_{h_i}) = \begin{cases} \eta_i, & \text{for } \vartheta_{h_i} = 1 \\ 1 - \eta_i, & \text{for } \vartheta_{h_i} = 0 \end{cases}, \tag{6}$$

using the symbols θ for the random variable and ϑ for its realization. The expected utility is obtained by summing over ϑ_{h_i}, yielding

$$\begin{aligned} p_{R_i}(u_i) &= \sum_{\vartheta_{h_i}} p_{R_i|\theta_{h_i}}(u_i|\vartheta_{h_i}) \cdot p_{\theta_{h_i}}(\vartheta_{h_i}) \\ &= \begin{cases} \epsilon_i, & \text{for } u_i = s \\ 1 - \epsilon_i, & \text{for } u_i = h \end{cases}, \end{aligned} \tag{7}$$

where $\epsilon_i = \mu_i \eta_i$ is the expected utility of hunting hare.

We next define the dependence relationships for θ_s. The distribution of this random variable, which is conditioned upon the rejecting preferences of both players, represents the degree to which a successful stag hunt is available to the selecting selves. The distribution of θ_s incorporates both the degree to which R_1 and R_2 reject cooperation as well as how well the players hunt stag. We model the latter consideration by defining $0 \leq \sigma \leq 1$, which represents the probability of catching a stag given that the players cooperate. If neither R_1 or R_2 rejects stag hunting, then a successful stag hunt is possible — from the perspective of the selecting selves — with probability σ. We characterize this by defining

$$p_{\theta_s|R_1R_2}(\vartheta_s|h,h) = \begin{cases} \sigma, & \text{for } \theta_s = 1 \\ 1 - \sigma, & \text{for } \theta_s = 0 \end{cases}. \tag{8}$$

If, however, either player unilaterally rejects stag-hunting, the probability of catching a stag is zero, thus

$$\begin{aligned} p_{\theta_s|R_1R_2}(\vartheta_s|s,s) = p_{\theta_s|R_1R_2}(\vartheta_s|s,h) = p_{\theta_s|R_1R_2}(\vartheta_s|h,s) \\ = \begin{cases} 0, & \text{for } \vartheta_s = 1 \\ 1, & \text{for } \vartheta_s = 0 \end{cases}. \end{aligned}$$

The selectability of each player is influenced by probability of a successful stag hunt. Here, the selecting self is concerned with cooperating successfully if possible, but avoiding failure if it is not. Thus, the players have a simple definition for selectability:

$$p_{S_i|\theta_s}(u_i|\vartheta_s) = \begin{cases} 1 & \text{for } u_i = s, \vartheta_s = 1 \\ 0 & \text{for } u_i = h, \vartheta_s = 1 \\ 0 & \text{for } u_i = s, \vartheta_s = 0 \\ 1 & \text{for } u_i = h, \vartheta_s = 0 \end{cases}. \tag{9}$$

The selecting self prefers stag-hunting inasmuch as a stag hunt is available, but prefers a hare-hunt if it is not. The marginal $p_{S_i}(u_i)$ is given by

$$p_{S_i}(u_i) = \begin{cases} \sigma(1 - \epsilon_1)(1 - \epsilon_2), & \text{for } u_i = s \\ 1 - \sigma(1 - \epsilon_1)(1 - \epsilon_2), & \text{for } u_i = h \end{cases}. \tag{10}$$

The satisficing rectangle is the set of action vectors which are simultaneously satisficing to each player individually, i.e., for which $p_{S_i} \geq q p_{R_i}$, $i = 1, 2$. In Figure 2, we set q to unity and plot the regions of the satisficing rectangle as functions of ϵ_1 and ϵ_2. To account for all possible values of these parameters, we impose the maximum range $0 \leq \mu_i < \frac{1}{2}$, $i = 1, 2$. The expected utility must obey this same constraint. Thus, only the region for which $0 \leq \epsilon_i < \frac{1}{2}$ is meaningful for the Stag Hunt. There are four possibilities for the satisficing rectangle. In Region (s, s), both players have low expected utility of hare hunting. In Region (s, s), both players have low risk-aversion. In Region (h, h) both players consider the risk unacceptable and separately hunt hare. In regions (h, s) and (s, h), however, one player is strongly risk-averse while the other strongly seeks payoff, with the result that one tries to cooperate and the other does not.

