A Praxeology for Rational Negotiation

Wynn C. Stirling
Electrical & Computer Engineering Department
Brigham Young University, Provo, UT 84602 USA
voice: 801-378-7669 fax: 801-378-6586
email: wynn@ee.byu.edu

Todd K. Moon
Electrical & Computer Engineering Department
Utah State University, Logan, UT 84332 USA
voice: 435-797-2970 fax: 435-797-3054
email: todd.moon@ece.usu.edu

Abstract

Multi-agent artificial decision systems require a praxeology, or science of efficient action, that accommodates complex interactions between decision makers. Conventional praxeologies are built on the paradigm of rational choice, which comprises the two companion premises of totally-ordered preferences and individual rationality. Exclusive self-interest when negotiating, however, engenders a pessimistic and defensive attitude, and limits the ability of a decision maker to accommodate the interests of others, and therefore may unnecessarily constrain the negotiability of a decision maker, particularly in cooperative environments. This paper provides a distinct alternative to the hyperrationality of conventional rational choice by waiving reliance on the individual rationality premise and offering an approach to negotiatory decision making that is based on a well-defined mathematical notion of satisficing, or being good enough, that permits the modeling of complex interrelationships between agents, including cooperation, unselfishness, and altruism.

Introduction

Negotiation is a branch of multi-agent decision making that involves the opportunity for repeated interaction between independent entities as they attempt to reach a joint decision that is acceptable to all participants. But unless the interests of the decision makers are extremely compatible, achieving such a compromise will usually require them to be willing to consider lowering their standards of what is acceptable if they are to avert an impasse. For an agent to consider lowering its standards, it must be willing to relax the demand for the best possible outcome for itself, and instead be willing to settle for an outcome that is merely good enough, in deference to the interests of others. Defining what it means to be good enough, however, is much more subtle than defining what it means to be optimal, and any such definition must be firmly couched in and consistent with the decision maker's concept of rationality.

Rational Choice

Fundamental rationality requires a decision maker to choose between alternatives in a way that is consistent with its preferences. Consequently, before a rational decision is possible, a decision maker must have some way to order its preferences.

Definition 1 Let the symbols "$\succeq$" and "$\cong$" denote binary relationships meaning "is at least as good as" and "is equivalent to," respectively, between members of a set $\mathcal{X} = \{x, y, z, \dots\}$. The set $\mathcal{X}$ is *totally ordered* if relationships between elements of $\mathcal{X}$ are *reflexive* ($x \succeq x$), *antisymmetric* ($x \succeq y \;\&\; y \succeq x \implies x \cong y$), *transitive* ($x \succeq y \;\&\; y \succeq z \implies x \succeq z$), and *linear* (either $x \succeq y$ or $y \succeq x \; \forall \, x, y \in \mathcal{X}$). If the linearity condition is relaxed, then the set is *partially ordered.* □

Once in possession of a preference ordering, a rational decision maker must employ general principles that govern the way the orderings are to be used to formulate decision rules. Perhaps the most well-known principle is the classical economics hypothesis of (Bergson 1938) and (Samuelson 1948), which asserts that individual interests are fundamental; i.e., that social welfare is an aggregation of individual welfares. This hypothesis leads to the doctrine of *rational choice,* the favorite paradigm of conventional decision theory. Rational choice is based upon two premises.

P-1 *Total ordering:* a decision maker is in possession of a total preference ordering for all of its possible choices under all conditions (in multi-agent settings, this includes knowledge of the total orderings of all other participants).

P-2 *The principle of individual rationality*: a decision maker should make the best possible decision for itself, that is, it should optimize with respect to its own total ordering (in multi-agent settings, this ordering will be influenced by the preferences of others).

A praxeology, or science of efficient action, is the philosophical underpinning that governs the actions of a decision-making entity. Conventional praxeologies are founded on the paradigm of rational choice. For single-agent systems, this equates to optimization, which typically results in maximizing expected utility. For multi-agent systems, rational choice equates to equilibration: a joint decision is an equilibrium if, were any individual to change its decision unilaterally, it would decrease its own expected utility. Rational choice has a strong normative appeal; it tell us what exclusively self-interested decision makers should do, and is the praxeological basis for much of current artificial decision system synthesis methodology. The ratiocination for this approach, as expressed by Sandholm, is that each decision maker should

> maximize its own good without concern for the global good. Such self-interest naturally prevails in negotiations among independent businesses or individuals ... Therefore, the protocols must be designed using a *noncooperative, strategic* perspective: the main question is what social outcomes follow given a protocol which *guarantees that each agent's desired local strategy is best for that agent—and thus the agent will use it* (Sandholm 1999, pp. 201,202).

This rationale is consistent with the conventional game-theoretic notion that society should *not* be viewed as a generalized agent, or superplayer, who is capable of making choices on the basis of some sort of group-level welfare function. So doing, (Shubik 1982) argues, creates an "anthropomorphic trap" of failing to distinguish between group choices and group preferences.

Anthropomorphisms aside, it is far from obvious that exclusive self interest is the appropriate characterization of agent systems when coordinated behavior is desirable. Granted, it is possible under the individual rationality regime for a decision maker to suppress its own egoistic preferences in deference to others by redefining its utilities, but doing so is little more than a device to trick individual rationality into providing a response that can be interpreted as unselfish. Such an artifice provides only an indirect way to simulate socially useful attributes of cooperation, unselfishness, and altruism under a regime that is more naturally attuned to competition, exploitation, and avarice. Luce and Raiffa summarized the situation succinctly when they observed that

> general game theory seems to be in part a sociological theory which does not include any sociological assumptions ... it may be too much to ask that any sociology be derived from the single assumption of individual rationality (Luce & Raiffa 1957, p. 196).

Often, the most articulate advocates of a theory are also its most insightful critics. Perhaps the essence of this criticism is that rational choice does not provide for the ecological balance that a society must achieve if it is to accommodate the variety of relationships that may exist between agents and their environment. But achieving such a balance should not require fabrication of a superplayer to aggregate individual welfare into group welfare. What it may require, however, is reconsideration of the claim that rational choice provides the appropriate praxeology for synthesizing cooperative social systems.

State of the Art

There are many proposals for artificial negotiatory systems under the rational choice paradigm, bounded in various ways to account for limitations in knowledge, computational ability, and negotiation time. (Stahl 1977) and (Rubenstein 1982) propose models of alternating offers; these approaches are refined by (Shaked & Sutton 1984) to account for time constraints, and are further developed by (Kraus & Wilkenfeld 1991a; 1991b; 1993), (Kraus, Wilkenfeld, & Zlotkin 1995), and (Kraus 1996) to incorporate a time discount rate and to account for incomplete information via the introduction of a revelation mechanism. These approaches are based on a notion of *perfect equilibrium*, which is stronger Nash equilibrium is that it requires that an equilibrium must be induced at any stage of the negotiation process. Similar manifestations of bounded rationality occur with (Russell & Wefald 1991), who present a general framework for metareasoning via decision theory to define the utility of computation. Others have followed these same lines (see, for example, (Sandholm & Lesser 1997), (Zilberstein 1998), and (Kaufman 1990)), and yield optimal solutions according to performance criteria that is modified to account for resource limitations. Additional approaches to bounded rationality occur with (Oaksford & Chater 1994), who provide a rational analysis framework that accounts for environmental constraints regarding what is optimal behavior in a particular context. Another individual rationality-based approach is to involve market price mechanisms, as is done by (Wellman 1993; Mullen & Wellman 1995), resulting in a competition between agents in a context of information service provision. (Ephrati & Rosenschein 1996) use the Clarke Tax voting procedure to obtain the highest sum of utilities in an environment of truthful voting. (Wangermann & Stengel 1999) present a method of "principled negotiation" involving proposed changes to an original master plan as a means of finding a distributed optimal negotiated solution.

