Socially Rational Models for Autonomous Agents

Wynn C. Stirling
wynn_stirling@byu.edu

James K. Archibald
james_archibald@byu.edu

Matthew S. Nokleby

Socially Rational Models for Autonomous Agents

James K. Archibald,* Wynn C. Stirling, and Matthew S. Nokleby

Electrical and Computer Engineering Department, Brigham Young University, Provo, Utah, 84602, USA

Abstract: Autonomous multi-agent systems that are to coordinate must be designed according to models that accommodate such complex social behavior as compromise, negotiation, and altruism. In contrast to *individually rational* models, where each agent seeks to maximize its own welfare without regard for others, *socially rational* agents have interests beyond themselves. Such models require a new type of utility function—a *social utility*—to ensure three desirable properties: (a) *conditional preferences*—agents may adjust their preferences to account for the preferences of others; (b) *endogeny*—group preferences are determined internally by interactions between individual agents; (c) *framing invariance*—reformulations of the social model using exactly the same information should not alter the conclusions; and (d) *social coherence*—no individual's welfare is categorically subjugated to the welfare of the group. Social utilities in turn require a compatible solution concept—*optimal failure-avoidance*. *Satisficing game theory* embodies both social rationality and optimal failure-avoidance and provides a formal mathematical framework in which to balance group and individual interests in mixed-motive societies. The satisficing approach is applied to two scenarios: the Ultimatum game and a random graph search problem. The Ultimatum game is one for which game-theoretic analysis does not correspond well to empirical data regarding human behavior; it is thus an important test case for a new theory. The graph search scenario is an idealization of an important resource allocation problem in which the ability to compromise and negotiate can greatly facilitate the search for a solution.

Keywords: Decision theory, rationality, multi-agent coordination, satisficing game theory.

INTRODUCTION

Multi-agent system design is essentially a problem in social design. That is, successful systems must coordinate the behavior of individual members so as to accomplish tasks that are beyond the reach of any individual. In the most interesting systems, individual agents are assumed to be autonomous and to have private as well as public concerns. For such systems a successful design methodology must permit agents to make decisions that balance individual and group welfare.

To make rational decisions agents must have (a) a set of criteria to evaluate the effects that a choice has on its welfare and, (b) a solution concept that selects the most suitable choices. Agents are said to be *individually rational* when the criteria are defined by utility functions and the solution concept is to maximize expected utility. This is the perspective of von Neumann-Morgenstern game theory [1], in which agents form their preferences prior to any social interaction and then seek to maximize their individual welfare, subject to the constraint that other agents are doing the same. The resulting solution is a Nash equilibrium [2].

The assumption of individual rationality is well-suited to competitive societies, but in mixed-motive and potentially cooperative situations the choices it prescribes can be both individually and socially dysfunctional [3–5]. Arrow has observed that "rationality in application is not merely a property of the individual. Its useful and powerful implications derive from the conjunction of individual rationality and other basic concepts of neoclassical theory—equilibrium, competition, and completeness of markets . . . When these assumptions fail, the very concept of rationality becomes threatened, because perceptions of others and, in particular, their rationality become part of one's own rationality" [6, p. 203]. Thus it is not at all clear that individual rationality is an appropriate model for the synthesis of artificial multi-agent systems.

Luce and Raiffa observed a half-century ago 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" [7, p. 196]. Since then, numerous studies in experimental psychology and behavioral economics have cast doubt on the adequacy of the individual rationality hypothesis as a general model for social behavior [3–5, 8–11].

Schelling has observed [12] that societies take many forms, ranging from purely competitive societies, in which it is difficult to define a coherent notion of group preference and for which individual rationality is very appropriate, to purely coordinative societies, where all participants are in complete agreement regarding their preferences and actions. For these societies a group preference is easily defined by the *Pareto principle*: if an action is simultaneously preferred by all players, then it is preferred by the group.

Between these two extremes lies the vast and important family of mixed-motive societies, where opportunities for both competition and coordination are present. In such societies the assumption that individual preferences are completely determined without social interaction is not self-evident. Certainly, competition implies no concern for the welfare of others, but the very term *coordination* implies "the harmonious functioning together of different interrelated parts" [13]. If an agent truly has no interest in others, it seems that any coordination

*Address correspondence to this author at the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, USA; E-mail: jka@ee.byu.edu

1874-110X/08 2008 Bentham Open

it might achieve would be accidental. On the other hand, we would expect a society either disposed (human systems) or intended (artificial systems) to coordinate to benefit from such social behaviors such cooperation, compromise, negotiation, and even altruism.