IV. Learning Dynamics in the Stag Hunt

A. Learning in Satisficing Games

In our learning dynamics, we make several assumptions. We first consider the satisficing model to be common knowledge. That is, all players are known to be socially rational with utilities that conform to (7) and (10). We assume that the environmental parameters σ, η_1, η_2 are common knowledge. Rather than trying to construct a model through observations,

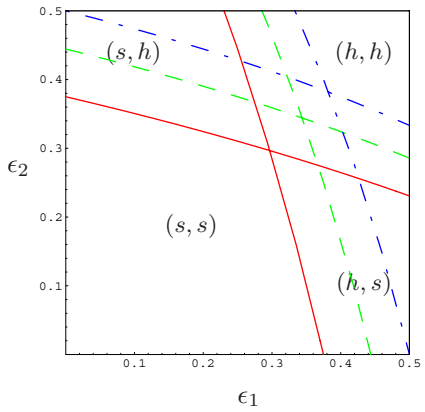

Fig. 2. The satisficing rectangle sets for $\sigma = 0.6$ (red line), $\sigma = 0.8$ (green dashed line), and $\sigma = 1$ (blue dot-dashed line).

players need to estimate others' risk-aversion preferences μ_i as model parameters. We also assume a continuous-time model. Players continuously "play" the Stag Hunt, revising their decisions and observing others' actions. Although the dynamics described here could be applied to an arbitrary satisficing game with different parameters, we will use notation associated with the Stag Hunt for simplicity.

Our learning mechanism employs a passive approach: players keep a probabilistic estimate of the other players' μ_i, and update their estimates according to their observations. First, we quantize the possible values of μ by defining $K = \{\nu_1, \nu_2, \cdots, \nu_n\}$, a set of n evenly spaced values of μ over $[0, \frac{1}{2})$. Each player posesses a probability mass function through which he models his perception of the others' μ. We express this mass function as a time-varying probability vector $\mathbf{p_{ij}}(t) = \{p_{ij_1}(t), p_{ij_2}(t), \cdots, p_{ij_n}(t)\}$, where each element $p_{ij_k}(t)$ represents X_i's belief that $\mu_j = \nu_k$. At each decision step, X_i determines its best estimate as to μ_j by taking the expectation over $\mathbf{p_{ij}}(t)$:

$$\hat{\mu}_{ij}(t) = \sum_{k=1}^{n} \nu_k p_{ij_k}(t). \tag{11}$$

X_i uses this estimated value to form his own p_{S_i} and p_{R_i}, since his decisions depend on X_j's preferences as well as his own. Since $\hat{\mu}_{ij}(t)$ varies in time as $\mathbf{p_{ij}}(t)$ evolves, a player's social utilities also change with time. Specifically, players' utility functions are given by

$$p_{S_i}(u_i, t) = \begin{cases} \sigma(1-\epsilon_i)(1-\hat{\mu}_{ij}(t)\eta_j), & \text{for } u_i = s \\ 1-\sigma(1-\epsilon_i)(1-\hat{\mu}_{ij}(t)\eta_j), & \text{for } u_i = h \end{cases} \tag{12}$$

and

$$p_{R_i}(v_i, t) = \begin{cases} \epsilon_i, & \text{for } u_i = s \\ 1-\epsilon_i, & \text{for } u_i = h \end{cases}. \tag{13}$$

As both players make their decisions, they can observe each other's actions and refine their beliefs as expressed by $\mathbf{p_{ij}}(t)$. In our model, players update their beliefs according to a system of differential equations similar to the replicator dynamics. Replicator dynamics is an evolutionary method which promotes strategies that are well-suited to the game's environment [12]. The population which can implement n different strategies is modeled as an n-dimensional vector of population shares. A strategy's population share increases or decreases according to the strategies' relative expected payoff.