Another stream of research for the design of negotiatory systems is to rely more heavily upon heuristics than upon formal optimization procedures. The approach taken by Rosenschein and Zlotkin is to emphasize special compromise protocols involving pre-computed solutions to specific problems (Rosenschein & Zlotkin 1994; Zlotkin & Rosenschein 1996c; 1996b; 1996a). Formal models which describe the mental states of agents based upon representations of their beliefs, desires, intentions, and goals can be used for communicative agents (Cohen & Levesque 1990; Cohen, Morgan, & Pollack 1990; Kraus & Lehmann 1999; Lewis & Sycara 1993; Shoham 1993; Thomas *et al.* 1991; Wellman & Doyle 1991). In particular, Sycara develops a negotiation model that accounts for human cognitive characteristics, and models negotiation as an iterative process involving case-based learning and multi-attribute utilities (Sycara 1990; 1991). (Kraus, Sycara, & Evenchik 1998) provide logical argumentation models as an iterative process involving exchanges among agents to persuade each other and bring about a change of intentions. (Zeng & Sycara 1997; 1998) develop a negotiation framework that employs a Bayesian belief update learning process through which the agents update their beliefs about their opponent. (Durfee & Lesser 1989) advance a notion of partial global planning for distributed problem solving in an environment of uncertainty regarding knowledge and abilities.

The above approaches offer realistic ways to deal with the exigencies under which decisions must be made in the real world. They represent important advances in the theory of decision making, and their importance will increase as the scope of negotiatory decision making grows. They all appear, however to have a common theme, which is, that if a decision maker could maximize its own private utility subject to the constraints imposed by other agents, it should do so. Exclusive self-interest is a very simple concept. It is also a very limiting concept, since it justifies ignoring the preferences of others when ordering one's own preferences. The advantage of invoking exclusive self-interest is that it may drastically reduce the complexity of a model of the society. The price for doing so is the risk of compromising group interests when individual preferences dominate, or of distorting the real motives of the individuals when group interests dominate. The root of the problem, in both of these extreme cases, is the lack of a way to account for both group and individual interests in a seamless, consistent way.

Middle Ground

Rather than searching for or approximating a narrowly defined theoretical ideal, an alternative is to focus on an approach that, even though it may not aspire to such an ideal, is ecologically tuned to the environment in which the agents must function. If it is to function in a coordinative environment, it should not ignore the possibility of distinct group interests, yet it must respect individual interests. It should be flexible with respect to evaluations of what is acceptable, yet it must not abandon all qualitative measures of performance. Kreps seems to be seeking such an alternative when he observes that

> ... the real accomplishment will come in finding an interesting middle ground between hyperrational behaviour and too much dependence on *ad hoc* notions of similarity and strategic expectations. When and if such a middle ground is found, then we may have useful theories for dealing with situations in which the rules are somewhat ambiguous (Kreps 1990, p. 184).

Is there really some middle ground, or is the lacuna between strict rational choice and pure heuristics bridgeable only by forming hybrids of these extremes? If non-illusory middle ground does exist, few have staked claims to that turf. Literature involving rational choice (bounded or unbounded) is overwhelmingly vast, reflecting many decades of serious study. Likewise, heuristics, rule-based decision systems, and various *ad hoc* techniques are well-represented in the literature. Rationality paradigms that depart from these extremes or blends thereof, however, are not in substantial evidence. One who has made this attempt, however, is Slote (Slote 1989), who argues that it is not even necessary to define a notion of optimality in order to define a common sense notion of adequacy. He suggests that it is rational to choose something that is merely adequate rather than something that is best, and that moderation in the short run may actually be instrumentally optimal in the long run. Unfortunately, Slote does not metrize the notion of being adequate. It is far easier to quantify the the notion of bestness than it is to quantify the notion of adequacy. Striving for the best may be the most obvious way to use ordering information, but it is not the only way. This paper presents a notion of adequacy that is not an approximation to bestness—it is a distinct concept that admits a precise mathematical definition in terms of utility-like quantities. The motivation for pursuing this development is to soften the strict egoism of individual rationality and open the way for consideration of a more socially compatible view of rationality that does not rely upon optimization, heuristics, or hybrids of these extremes.

A New Praxeology

The assumption that a decision-maker possesses a total preference ordering that accounts for all possible combinations of choices for all agents under all conditions is a very strong condition, particularly when the number of possible outcomes is large. In multi-agent decision scenarios, individuals may not be able to comprehend, or even to care about, a full understanding of their environment. They may be concerned mostly about issues that are closest to them, either temporally, spatially, or functionally. A praxeology relevant to this situation must be able to accommodate preference orderings that may be limited to proper subsets of the community or to proper subsets of conditions that may obtain.

In societies that value cooperation, it is unlikely that the preferences of a given individual will be formed independently of the preferences of others. Knowledge about one agent's preferences may alter another agent's preferences. Such preferences are *conditioned on the preferences of others*. Individual rationality does not accommodate such conditioning. The only type of conditioning supported by individual rationality is for each agent to express its preferences conditioned on the choices of the others but not on their *preferences* about their choices. Each agent then computes its own expected utility as a function of the possible options of all agents, juxtaposes these expected utilities into a payoff array, and searches for an equilibrium. Although the equilibrium itself is governed by the utilities of all agents, the individual expected utilities that define the equilibrium do not consider the preferences of others. A praxeology for a complex society, however, should accommodate notions of cooperation, unselfishness, and even altruism. One way to do this is to permit the preferences (not just the choices) of decision makers to influence each other.

Tradeoffs

At present, there does not appear to be a body of theory that supports the systematic synthesis of multi-agent decision systems that does not rely upon the individual rationality premise. It is a platitude that decision makers should make the best choices possible, but we cannot rationally choose an option, even if we do not know of anything better, unless we know that it is good enough. Being good enough is the fundamental obligation of rational decision makers—being best is a bonus.

Perhaps the earliest notion of being "good enough" is Simon's concept of *satisficing*. His approach is to blend rational choice with heuristics by specifying aspiration levels of how good a solution might reasonably be achieved, and halting search for the optimum when the aspirations are met (Simon 1955; 1990; 1996). But it is difficult to establish good and practically attainable aspiration levels without first exploring the limits of what is possible, that is, without first identifying optimal solutions—the very procedure this notion of satisficing is designed to circumvent. Aspiration levels at least superficially establish minimum requirements, and specifying them for simple single-agent problems may be noncontroversial. But with multi-agent systems, interdependencies between decision makers can become complex, and aspiration levels can be conditional (what is satisfactory for me may depend upon what is satisfactory for you). The current state of affairs regarding aspiration levels does not address the problem of specifying them in multi-agent contexts. It may be that what is really needed is a notion of satisficing that does not depend upon arbitrary aspiration levels or stopping rules.