Consider an alternate point of view, one admittedly not as neat as individual rationality but far more amenable to modeling social behavior. Let the proximate aim of the agents be to *avoid failure*, both for themselves individually and for the group as a whole. Failure-avoidance, while seemingly a more modest goal than utility maximization, has important advantages:

- Optimal behavior in multi-agent systems, either for individuals or the group, may be difficult or impossible to define. However, failure-avoidance concepts such as functionality, reliability, robustness, flexibility, and survivability can always be formed.
- Even when well defined, the optimization of utility is intrinsically an individual enterprise. If each individual were to optimize, the resulting behaviors would not necessarily be optimal, or even acceptable, to the group. The opposite is also true. In contrast, as will be established, failure-avoidance can apply simultaneously to both groups and individuals.
- Optimization produces a single solution, but failure-avoidance can often be achieved in more than one way. This provides alternatives for negotiation and compromise which can often reconcile group and individual interests to the satisfaction of both.

In the following we propose a model of *social rationality* for agents in mixed-motive societies. It differs from the individual rationality model in both (a) the criteria used to evaluate choices, and (b) the solution concept used to identify acceptable decisions. Social rationality requires a *social utility function* that permits agents to define their preference orderings in a way that accounts for the preferences, as well as the possible actions, of others. We further present a formalization of the intuitive notion of failure-avoidance. This new solution concept—satisficing game theory—accommodates sophisticated social relationships such as coordination, compromise, negotiation, and altruism. A remarkable feature of this theory is that coherent group and individual preferences emerge as a result of agent interactions.

We illustrate satisficing game theory with two applications. The first, the Ultimatum game, gives results consistent with experimental data that departs from the game-theoretic analysis. The second is a simulated graph search by a multi-agent team; n autonomous agents must coordinate to visit all nodes in a random graph with the constraint that they maintain communication with each other. We show that satisficing game theory is ideally suited to this problem since it accommodates multiple solutions for each player, providing opportunities for negotiation and compromise. We present simulation results that quantify performance and demonstrate the emergence of important coordinated behaviors.

SOCIAL RATIONALITY

A *society* is any collection of agents such that its members have the ability to influence each other's behavior or welfare. A *sub-society* is any subset of a society. In particular, an individual is a singleton sub-society. Let $G = \{X_1, \dots, X_n\}$ be a society with joint action space $\mathbf{U} = U_1 \times \cdots \times U_n$, and let $G^i = \{X_{i_1}, \dots, X_{i_k}\}$ be any sub-society, with corresponding action space $\mathbf{U}_i = U_{i_1} \times \cdots \times U_{i_k}$. We assume four axioms for autonomous, *socially rational*, mixed-motive societies.

Axiom 1 (Conditional Preferences) *Sub-societies may condition their preferences on the preferences of other sub-societies.*

Conditional preferences are hypothetical statements that enable a sub-society to define its preferences as a function of the possible preferences of another sub-society without actually knowing the preferences of the other, and without making a categorical commitment to the other.

To develop the concept of conditioning, we must first define a preference order. A sub-society G^i possesses a total preference ordering if there exists a reflexive, antisymmetric, transitive, and complete preference relation pair $(\succ_{G^i}, \sim_{G^i})$ over $\mathbf{u}_i \in \mathbf{U}_i$, where $\mathbf{u}_i \succ_{G^i} \mathbf{u}'_i$ means that G^i strictly prefers $\mathbf{u}_i$ to $\mathbf{u}'_i$, and $\mathbf{u} \sim_{G^i} \mathbf{u}'_i$ means preferences for the two actions are equivalent. The expression $\mathbf{u}_i \succeq_{G^i} \mathbf{u}'_i$ means that either $\mathbf{u}_i \succ_{G^i} \mathbf{u}'_i$ or $\mathbf{u}_i \sim_{G^i} \mathbf{u}'_i$, in which case $\mathbf{u}_i$ is preferred to $\mathbf{u}'_i$ (but not strictly).