Our learning dynamics similarly models a player's beliefs by letting $\mathbf{p_{ij}}(t)$ vary according to the dynamics. Whereas in the replicator dynamics popuation shares vary with relative payoff, the probabilities for a preference value in the learning dynamics grow and shrink according to its relative "plausibility". We define this plausibility with a *plausibility function* $u_{p_j}(\nu_k, t)$, which represents the plausibility that $\mu_j = \nu_k$.

To construct the plausibility fuction, X_i first constructs estimates $\hat{p}_{S_j}(u_j, \nu_k)$ and $\hat{p}_{R_j}(u_j, \nu_k)$ for all $\nu_k \in K$. $\hat{p}_{S_j}$ and $\hat{p}_{R_j}$ are based upon (a) ν_k, and (b) X_i's true preference μ_i. That is, in constructing his estimates of X_j's utilities, X_i assumes that X_j has complete knowledge of μ_i. While this assumption introduces further inaccuracies into our estimates, relaxing it would result in an infinite regress of players' estimating other players' estimates.

Once X_i has estimated X_j's utilities, he can determine how well they agree with observation: is X_j's observed action individually satisficing according to $\hat{p}_{R_j}$ and $\hat{p}_{S_j}$? Also, *how strongly* would X_j prefer the observed action according to the estimates? If, according to $\hat{p}_{R_j}$ and $\hat{p}_{S_j}$, X_j would strongly reject an action u_j which we observe him taking, then ν_k is significantly less plausible than if X_j would only mildly reject u_j. To account for this, given an observed action u_j, we define the plausibility for ν_k as

$$u_{p_j}(\nu_k) = \begin{cases} C + k_c \left|\hat{p}_{S_j}(u_j) - \hat{p}_{R_j}(u_j)\right|, \text{for } \hat{p}_{S_j}(u_j) \geq \hat{p}_{R_j}(u_j) \\ C - k_i \left|\hat{p}_{S_j}(u_j) - \hat{p}_{R_j}(u_j)\right|, \text{for } \hat{p}_{S_j}(u_j) < \hat{p}_{R_j}(u_j) \end{cases} \tag{14}$$

where $k_c, k_i \geq 0$ specify how much to "reward" correct predictions and "punish" incorrect predicitons, respectively, and C represents a constant chosen to ensure non-negativity. At each decsion step, players compute the plausibility function for each $\nu_k \in K$. They then update their probability vectors according to

$$\dot{p}_{ij_k} = \left[u_{p_j}(\nu_k) - u_{p_j}(\mathbf{p_{ij}})\right] p_{ij_k}, \tag{15}$$

where $u_j(\mathbf{p_{ij}})$ represents expected preference plausibility, or

$$u_{p_j}(\mathbf{p_{ij}}) = \sum_{k=1}^{n} p_{ij_k} u_{p_j}(\nu_k). \tag{16}$$

Thus, if a value ν_k is more plausible than the average plausibility, its probability mass increases, while values less plausible than average diminish.

The final consideration in our model is the initial state of $\mathbf{p_{ij}}$. Since players' actions may change with others' preferences, a player's initial distribution might cause him to take a vastly different action than he would with correct information. These poor "first impressions" can cause players to lock into incorrect

assumptions from which the dynamics cannot escape. We therefore must choose our initial conditions with care.