Let us replace the premise of individual rationality with a concept of being good enough that is distinct from being approximately best. Mathematically formalizing a concept of being good enough, however, is not as straightforward as optimizing or equilibrating. Being best is an absolute concept—it does not come in degrees. Being good enough, however, is not an absolute, and does come in degrees. Consequently, we must not demand a unique good-enough solution, but instead be willing to accept varying degrees of adequacy.

This paper proposes a notion for being good enough that is actually more primitive and yet more complicated to quantify than doing the best thing possible. It is a benefit-cost tradeoff paradigm of getting at least what one pays for. The reason it is more complicated to quantify is that it requires the application of two distinct metrics to be compared, whereas doing the best thing requires only one metric to be maximized. As a formalized means of decision making, this approach has appeared in at least two very different contexts: economics and epistemology—the former is intensely practical and concrete, the latter is intensely theoretical and abstract. Economists implemented the formal practice of benefit-cost analysis to evaluate the wisdom of implementing flood control policies (Pearce 1983). The usual procedure is to express all benefits and costs in monetary units and to sanction a proposition if the benefits are in excess of the estimated costs. The problem with this concept, however, is that the individual interests are aggregated into a single monolithic interest by comparing the total benefits with the total costs. Despite its flaws, benefit-cost analysis has proven to be a useful way to reduce a complex problem to a simpler, more manageable one. One of its chief virtues is its fundamental simplicity.

A more sophisticated notion of benefit-cost appears in philosophy. Building upon the American tradition of pragmatism fostered by Peirce, James, and Dewey, (Levi 1980) has developed a distinctive school of thought regarding the evolution of knowledge corpora. Unlike the conventional doctrine of expanding a knowledge corpus by adding information that has been justified as true, Levi proposes the more modest goal of avoiding error. This theory has been detailed elsewhere (see (Levi 1980; Stirling & Morrell 1991; Stirling & Goodrich 1999; Goodrich, Stirling, & Frost 1998; Stirling, Goodrich, & Packard 2001)). The gist is that, given the task of determining which, if any, of a set of propositions should be retained in an agent's knowledge corpus, the agent should evaluate each proposition on the basis of two distinct criteria—first, the credal, or subjective, probability of it being true, and second, the informational value[1] of rejecting it, that is, the degree to which discarding the option focuses attention on the kind of information that is demanded by the question. Thus, for an option to be admissible, it must be both believable and informative—all implausible or uninformative option should be rejected. Levi constructs an *expected epistemic utility function* and shows that it is the difference between credal probability and a constant (the index of caution) times another probability function, termed the informational-value-of-rejection probability. The set of options that maximizes this difference is the admissible set.

Single-Agent Satisficing

Levi's epistemology is to employ two separate and distinct orderings—one to characterize belief, the other to characterize value. This approach, originally developed for epistemological decision-making (committing to beliefs), may easily be adapted to the praxeological domain (taking action) by formulating praxeological analogs to the epistemological notions of truth and informational value. A natural analog for *truth* is *success*, in the sense of achieving the fundamental goals of taking action. To formulate an analog for informational value, observe that, just as the management of a finite amount of relevant information is important when inquiring after truth in the epistemological context, taking effective action requires the management of finite resources, such as conserving wealth, materials, energy, safety, or other assets. An apt praxeological analog to the informational value of rejection is the conservational value of rejection. Thus, the context of the decision problem changes from the epistemological issue of acquiring information while avoiding error to the praxeological issue of conserving resources while avoiding failure. To emphasize the context shift, the resulting utility function will be termed *praxeic utility*.

Let us refer to the degree of resource consumption as *rejectability* and require the rejectability function to conform to the axioms of probability. This new terminology emphasizes the semantic distinction of using the mathematics of probability in a non-conventional way. Thus, for a finite action space U, rejectability is expressed in terms of a mass function $p_R\colon U \to [0,1]$, such that $p_R(u) \geq 0$ for all $u \in U$ and $\sum_{u \in U} p_R(u) = 1$. Inefficient options (those with high resource consumption) should be highly rejectable; that is, if considerations of success are ignored, one should be prone to reject options that result in large costs, high energy consumption, exposure to hazard, etc. Normalizing p_R to be a mass function, termed the *rejectability mass function,* insures that the agent will have a unit of resource consumption

[1] Informational value, as used here, is distinct from the notion of "value of information" of conventional decision theory, which deals with the change in expected utility if uncertainty is reduced or eliminated from a decision problem.

to apportion among the elements of U. The function p_R is the dis-utility of consuming resources; that is, if $u \in U$ is rejected, then the agent conserves $p_R(u)$ worth of its unit of resources.

The degree that u contributes toward the avoidance of failure is the *selectability* of u. Let us define the *selectability mass function*, $p_S\colon U \to [0,1]$ as the normalized amount of success support associated with each $u \in U$. Suppose that implementing $u \in U$ would avoid failure. For any $A \subset U$, the utility of not rejecting A in the interest of avoiding failure is the indicator function $I_A(u) = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{otherwise} \end{cases}$. The *praxeic utility* of not rejecting A when u avoids failure is the convex combination of the utility of avoiding failure and the utility of conserving resources:

$$\phi(A,u) = \alpha I_A(u) + (1-\alpha)\left(1 - \sum_{v\in A} p_R(v)\right),$$

where $\alpha \in [0,1]$ is chosen to reflect the agent's personal weighting of these two desiderata—setting $\alpha = \frac{1}{2}$ associates equal concern for avoiding failure and conserving resources.

Generally, the decision-maker will not know precisely which u will avoid failure, and so must weight the utility for each u by the corresponding selectability, and sum over U to compute the expected praxeic utility.

$$\begin{aligned}\bar{\phi}(A) =& \sum_{u\in U}\left[\alpha I_A(u) + (1-\alpha)\left(1 - \sum_{v\in A} p_R(v)\right)\right] p_S(u) \\ =& \alpha \sum_{v\in A} p_S(v) - (1-\alpha)\sum_{v\in A} p_R(v) + (1-\alpha).\end{aligned}$$

Dividing by α and ignoring the constant term yields a more convenient but equivalent form:

$$\varphi(A) = \sum_{u\in A}\left[p_S(u) - q p_R(u)\right],$$

where $q = \frac{1-\alpha}{\alpha}$. The term q is the *index of caution*, and parameterizes the degree to which the decision maker is willing to accommodate increased costs to achieve success. An equivalent way of viewing this parameter is as an index of boldness, characterizing the degree to which the decision maker is willing risk rejecting successful options in the interest of conserving resources. Nominally, $q = 1$, which attributes equal weight to success and resource conservation interests.