Let G^i and G^j be disjoint sub-societies of G with corresponding action subspaces $\mathbf{U}_i$ and $\mathbf{U}_j$, respectively. A conditional preference ordering for G^i given G^j corresponds to a hypothetical proposition involving an *antecedent* and a *consequent*. The antecedent corresponds to a hypothesized preference ordering for G^j; namely, that $\mathbf{u}_j \succeq_{G^j} \mathbf{u}'_j$ for all $\mathbf{u}'_j \in \mathbf{U}_j$. Such a hypothesized preference ordering is denoted a *commitment* to $\mathbf{u}_j$ by G^j. Given this antecedent, the consequent corresponds to the resulting preferences of G^i, expressed as a *conditional preference ordering* $(\succ_{G^i|G^j}, \sim_{G^i|G^j})$. The expression $\mathbf{u}_i|\mathbf{u}_j \succ_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$ means that G^i strictly prefers $\mathbf{u}_i$ given that G^j is committed to $\mathbf{u}_j$, to $\mathbf{u}'_i$ given that G^j is committed to $\mathbf{u}'_j$, and $\mathbf{u}_i|\mathbf{u}_j \sim_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$ means that G^i is conditionally indifferent. The expression $\mathbf{u}_i|\mathbf{u}_j \succeq_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$ signifies either strict conditional preference or conditional indifference. A conditional preference forms a total ordering when $\mathbf{u}_j = \mathbf{u}'_j$.

To illustrate the notion of conditional preference, suppose G^i's action space is to choose the model of a car (i.e., $U_j = $ convertible, sedan), and G^j's action space is to choose the color (i.e., $U_j = $ red, green). The expression convertible|red $\succ_{G^i|G^j}$ sedan|green means that G^i prefers convertibles, given that G^j prefers red to green, over sedans, given that G^j prefers green to red. The expression convertible|red $\succ_{G^i|G^j}$ sedan|red is somewhat simpler to interpret; it means that G^i prefers convertibles to sedans if G^j prefers red to green.

Conditional preferences permit the specification of well-formed group goals, as well as operationally meaningful group preferences.

Axiom 2 (Endogeny) *If preference orderings exist for a so-*

ciety, they must be determined by interactions between sub-societies.

Axiom 2 prohibits the exogenous or external imposition of a social preference. In particular, it prohibits the use of a classical social welfare function. If some notion of collective preference exists for a society, it must emerge internally. This is a rather broad interpretation of autonomy, broader perhaps than many applications currently require, but one that we believe will become increasingly important as multi-agent systems are deployed for long periods of time in dynamic environments without human supervision.

Axiom 3 (Framing Invariance) *If a social model can be framed in more than one way using exactly the same information, all framings should yield the same decisions.*

A consequence of framing invariance is that there must be some concept of reciprocity; that is, if the preference ordering over one sub-society is conditioned on the preferences of another sub-society, then it is theoretically possible to re-frame the problem such that the preference ordering of the second sub-society can be expressed in terms of conditional preference orderings given the preferences of the first sub-society. If this assumption is violated, then either (a) some information has been lost or ignored in the re-framing, (b) the information has not been applied consistently, or (c) one of the sub-societies is intransigent and unwilling to take into consideration a different context when defining its preferences. This last concern can be problematic if multiple decision makers are involved. However, if a society is to be robust and enduring, it is reasonable to assume that the individuals do not form their preferences in a social vacuum, but rather in a social context that takes into account the fact that they must interact with others. This is particularly true for an artificial society that is designed to be cooperative.

Axiom 4 (Social Coherence) *No sub-society must be categorically required to subjugate its own welfare to the society in all situations in order to benefit the society.*

Axiom 4 is a weak notion of social equity which requires only that a society must allow for the possibility (but not the guarantee) that each sub-society can get its way, at least some of the time. It is tantamount to avoiding sure subjugation, whereby an individual or sub-society is required to sacrifice its welfare *in all circumstances* in order to benefit the larger group.

None of these axioms is compatible with the classical formulation. To accommodate these four axioms we require a new utility function, called a *social utility*.

Social Utilities

The function p_{G^i} is a *social utility* if $p_{G^i}(\mathbf{u}_i) \geq p_{G^i}(\mathbf{u}'_i) \iff \mathbf{u}_i \succeq_{G^i} \mathbf{u}'_i$ [14,15]. If $G^i = G$, p_G is a *joint social utility*, and if $G^i = \{X_i\}$, a singleton sub-society, then $p_{G_i} = p_{X_i}$ is the *marginal social utility* for X_i. The function $p_{G^i|G^j}$ is a *conditional social utility* if $p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) \geq p_{G^i|G^j}(\mathbf{u}'_i|\mathbf{u}'_j) \iff \mathbf{u}_i|\mathbf{u}_j \succeq_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$. This structure allows G^i to define its preferences as functions of each of G^j's possible most-preferred choices, and it is the mechanism by which G^i can extend its sphere of interest beyond itself.