In the case of the Stag Hunt, we opt for a "reflexive" distribution. That is, X_i's distribution $\mathbf{p_{ij}}$ initially centers around his own μ_i. In the Stag Hunt, at least, this assumption makes intuitive sense: given no preference information, a risk-averse X_i would likely play cautiously and initially assume that X_j is also risk-averse. Similarly, a payoff-seeking X_i would be somewhat willing to ignore risk and assume that X_j is prone to stag-hunting. We choose a maximum-entropy initial distribution over K, with the constraint that the expected value of $\mathbf{p_{ij}}$ is μ_i. Using the method of Lagrangian multipliers, one can show that this distribution is given by the form

$$p_{ij_k} = e^{\nu_k \lambda_0 + \lambda_1 - 1}, \tag{17}$$

where λ_0 and λ_1 are chosen to satisfy the constraints

$$\sum_{k=1}^{n} p_{ij_k} = 1, \qquad \sum_{k=1}^{n} \nu_k p_{ij_k} = \mu_i. \tag{18}$$

B. Simulation Results

For our simulations, we let $\eta_i = \sigma = 1$. We approximate the dynamics numerically by converting the system of differential equations in (15) into an equivalent system of difference equations. We begin with a simple example. Let $\mu_1 = 0.1$, and $\mu_2 = 0.4$. This pair should hunt stag cooperatively, but initially they do not. From X_1's perspective, they are deep in the (s, s) region: $\mu_1 = \hat{\mu}_{12} = 0.1$. X_2, however, initially assumes (h, h) since $\hat{\mu}_{21} = \mu_2 = 0.4$. Thus, X_1 hunts stag while X_2 hunts hare. As they observe each other's actions, they begin to update their beliefs according to (15). We let $k_i = 1$ and $k_c = 0$, meaning that the dynamics only correct erroneous estimates, and will not explicitly reward estimates whose predictions agree with observations. This most conservative configuration gives us slower dynamics, but it reduces the likelihood that players will lock in to an incorrect decision region.

We plot the dynamics for the first thirty time steps in Figure 3. We can see that X_1's estimates (blue dots), which begin at $(\mu_1, \hat{\mu}_{12}) = (0.1, 0.1)$, immediately begin shooting upward. Since X_2 initially hunts hare, X_1 realizes that μ_2 cannot be as low as 0.1, and adjusts accordingly. X_2 also corrects its estimate (green dots) as it observes X_1's stag hunting. Fortunately, X_2 also begins hunting stag before X_1's estimate moves too far, and both estimates remain in the (s, s) region, settling down approximately to $(\mu_1, \hat{\mu}_{12}) = (0.1, 0.43)$ and $(\hat{\mu}_{21}, \mu_2) = (0.31, 0.4)$. Note that while the players' estimates have converged to the appropriate decision region, they do not converge to the correct values. Since the players have no information beyond their observations, this is the best that we can hope for in a passive learning scheme. However, players can overcome their initial cynicism and learn to cooperatively hunt stag.

For an exhaustive simulation, we run the learning dynamics for all ordered pairs $(\mu_1, \mu_2) \in K^2$. We again let $k_i = 1$ and $k_c = 0$. While we cannot analyze the dynamics of each ordered pair, we can plot each pair's steady-state decision regions. To show the benefit of the reflexive initial distrubution discussed in section IV-A, we show results of two exhaustive simulations. In the first simulation (Figure 4(a)) players initially assume a uniform distribution over K. Their initial actions, therefore, assume a $\hat{\mu} = .25$ for their partner. In the second simulation (Figure 4(b)), players employ the reflexive distributions. Each dot in the figures represents the true pair (μ_1, μ_2), and the dots' colors correspond with the different decision regions: a green dot means that the dynamics converge to the (s, s) region, red represents (h, h), blue represents (s, h), and cyan (h, s).