Definition 2 A decision maker is *satisficingly rational* if it chooses an option for which the selectability is greater than or equal to the index of caution times rejectability. □

We adopt this notion of satisficing as the mathematical definition of being *good enough*. The largest set of satisficing options is the *satisficing set*:

$$\Sigma_q = \arg\max_{A\subset U} \varphi(A) = \{u \in U\colon p_S(u) \geq q p_R(u)\}. \quad (1)$$

Notice that (1) is in the form of a likelihood ratio test, since the selectability and rejectability functions are mass functions. Equation (1) is the *praxeic likelihood ratio test* (PLRT).

This concept of satisficing does not require that the set of good-enough solution be non-empty. If it is non-empty, however, fundamental consistency requires that the best solution, if it exists (under the same criteria), must be a member of that set.

Theorem 1 (a) $q \leq 1 \implies \Sigma_q \neq \emptyset$. (b) *If* $\Sigma_q \neq \emptyset$ *then there exists an optimality criterion that is consistent with* p_S *and* p_R *such that the optimal choice is an element of* Σ_q.

Proof (a) If $\Sigma_q = \emptyset$, then $p_S(u) < q p_R(u)\ \forall u \in U$, and hence $1 = \sum_{u\in U} p_S(u) < q \sum_{u\in U} p_R(u) = q$, a contradiction. (b) Define $J(u) = p_S(u) - q p_R(u)$, and let $u^* = \arg\max_{u\in U} J(u)$. But $J(u) \geq 0\ \forall\ u \in \Sigma_q$, and since $\Sigma_q \neq \emptyset$, $J(u^*) \geq \max_{u\in\Sigma_q} J(u) \geq 0$, which implies $u^* \in \Sigma_q$. □

Individual rationality requires that a single ordering be defined for each agent, and that all of its options be rank-ordered with the best one surviving. This is an inter-option, or *extrinsic*, comparison, since it requires the evaluation of an option with respect to quantities other than those associated with itself (namely, ranking of all other options). The PLRT provides another way to order, using two preference orderings: one to characterize the desirable, or selectable, attributes of the options, while the other characterizes the undesirable, or rejectable, attributes, and compares these two orderings for each option, yielding a binary decision (reject or retain) for each. Such intra-option comparisons are *intrinsic*, since they do not require the evaluation of an option with respect to quantities other than those associated with itself. This intrinsic comparison identifies all options for which the benefit derived from implementing them is at least as great as the cost incurred. This notion of satisficing is compatible with Simon's original notion in that it addresses exactly the same issue that motivated Simon—to identify options that are good enough by directly comparing attributes of options. This notion differs only in the standard used for comparison. The standard for satisficing *à la* Simon, as with individual rationality in general, is imposed from without—it is extrinsic, since it relies upon external information (the aspiration level). In contrast, the standard for satisficing *à la* the PLRT is set up from within—it is intrinsic, and compares the positive attributes to the negative attributes of each option.

Intrinsic satisficing may be blended with Simon's extrinsic approach by specifying the aspiration level via the PLRT, rather than a fixed threshold. Searching then may stop when the first element of Σ_q is identified. On the other hand, searching may continue to exhaustion, and additional ordering constraints can be imposed on the elements of Σ_q to identify an optimal solution (for example, see (Stirling, Goodrich, & Packard 2001)).

Extension to Multiple Agents

Individual satisficing is defined in terms of univariate selectability and rejectability mass functions that provide separate orderings for success and resource consumption, respectively. Just as univariate probability theory extends to multivariate probability theory, we may extend single-agent selectability and rejectability mass functions to the multi-agent case by defining a multi-agent (joint) selectability mass function to characterize group selectability and a joint rejectability function to characterize group rejectability. Given such functions, we may define a concept of multi-agent satisficing, or jointly satisficing, as follows:

Definition 3 A decision-making group is *jointly satisficingly rational* if the members of the group choose a vector of options for which joint selectability is greater than or equal to the index or caution times joint rejectability. □

For this definition to be useful we must be able to construct the joint selectability and rejectability functions in a way that accommodates partial preference orderings and conditional preferences. To establish this utility, we first introduce the notion of interdependence and define a satisficing game. We then describe how the interdependence function can be constructed from local orderings, leading to emergent total preference orderings.

Interdependence

An act by any member of a multi-agent system has possible ramifications throughout the entire community. Some agents may be benefited by the act, some may be damaged, and some may be unaffected. Furthermore, although the single agent may perform the act in its own interest, or for the benefit (or detriment) of other agents, the act is usually not implemented free of cost. Resources are expended, or risk is taken, or some other cost, penalty, or unpleasant consequence is incurred by the agent itself or by other agents. Although these undesirable consequences may be defined independently from the benefits, the measures associated with benefits and costs cannot be specified independently of each others due to the possibility of interaction. A critical aspect of modeling the behavior of such a society, therefore, is the means of representing the interdependence of both positive and negative consequences of all possible joint actions that could be undertaken.

Definition 4 Let $\{X_1, \dots, X_N\}$ be an N-member multi-agent system. A *mixture*[2] is any subset of agents considered in terms of their interaction with each other, exclusively of possible interactions with other agents not in the subset.

A *selectability mixture,* denoted $\mathcal{S} = S_{i_1} \dots S_{i_k}$, is a mixture consisting of agents $X_{i_1}, \dots X_{i_k}$ being considered from the point of view of success. The *joint selectability mixture* is the selectability mixture consisting of all agents in the system, denoted $\mathbf{S} = S_1 \dots S_N$.

A *rejectability mixture,* denoted $\mathcal{R} = R_{j_1} \dots R_{j_\ell}$, is a mixture consisting of agents $X_{j_1}, \dots X_{j_\ell}$ being considered from the point of view of resource consumption. The *joint rejectability mixture* is the rejectability mixture consisting of all agents in the system, denoted $\mathbf{R} = R_1 \dots R_N$.

An *intermixture* is the concatenation of a selectability mixture and a rejectability mixture, and is denoted $\mathcal{SR} = S_{i_1} \dots S_{i_k} R_{j_1} \dots R_{j_\ell}$. The *joint intermixture* is the concatenation of the joint selectability and joint rejectability mixtures, and is denoted $\mathbf{SR} = S_1 \dots S_N R_1 \dots R_N$. □

[2]Not to be confused with a mixture of distributions, which is a convex combination of probability distributions.

Definition 5 Let U_i be the action space for X_i, $i = 1, \dots, N$. The *product action space,* denoted $\mathbf{U} = U_1 \times \cdots \times U_N$ is the set of all N-tuples $\mathbf{u} = (u_1, \dots, u_N)$ where $u_i \in U_i$. The *selectability action space* associated with a selectability mixture $\mathcal{S} = S_{i_1} \dots S_{i_k}$ is the product space $\mathbf{U}_\mathcal{S} = U_{i_1} \times \cdots \times U_{i_k}$. The *rejectability action space* associated with a rejectability mixture $\mathcal{R} = R_{j_1} \dots R_{j_\ell}$ is the product space $\mathbf{U}_\mathcal{R} = U_{j_1} \times \cdots \times U_{j_\ell}$. The *interaction space* associated with an intermixture $\mathcal{SR} = S_{i_1} \cdots S_{i_k} R_{j_1} \cdots R_{j_\ell}$ is the product space $\mathbf{U}_{\mathcal{SR}} = \mathbf{U}_\mathcal{S} \times \mathbf{U}_\mathcal{R} = U_{i_1} \times \cdots \times U_{i_k} \times U_{j_1} \times \cdots \times U_{j_\ell}$. The *joint interaction space* is $\mathbf{U_{SR}} = \mathbf{U} \times \mathbf{U}$. □

Definition 6 A *selectability mass function* (smf) for the mixture $\mathcal{S} = \{S_{i_1} \dots S_{i_k}\}$ is a mass function denoted $p_\mathcal{S} = p_{S_{i_1}, \dots, S_{i_k}} \colon \mathbf{U}_\mathcal{S} \to [0,1]$. The *joint smf* is an smf for $\mathbf{S}$, denoted $p_\mathbf{S}$.