Consider the problem where we are given arbitrary disjoint sub-societies G^i and G^j of society G, utility p_{G^j} (for G^j) and conditional utility $p_{G^i|G^j}$, and we wish to aggregate the given utilities to determine the social utility of sub-society $G^i \cup G^j$. With respect to Axiom 2, we say that G possesses *endogenously aggregated* utilities if, for any two disjoint sub-societies G^i and G^j, $p_{G^i \cup G^j}$, the social utility of sub-society $G^i \cup G^j$, is a function of the given utility p_{G^j} and the given conditional utility $p_{G^i|G^j}$. That is, there exists a function F, non-decreasing in both arguments, such that

$$p_{G^i \cup G^j}(\mathbf{u}_i, \mathbf{u}_j) = F[p_{G^j}(\mathbf{u}_j), p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)]. \quad (1)$$

(Note that, according to Axiom 3, reversing the roles of i and j would yield an identical aggregated utility for the new sub-society.) For this form of aggregation to correspond to a group preference it must be a total ordering of $\mathbf{U}_i \times \mathbf{U}_j$. Fortunately this property is immediate, since the binary relation $(\mathbf{u}_i, \mathbf{u}_j) \succeq_{G^i \cup G^j} (\mathbf{u}'_i, \mathbf{u}'_j)$ for the sub-society $G^i \cup G^j$ is induced by the function F, and hence is reflexive, antisymmetric, transitive, and complete. A reasonable and intuitively important property of a group ordering is that if $\mathbf{u}_i|\mathbf{u}_j \succeq_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$ and $\mathbf{u}_j \sim_{G^j} \mathbf{u}'_j$, or if $\mathbf{u}_i|\mathbf{u}_j \sim_{G^i|G^j} \mathbf{u}'_i|\mathbf{u}'_j$ and $\mathbf{u}_j \succeq_{G^j} \mathbf{u}'_j$, then $(\mathbf{u}_i, \mathbf{u}_j) \succeq_{G^i \cup G^j} (\mathbf{u}'_i, \mathbf{u}'_j)$. This condition obtains if and only if F is non-decreasing in both arguments.

Framing invariance requires that $p_{G^i \cup G^j}(\mathbf{u}_i, \mathbf{u}_j) = p_{G^j \cup G^i}(\mathbf{u}_j, \mathbf{u}_i)$ or, in other words,

$$F[p_{G^j}(\mathbf{u}_j), p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)] = F[p_{G^i}(\mathbf{u}_i), p_{G^j|G^i}(\mathbf{u}_j|\mathbf{u}_i)]. \quad (2)$$

It is important to appreciate that endogenous aggregation is strictly a mathematical operation, and may or may not correspond to harmonious, or even purposeful, group behavior. Such a preference ordering is emergent, in that it may not be anticipated or explicitly modeled, and it can take many forms. If the members of a society are in total opposition, such as in a zero-sum game, then the group "preference" should be to compete. At the other extreme, a society in which coordinated behavior yields high rewards for participating agents, the group preference should be consistent with that behavior. The absence of any meaningful concept of group preference would be an indication that the group is dysfunctional.

Compliance with Axiom 4 requires that social utilities possess the additional property of social coherence. A society G has *socially coherent* utilities if, for any two disjoint sub-societies G^i and G^j, $p_{G^i}(\mathbf{u}_i) \geq p_{G^i}(\mathbf{u}'_i)$ implies that there exists $\mathbf{u}^*_j \in \mathbf{U}_j$ such that $p_{G^i \cup G^j}(\mathbf{u}_i, \mathbf{u}^*_j) \geq p_{G^i \cup G^j}(\mathbf{u}'_i, \mathbf{u}^*_j)$ (where $\mathbf{u}^*_j$ may depend on $\mathbf{u}_i$). Social coherence means that if one sub-society prefers one action over another, then, for any other sub-society, there exists some action such that the resulting joint action is preferred by the combined sub-society. The violation of social coherence would result in a condition of *sure subjugation*, in the sense that under no condition would such a sub-society be able simultaneously to benefit itself and the society.