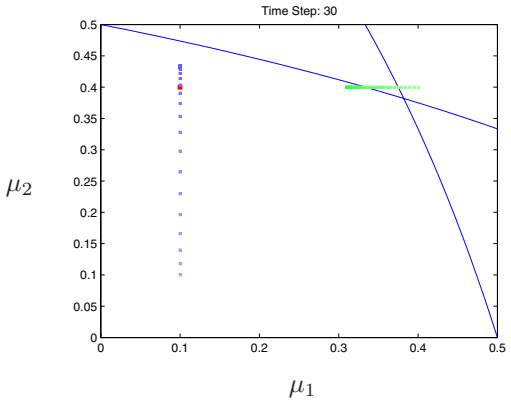

Fig. 3. Learning dynamics for $\mu_1 = .1, \mu_2 = .4$ with $k_i = 1$ and $k_c = 0$. Blue represents X_1's estimates, green represents X_2's, and red represents the true ordered pair.

In Figure 4(a), players who should end up in (h, h) often end up hunting stag cooperatively. Because they assume a low initial $\hat{\mu}$, the players both initially hunt stag. These initial actions reinforce the incorrectly low estimates of $\hat{\mu}$, and the players never observe contradictory actions which would lead them toward hare-hunting. Thus, this learning environment artificially promotes stag-hunting.

In Figure 4(b), however, the boundaries around (s, s) and (h, h) resolve correctly: players never converge to mutual hare-hunting when they both should be stag-hunting, and vice versa. The reflexive distributions cause risk-averse players to make conservative initial estimates, preventing them from inadvertently locking into stag-hunting. This learning environment, however, artificially extends the (s, h) and (h, s) regions. Here, the risk-averse player initially assumes that his partner hunts hare. He observes, however, the other player attempting to hunt stag. The dynamics shift his preferences just enough to correctly predict that his partner will hunt stag, but his estimate $\hat{\mu}$ of the other player is still sufficiently conservative that he remains hunting hare.

Finally, leaving all else constant, we set $k_c = 0.2$, which causes the plausibility function to favor correct predictions rather than simply punish incorrect prediction. We plot the results in Figure 5. This plot is similar to Figure 4(b) except that the (s, h) and (h, s) regions are entirely replaced with mutual stag-hunting. The definition for the plausibility function in (14) explains this phenomenon. With $k_c > 0$, the plausibility function not only favors correct answers; it also favors

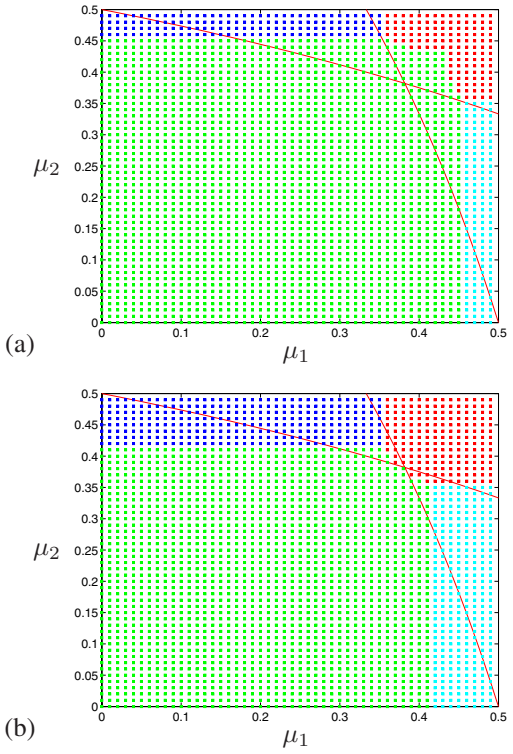

Fig. 4. Steady-state decision regions for different (μ_1, μ_2) pairs, with (a) uniform and (b) reflexive initial distributions. Green represents (s, s), red (h, h), blue (s, h), and cyan (h, s).

estimates that give the most extreme prediction of the observed behavior. That is, if X_j is observed hunting stag, $u_{p_j}(\nu_k)$ is maximized by $\nu_k = 0$, since this value will maximize the difference $\left|\hat{p}_{S_j}(u_j) - \hat{p}_{R_j}(u_j)\right|$. In the limit, X_i does not see X_j as merely a stag-hunter who will hunt stag with X_i. He sees him as a "die-hard" stag hunter who will hunt stag regardless of the other's preferences, and with whom even the most risk-averse should hunt stag. The (s, h) and (h, s) regions cannot survive such a learning environment.