A *rejectability mass function* (rmf) for the mixture $\mathcal{R} = \{R_{j_1}, \dots, R_{j_\ell}\}$ is a mass function denoted $p_\mathcal{R} = p_{R_{j_1}, \dots, R_{j_\ell}} \colon \mathbf{U}_\mathcal{R} \to [0,1]$. The *joint rmf* is a rmf for $\mathbf{R}$, denoted $p_\mathbf{R}$.

An *interdependence mass function* (IMF) for the intermixture $\mathcal{SR} = \{S_{i_1} \dots S_{i_k} R_{j_1} \dots R_{j_\ell}\}$ is a mass function denoted $p_{\mathcal{SR}} = p_{S_{i_1}, \dots, S_{i_k}, R_{j_1}, \dots, R_{j_\ell}} \colon \mathbf{U}_\mathcal{S} \times \mathbf{U}_\mathcal{R} \to [0,1]$. The *joint IMF* is an IMF for $\mathbf{SR}$, denoted $p_\mathbf{SR}$. □

Let $\mathbf{v} \in \mathbf{U}_\mathcal{S}$ and $\mathbf{w} \in \mathbf{U}_\mathcal{R}$ be two option vectors. Then $p_{\mathcal{S},\mathcal{R}}(\mathbf{v}, \mathbf{w})$ is a representation of the success support associated with $\mathbf{v}$ and the resource consumption associated with $\mathbf{w}$ when the two option vectors are viewed simultaneously. In other words, $p_{\mathcal{S},\mathcal{R}}(\mathbf{v}, \mathbf{w})$ is the mass associated with selecting $\mathbf{v}$ in the interest of success and rejecting $\mathbf{w}$ in the interest of conserving resources.

Satisficing Games

The interdependence function incorporates all of the information relevant to the multi-agent decision problem. From this function we may derive the joint selectability and rejectability marginals as

$$p_\mathbf{S}(\mathbf{u}) = \sum_{\mathbf{v} \in \mathbf{U}} p_{\mathbf{S},\mathbf{R}}(\mathbf{u}, \mathbf{v}) \tag{2}$$

$$p_\mathbf{R}(\mathbf{v}) = \sum_{\mathbf{u} \in \mathbf{U}} p_{\mathbf{S},\mathbf{R}}(\mathbf{u}, \mathbf{v}) \tag{3}$$

for all $(\mathbf{u}, \mathbf{v}) \in \mathbf{U} \times \mathbf{U}$. Once these quantities are in place, a satisficing game can be formally defined.

Definition 7 A *satisficing game* for a set of decision makers $\{X_1, \dots, X_N\}$, is a triple $\{\mathbf{U}, p_\mathbf{S}, p_\mathbf{R}\}$, where $\mathbf{U}$ is a joint action space, $p_\mathbf{S}$ is the joint selectability function, and $p_\mathbf{R}$ is the joint rejectability function. The *joint solution* to a satisficing game with index of caution q is the set

$$\Sigma_q = \{\mathbf{u} \in \mathbf{U} \colon p_\mathbf{S}(\mathbf{u}) \geq q p_\mathbf{R}(\mathbf{u})\}. \tag{4}$$

$\boldsymbol{\Sigma}_q$ is termed the *joint satisficing* set, and elements of $\boldsymbol{\Sigma}_q$ are *jointly satisficing actions*. Equation (4) is the *joint praxeic likelihood ratio test* (JPLRT). □

The JPLRT establishes group preferences and identifies the joint option vectors that are satisficing from the group perspective. The marginal selectability and rejectability mass functions for each X_i may be obtained from (2) and (3), yielding:

$$p_{S_i}(u_i) = \sum_{\substack{u_j \in U_j \\ j \neq i}} p_{S_1,\ldots,S_N}(u_1,\ldots,u_N) \quad (5)$$

$$p_{R_i}(u_i) = \sum_{\substack{u_j \in U_j \\ j \neq i}} p_{R_1,\ldots,R_N}(u_1,\ldots,u_N). \quad (6)$$

Definition 8 The *individual solutions* to the satisficing game $\{\mathbf{U}, p_{\mathbf{S}}, p_{\mathbf{R}}\}$ are the sets

$$\Sigma_q^i = \{u_i \in U_i \colon p_{S_i}(u_i) \geq q p_{R_i}(u_i)\}, \quad (7)$$

where p_{S_i} and p_{R_i} are given by (5) and (6), respectively, for $i = 1, \ldots, N$. The product of the individually satisficing sets is the *satisficing rectangle*:

$$\mathfrak{R}_q = \Sigma_q^1 \times \cdots \times \Sigma_q^N = \{(u_1, \ldots, u_N) \colon u_i \in \Sigma_q^i\}.$$

□

It remains to determine the relationship between the jointly satisficing set $\boldsymbol{\Sigma}_q$ and the individually satisficing sets, $\Sigma_q^i, i = 1, \ldots, N$. Unfortunately, it is not generally true that either $\boldsymbol{\Sigma}_q \subset \mathfrak{R}_q$ or $\mathfrak{R}_q \subset \boldsymbol{\Sigma}_q$. The following result, however, is very useful.

Theorem 2 (The Negotiation Theorem) *If u_i is individually satisficing for X_i, that is, if $u_i \in \Sigma_q^i$, then it must be the ith element of some jointly satisficing vector $\mathbf{u} \in \boldsymbol{\Sigma}_q$.*

Proof This theorem is proven by establishing the contrapositive, namely, that if u_i is not the ith element of any $\mathbf{u} \in \boldsymbol{\Sigma}_q$, then $u_i \notin \Sigma_q^i$. Without loss of generality, let $i = 1$. By hypothesis, $p_{\mathbf{S}}(u_1, \mathbf{v}) < q p_{\mathbf{R}}(u_1, \mathbf{v})$ for all $\mathbf{v} \in U_2 \times \cdots \times U_N$, so $p_{S_1}(u_1) = \sum_{\mathbf{v}} p_{\mathbf{S}}(u_1, \mathbf{v}) < q \sum_{\mathbf{v}} p_{\mathbf{R}}(u_1, \mathbf{v}) = q p_{R_1}(u_1)$, hence $u_1 \notin \Sigma_q^1$. □

The content of this theorem is that no one is ever completely frozen out of a deal—every decision maker has, from its own perspective, a seat at the negotiating table. This is perhaps the weakest condition under which negotiations are possible. If $\boldsymbol{\Sigma}_q \cap \mathfrak{R}_q$ is empty, then there are no jointly satisficing options that are also individually satisficing for all players for the given value of q. The following corollary, whose proof is trivial and is omitted, addresses this situation.