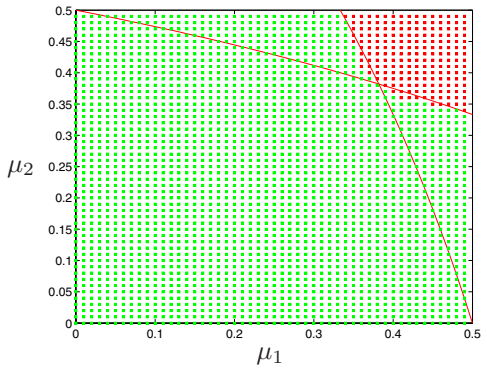

Fig. 5. Steady-state decision regions for $k_i = 1, k_c = .2$. Green represents (s, s), and red (h, h)

V. Conclusion

In satisficing game theory, it is important that players' have knowledge of each other's preferences. Indeed, this dependence is the primary strength of satisficing theory, permitting players to condition their utilities on the preferences of others. The learning dynamics described here therefore represent an important extension of satisficing game theory. It allows socially rational players who initially know nothing of each other's preferences to learn to act in a socially coherent way.

In applying the learning dynamics to the Stag Hunt, we found that players usually can estimate preferences sufficiently well that both players take the action that they would take if they had perfect knowledge. However, in some cases, players misjudge each other's actions and the dynamics converge to incorrect decision regions. While unfortunate, this result is a necessary consequence of the passive nature of the system. However, since players rely solely upon observation, no signaling is required, both reducing the overhead of an artificial learning system and ensuring that player's cannot attempt to deceive each other with false signals.

Finally, these results give the potential of applying the satisficing framework to a wider variety of circumstances. Through the learning dynamics, cooperation can emerge not only in the risk of failure, but also under complete uncertainty of others' preferences. We also note the possibility of running the evolutionary dynamics discussed in [5] in parallel with satisficing learning. Such a system would give a complete adaptive mechanism whereby players may simultaneously learn each other's preferences and adapt their own preferences in response to the surrounding community.

References

[1] J. von Neumann and O. Morgenstern, *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press, 1944, (2nd ed., 1947).

[2] J. F. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, pp. 289–295, 1951.

[3] W. C. Stirling, "Social Utility Functions, Part 1 — Theory," *IEEE Transactions on Systems, Man, and Cybernetics (Part C)*, vol. 35, no. 4, pp. 522–532, 2005.

[4] ——, *Satisficing Games and Decision Making: with applications to engineering and computer science*. Cambridge, UK: Cambridge Univ. Press, 2003.

[5] M. S. Nokleby and W. C. Stirling, "The stag hunt: A vehicle for evolutionary cooperation," in *Proceedings of The IEEE World Congress on Computational Intelligence*, Vancouver, BC, Canada, July 16–21, 2006.

[6] B. Skyrms, *The Stag Hunt and the Evolution of Social Structure*. Cambrige, UK: Cambrige Univ. Press, 2004.

[7] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Ca: Morgan Kaufmann, 1988.

[8] S. L. Lauritzen, *Graphical Models*. New York: Springer Verlag, 1996.

[9] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, 1973.

[10] W. Wu and J. Jian, "Essential equilibrium points of n-person non-cooperative games," *Sci. Sinica*, vol. 11, pp. 1307–22, 1962.

[11] E. Kohlberg and J. Mertens, "On the strategic stability of equilibria," *Econometrica*, vol. 54, pp. 1003–1037, 1986.

[12] J. W. Weibull, *Evolutionary Game Theory*. Cambridge, MA: MIT Press, 1995.