Corollary 1 *There exists an index or caution value $q_0 \in [0, 1]$ such that $\boldsymbol{\Sigma}_{q_0} \cap \mathfrak{R}_{q_0} \neq \emptyset$.*

Thus, if the players are each willing to lower their standards sufficiently by decreasing the index of caution, q, they may eventually reach a compromise that is both jointly and individually satisficing, according to a reduced level of what it means to be good enough. The parameter q_0 is a measure of how much they must be willing to compromise to avoid an impasse. Note that willingness to lower one's standards is not total capitulation, since the participants are able to control the degree of compromise by setting a limit on how small of a value of q they can tolerate. Thus, a controlled amount of altruism is possible with this formulation. But, if any player's limit is reached without a mutual agreement being obtained, the game has reached an impasse.

It may be observed that the negotiation theorem does not provide for solutions which are both individually and jointly satisficing for all agents. This requires separate efforts at coordination in an active process of working toward an accord. This process is explored in (Moon & Stirling 2001).

Synthesis

The joint IMF provides a complete description of the individual and interagent relationships in terms of their positive and negative consequences, and provides a total ordering for both selectability and rejectability for the entire community as well as for each individual. Basing a praxeology on the IMF does not, at first glance, however, appear to conform to the requirement to accommodate partial orderings, but first glances can be misleading. Fortunately, the IMF, based as it is on the mathematics of probability theory, can draw upon a fundamental property of that theory, namely, the law of compound probability, to simplify its construction.

The law of compound probability says that joint probabilities can be constructed from conditional probabilities and marginal probabilities. For example, we may construct a joint probability mass function $p_{X,Y}(x, y)$ from the conditional mass function $p_{X|Y}(x|y)$ and the marginal $p_Y(y)$ according to Bayes rule, yielding $p_{X,Y}(x, y) = p_{X|Y}(x|y) p_Y(y)$. This relationship may be extended to the general multivariate case by repeated applications, yielding what is often termed the *chain rule*.

Definition 9 Given an intermixture $\mathcal{SR} = S_{i_1} \ldots S_{i_k} R_{j_1} \ldots R_{j_\ell}$, a *subintermixture* of $\mathcal{SR}$ is an intermixture formed by concatenating subsets of $\mathcal{S}$ and $\mathcal{R}$: $\mathcal{S}_1\mathcal{R}_1 = S_{i_{p_1}} \ldots S_{i_{p_q}} R_{j_{r_1}} \ldots R_{j_{r_s}}$, where $\{i_{p_1}, \ldots, i_{p_q}\} \subset \{i_1, \ldots, i_k\}$ and $\{j_{r_1}, \ldots, j_{r_s}\} \subset \{j_1, \ldots, j_\ell\}$. The notation $\mathcal{S}_1\mathcal{R}_1 \subset \mathcal{SR}$ indicates that $\mathcal{S}_1\mathcal{R}_1$ is a subintermixture of $\mathcal{SR}$.

The *$\mathcal{SR}$-complementary subintermixture* associated with a subintermixture $\mathcal{S}_1\mathcal{R}_1$ of an intermixture $\mathcal{SR}$, denoted $\mathcal{SR} \backslash \mathcal{S}_1\mathcal{R}_1$, is an intermixture created by concatenating the selectability and rejectability mixtures formed by the relative compliments of $\mathcal{S}_1$ and $\mathcal{R}_1$. Clearly, $\mathcal{SR} \backslash \mathcal{S}_1\mathcal{R}_1 \subset \mathcal{SR}$. $\mathcal{SR}$ is the union of $\mathcal{SR} \backslash \mathcal{S}_1\mathcal{R}_1$ and $\mathcal{S}_1\mathcal{R}_1$, denoted $\mathcal{SR} = \mathcal{SR} \backslash \mathcal{S}_1\mathcal{R}_1 \cup \mathcal{S}_1\mathcal{R}_1$. □

Definition 10 Let $\mathcal{SR}$ be an intermixture with subintermixture $\mathcal{S}_1\mathcal{R}_1$. A *conditional interdependence mass function*, denoted $p_{\mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1|\mathcal{S}_1\mathcal{R}_1}$, is a mapping of $(\mathbf{U}_{\mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1} \times$

$\mathbf{U}_{\mathcal{S}_1\mathcal{R}_1}$) into $[0,1]$ such that, for every $\mathbf{v} \in \mathbf{U}_{\mathcal{S}_1\mathcal{R}_1}$, $p_{\mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1|\mathcal{S}_1\mathcal{R}_1}(\cdot|\mathbf{v})$ is a mass function on $\mathbf{U}_{\mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1}$. □

All conditional interdependence mass functions must be be consistent with interdependence mass functions. That is, for $\mathcal{SR}$ an arbitrary intermixture with subintermixture $\mathcal{S}_1\mathcal{R}_1$ with $\mathbf{w} \in \mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1$ and $\mathbf{v} \in \mathcal{S}_1\mathcal{R}_1$, Bayes rule requires that

$$p_{\mathcal{S},\mathcal{R}}(\mathbf{v},\mathbf{w}) = p_{\mathcal{SR}\backslash\mathcal{S}_1\mathcal{R}_1|\mathcal{S}_1\mathcal{R}_1}(\mathbf{w}|\mathbf{v}) \cdot p_{\mathcal{S}_1\mathcal{R}_1}(\mathbf{v}). \quad (8)$$

This is the chain rule applied to intermixtures. Repeated applications of the chain rule provides a way to construct global behavior from local behavioral relationships. To illustrate, let $\{X_1, X_2, X_3\}$ be a multi-agent system and let $\mathcal{S} = S_1S_2$ and $\mathcal{R} = R_3$. Then $\mathcal{SR} = S_1S_2R_3$ and $\mathbf{SR}\backslash\mathcal{SR} = S_3R_1R_2$. The IMF is

$$p_{S_1,S_2,S_3,R_1,R_2,R_3}(v_1,v_2,v_3,w_1,w_2,w_3) =$$
$$p_{S_3,R_1,R_2|S_1,S_2,R_3}(v_3,w_1,w_2|v_1,v_2,w_3) \cdot p_{S_1,S_2,R_3}(v_1,v_2,w_3).$$

Now let $\mathcal{S}_1 = S_1$ be a subintermixture of $S_1S_2R_3$, so that $\mathcal{SR}\backslash\mathcal{S}_1 = S_2R_3$. We may apply the chain rule to this subintermixture to obtain

$$p_{S_1,S_2,R_3}(v_1,v_2,w_3) = p_{S_1|S_2,R_3}(v_1|v_2,w_3) \cdot p_{S_2,R_3}(v_2,w_3),$$

yielding

$$\begin{aligned} p_{S_1,S_2,S_3,R_1,R_2,R_3}(v_1,v_2,v_3,w_1,w_2,w_3) &= \\ p_{S_3,R_1,R_2|S_1,S_2,R_3}&(v_3,w_1,w_2|v_1,v_2,w_3) \\ &\cdot p_{S_1|S_2,R_3}(v_1|v_2,w_3) \cdot p_{S_2,R_3}(v_2,w_3). \quad (9) \end{aligned}$$

The term $p_{S_3,R_1,R_2|S_1,S_2,R_3}(v_3,w_1,w_2|v_1,v_2,w_3)$ is the conditional selectability/rejectability associated with X_3 selecting v_3, X_1 rejecting w_1, and X_2 rejecting w_2, given that X_1 prefers to select v_1, X_2 prefers to select v_2, and X_3 prefers to reject w_3; $p_{S_1|S_2,R_3}(v_1|v_2,w_3)$ characterizes X_1's selectability for v_1 given X_2 prefers to select v_2 and X_3 prefers to reject w_3; and $p_{S_2,R_3}(v_2,w_3)$ is the joint selectability/rejectability of X_2 selecting v_2 and X_3 rejecting w_3. The various terms of this factorization may often be simplified further. For example, suppose that X_1 is indifferent to X_3's rejectability posture, in which case we may simplify $p_{S_1|S_2,R_3}(v_1|v_2,w_3)$ to become $p_{S_1|S_2}(v_1|v_2)$.

Clearly, there are many ways to factor the interdependence function according to the chain rule. The design issue, however, is to implement a factorization that allows the desired local interdependencies to be expressed through the appropriate conditional interdependencies. The construction of the interdependence function is highly application dependent, and there is no general algorithm or procedure that a designer should follow for its synthesis. There are, however, some general guidelines for the construction of interdependence functions.

1. Form operational definitions of selectability and rejectability for individuals or groups, as appropriate from the context of the problem.
2. Identify the local orderings that are desirable, and map these into conditional selectability and rejectability functions.
3. Factor the interdependence function such that the desired conditional selectability/rejectability relationships are products in the factorization.
4. Eliminate all irrelevant interdependencies in the factors.

Meso-Emergence

Although each of the conditional mass functions in the factorization of the interdependence function is a total ordering, it is a *local* total ordering, and involves only a subset of agents and concerns. Each of these local total orderings is only a partial ordering, however, if viewed from the global, or community-wide, perspective, since orderings are not defined for all possible option vectors. By combining such local total orderings together according to the chain rule, a global total ordering emerges. The joint selectability and rejectability mass functions then characterize emergent global behavior, and the individual selectability and rejectability marginals characterize emergent individual behavior. Thus, both individual and group behavior emerge as consequences of local conditional interests that propagate throughout the community from the interdependent local to the interdependent global and from the conditional to the unconditional.

Synthesizing the IMF exploits an emergence property that is quite different from the temporal, or evolutionary, emergence that can occur with repeated play games. To differentiate these two types of emergence, let us refer to the former as *spatial* emergence. Temporal emergence is an inter-game phenomenon that produces relationships between agents with repeated play as time propagates, and spatial emergence is an intra-game phenomenon that produces relationships between agents as interests propagate through the agent system with single-play. Perhaps the most common example of spatial emergence is the *micro-to-macro*, or *bottom-up* phenomenon of group behavior emerging as a consequence of individual interests, as occurs with social choice theory (Sen 1979; Sandholm 1999) and with evolutionary games (Axelrod 1997; Weibull 1995). A second approach is a macro-to-micro or *top-down* approach, where individual behaviors emerge as a consequence of group interests. Satisficing praxeology accommodates both of these approaches. It also points to a third approach, that of an *inside-out*, or meso-to-micro/macro view, where intermediate-level conditional preferences propagate up to the group level and down to the individual level. Let us term this type of spatial emergence *meso-emergence*.

The conditional selectability and rejectability mass functions are constructed as functions of the preferences of the other agents. For example, the local total ordering function $p_{S_1|S_2}(\cdot|v_2)$ characterizes X_1's ordering of its selectability preferences given that X_2 prefers v_2. This structure permits X_1 to ascribe some weight to X_2's interests without requiring X_1 to abandon its own interests in deference to X_2. By adjusting these weights, X_1 may control the degree two which it is willing to compromise its egoistic values to accommodate X_2.

Discussion

The group decision problem has perplexed researchers for decades. As (Raiffa 1968, pp. 233–237) put it over thirty years ago, "I find myself in that uncomfortable position in which the more I think the more confused I become." The source of Raiffa's concern, it seems, is that it is difficult to reconcile the notion of individual rationality with the belief that "somehow the group entity is more than the totality of its members." Yet, researchers have steadfastly and justifiably refused to consider the group entity itself as a decision-making superplayer.

Satisficing game theory offers a way to account for the group entity without the fabrication of a superplayer. This accounting is done through the conditional relationships that are expressed through the interdependence function due to its mathematical structure as a probability (but not with the usual semantics of randomness). Just as the a joint probability function is more than the totality of the marginals, the interdependence function is more than the totality of the individual selectability and rejectability functions. It is only in the case of stochastic independence that a joint distribution can be constructed from the marginal distributions, and it is only in the case of complete inter-independence that group welfare can be expressed in terms of the welfare of the individuals.

The current literature on negotiation concentrates heavily on ways to obtain just-in-time negotiated solutions that can be accomplished within real-time computational constraints, but it does so *primarily from the point of view of individual rationality*. There is no reason, however, to limit consideration to that perspective. This paper is an invitation to expand to a broader perspective, and consider dealing with the exigencies of practical decision making in the light of satisficing game theory as well as with conventional theory.

Negotiation under (bounded or unbounded) rational choice requires the decision maker to attempt to maximize its own benefit. This is a valid, and perhaps the only reliable, paradigm in extremely conflictive environments, such as zero-sum games, but when the opportunity for cooperation exists, the rational choice paradigm is overly pessimistic and unnecessarily limits the scope of negotiation.

The appeal of optimization, no matter now approximate, is a strongly entrenched attitude that dominates current decision making practice. There is great comfort in following traditional paths, especially when those paths are founded on such a rich and enduring tradition as rational choice affords. But when synthesizing an artificial negotiatory system, the designer has the opportunity to impose upon the agents a more socially accommodating paradigm. The satisficing game theory presented in this paper provides a sociological decision-making mechanism that seamlessly accounts for group and individual interests, and provides a rich framework for negotiation to occur between agents who share common interests and who are willing to give deference to each other. Rather than depending upon the non-cooperative equilibria defined (even if only approximately) by individual-benefit saddle points, this alternative may lead to the more socially realistic and valuable equilibria of shared interests and acceptable compromises.

References

Axelrod, R. 1997. *The Complexity of Cooperation*. Princeton, NJ: Princeton Univ. Press.

Bergson, A. 1938. A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52:310–334.

Cohen, P., and Levesque, H. 1990. Intention is choice with commitment. *Artificial Intelligence* 42:262–310.

Cohen, P.; Morgan, J.; and Pollack, M. E., eds. 1990. *Intentions in Communication*. Cambridge, MA: MIT Press.

Durfee, E. H., and Lesser, V. R. 1989. Negotiating task decomposition and allocation using partial global planning. In Gasser, L., and Huhns, M. N., eds., *Distributed Artificial Intelligence*, volume II. London: Pitman/Mortan Kaufmann. 229–244.

Ephrati, E., and Rosenschein, J. S. 1996. Deriving consensus in multiagent systems. *Artificial Intelligence* 87(1–2):21–74.

Goodrich, M. A.; Stirling, W. C.; and Frost, R. L. 1998. A theory of satisficing decision and control. *IEEE Trans. Systems, Man, Cybernet.* 28(6):763–779.

Kaufman, B. E. 1990. A new theory of satisficing. *The Journal of Behavorial Economics* 19(1):35–51.

Kraus, S., and Lehmann, D. 1999. Knowledge, belief and time. *Theoretical Computer Science* 58:155–174.

Kraus, S., and Wilkenfeld, J. 1991a. The function of time in cooperative negotiations. In *Proceedings of AAAI-91*, 179–184.

Kraus, S., and Wilkenfeld, J. 1991b. Negotiations over time in a mutiagent environment. In *Proceedings of IJCAI-91*, 56–61.

Kraus, S., and Wilkenfeld, J. 1993. A strategic negotiations model with applications to an international crisis. *IEEE Transactions on Systems, Man, and Cybernetics* 32(1):313–323.

Kraus, S.; Sycara, K.; and Evenchik, A. 1998. Reaching agreements thourh argumentation: a logical model and implementation. *Artificial Intelligence Journal* 104(1–2):1–69.

Kraus, S.; Wilkenfeld, J.; and Zlotkin, G. 1995. Multiagent negotiation under time constraints. *Artificial Intelligence* 75(2):297–345.

Kraus, S. 1996. Beliefs, time and incomplete information in multiple encounter negotiations among autonomous agents. *Annals of Mathematics and Artificial Intelligence* 20(1–4):111–159.

Kreps, D. M. 1990. *Game Theory and Economic Modelling*. Oxford: Clarendon Press.

Levi, I. 1980. *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

Lewis, M., and Sycara, K. 1993. Reaching informed agreement in multi-specialist cooperation. *Group Decision and Negotiation* 2(3):279–300.

Luce, R. D., and Raiffa, H. 1957. *Games and Decisions*. New York: John Wiley.

Moon, T. K., and Stirling, W. C. 2001. Satisficing negotiation for resource allocation with disputed resources. In *AAAI Symposiusm on Negotiation Methods for Autonomous Cooperative Systems.* (submitted).

Mullen, T., and Wellman, M. 1995. A simple computational market for network information services. In *Proc. of the First International Conference on Multiagent Systems*, 283–289.

Oaksford, M., and Chater, N. 1994. A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–631.

Pearce, D. W. 1983. *Cost-Benefit Analysis.* London: Macmillan, second edition.

Raiffa, H. 1968. *Decision Analysis.* Reading, MA: Addison-Wesley.

Rosenschein, J., and Zlotkin, G. 1994. *Rules of Encounter.* Cambridge, MA: MIT Press.

Rubenstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50(1):97–109.

Russell, S., and Wefald, E. 1991. Principles of metareasoning. *Artificial Intelligence* 49:361–395.

Samuelson, P. A. 1948. *Foundations of Economic Analysis.* Cambridge, MA: Harvard University Press.

Sandholm, T., and Lesser, V. 1997. Coalitions among computationally bounded agents. *Artificial Intelligence* 94(1):99–137.

Sandholm, T. W. 1999. Distributed rational decision making. In Weiss, G., ed., *Multiagent Systems.* Cambridge, MA: MIT Press. chapter 5, 201–258.

Sen, A. K. 1979. *Collective Choice and Social Welfare.* Amsterdam: North-Holland.

Shaked, A., and Sutton, J. 1984. Involuntary unemployment as a perfect equilibrium in a bargaining model. *Econometrica* 52(6):1351–1364.

Shoham, Y. 1993. Agent oriented programming. *Artificial Intelligence* 60(1):51–92.

Shubik, M. 1982. *Game Theory in the Social Sciences.* Cambridge, MA: MIT Press.

Simon, H. A. 1955. A behavioral model of rational choice. *Quart. J. Econ.* 59:99–118.

Simon, H. A. 1990. Invariants of human behavior. *Annu. Rev. Psychol.* 41:1–19.

Simon, H. A. 1996. *The Sciences of the Artificial.* Cambridge, MA: MIT Press, third edition.

Slote, M. 1989. *Beyond Optimizing.* Cambridge, MA: Harvard Univ. Press.

Stahl, I. 1977. An n-person bargaining game in an extensive form. In Henn, R., and Moeschlin, O., eds., *Lecture Notes in Economics and Mathematical Systems No. 141.* Berlin: Springer-Verlag.

Stirling, W. C., and Goodrich, M. A. 1999. Satisficing games. *Information Sciences* 114:255–280.

Stirling, W. C., and Morrell, D. R. 1991. Convex bayes decision theory. *IEEE Trans. Systems, Man, Cybernet.* 21(1):173–183.

Stirling, W. C.; Goodrich, M. A.; and Packard, D. J. 2001. Satisficing equilibria: A non-classical approach to games and decisions. *Autonomous Agents and Multi-Agent Systems Journal.* Forthcoming.

Sycara, K. 1990. Persuasive argumentation in negotiation. *Theory and Decision* 28:203–242.

Sycara, K. 1991. Problem restructuring in negotiation. *Management Science* 37(10):1248–1268.

Thomas, B.; Shoham, Y.; Schwartz, A.; and Kraus, S. 1991. Preliminary thoughts on an agent description lauguage. *International Journal of Intelligent Systems* 6(5):497–508.

Wangermann, J. P., and Stengel, R. F. 1999. Optimization and coordination of multiagent systems using principled negotiation. *J. Guidance, Control, and Dynamics* 22(1):43–50.

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

Wellman, M., and Doyle, J. 1991. Preferential semantics for goals. In *Proceedings of AAAI-91*, 698–703.

Wellman, M. 1993. A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of Artificial Intelligence Research* 1:1–23.

Zeng, D., and Sycara, K. 1997. Benefits of learning in negotiation. In *Proceedings of AAAI-97*, 36–41.

Zeng, D., and Sycara, K. 1998. Bayesian learning in negotiation. In *Int J. Human Computer Systems*, volume 48, 125–141.

Zilberstein, S. 1998. Satisficing and bounded optimality. In *Proceedings of the 1998 AAAI Symposium*, 91–94. March 23–25, Stanford California. Technical Report SS-98-05.

Zlotkin, G., and Rosenschein, J. S. 1996a. Compromise in negotiation: Exploiting worth functions over states. *Artificial Intelligence* 84(1-2):151–176.

Zlotkin, G., and Rosenschein, J. S. 1996b. Mechanism design for automated negotiation and its application to task oriented domains. *Artificial Intelligence* 86(2):195–244.

Zlotkin, G., and Rosenschein, J. S. 1996c. Mechanisms for automated negotiation in state oriented domains. *Journal of Artificial Intelligence Research* 5:163–238.