Axioms 2 and 4 guarantee an important common-sense result. Suppose that $p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) \geq p_{G^i|G^j}(\mathbf{u}_i'|\mathbf{u}_j)$ for all $\mathbf{u}_j \in \mathbf{U}_j$. Endogenous aggregation ensures that $p_{G^i \cup G^j}(\mathbf{u}_i, \mathbf{u}_j) \geq p_{G^i \cup G^j}(\mathbf{u}_i', \mathbf{u}_j)$ for all $\mathbf{u}_j \in \mathbf{U}_j$. If this condition holds then, by social coherence, $p_{G^i}(\mathbf{u}_i) \geq p_{G^i}(\mathbf{u}_i')$. Thus, these axioms ensure that, if G^i prefers $\mathbf{u}_i$ to $\mathbf{u}_i'$ no matter what action G^j prefers, it is better for both the larger society $G^i \cup G^j$ and G^i that G^i implement $\mathbf{u}_i$ rather than $\mathbf{u}_j'$.

The Sociality Theorem

There is in probability theory an important analogue to avoiding sure subjugation, namely, the wagering concept of avoiding sure loss—a betting situation such that, no matter what the outcome, the gambler loses. The Dutch Book theorem [16, 17] and its converse [18] state that it is not possible to construct a bet such that the individual will lose money no matter what happens if and only if the gambler acts in accordance with a probability measure that describes the individual's degrees of belief in the propositions under consideration. Such a belief system is said to be *coherent*.

Probability theory is chiefly concerned with the epistemological domain (i.e., the classification of propositions on the basis of knowledge and belief), and it is used to express the degrees of belief regarding the truth of propositions. However, the mathematical structure of probability can also be applied to the praxeological domain (i.e., the classification of actions on the basis of effectiveness and efficiency) to express the degrees of suitability of a set of actions. To this end, we define a *praxeological mass function* as a non-negative function that is normalized to sum to unity over its domain space; that is, p_{G^i} is a praxeological mass function if $p_{G^i}(\mathbf{u}_i) \geq 0\ \forall \mathbf{u}_i$ and $\sum_{\mathbf{u}_i \in \mathbf{U}_i} p_{G^i}(\mathbf{u}_i) = 1$, and $p_{G^i|G^j}$ is a conditional praxeological mass function if $p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) \geq 0\ \forall \mathbf{u}_i,\ \forall \mathbf{u}_j$ and $\sum_{\mathbf{u}_i \in \mathbf{U}_i} p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) = 1\ \forall \mathbf{u}_j$.

Theorem 1 (The sociality theorem) *Let $\{p_{G^i}\}$ be a family of social utilities for all sub-societies of G and let $\{p_{G^i|G^j}\}$ be a family of conditional social utilities associated with all pairs of disjoint sub-societies of G. These social utilities are endogenously aggregated if and only if they are praxeological mass functions, in which case*

$$F[p_{G^j}(\mathbf{u}_j), p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)] = p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) p_{G^j}(\mathbf{u}_j). \quad (3)$$

A proof of this theorem for F differentiable in both arguments is provided in the Appendix. Social utilities thus possess all of the syntactical properties of probability mass functions, albeit with praxeological, rather than epistemological, semantics.

- *Marginalization*: let G^i be an arbitrary sub-society of G and let $G^{\neg i}$ denote the complementary sub-society of G^i. Then the social utility of G^i is obtained by summing $p_G(\mathbf{u})$ over all actions in $G^{\neg i}$. It is convenient to express this sum via *"not-sum"* notation $\sum_{\neg\{\mathbf{u}_i\}}$, which sums over all elements not equal to $\mathbf{u}_i$, yielding

$$p_{G^i}(\mathbf{u}_i) = \sum_{\neg\{\mathbf{u}_i\}} p_G(\mathbf{u}). \quad (4)$$

To illustrate this notation, if $G = \{X_1, X_2, X_3, X_4\}$, the marginal social utility of $\{X_2, X_4\}$ is

$$\begin{aligned} p_{X_2X_4}(u_2, u_4) &= \sum_{\neg\{u_2,u_4\}} p_G(\mathbf{u}) \\ &= \sum_{\{u_1\}} \sum_{\{u_3\}} p_G(u_1, u_2, u_3, u_4). \quad (5) \end{aligned}$$

- *Independence*: if G^j's preferences have no influence on the preferences of G^i, then

$$p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j) = p_{G^i}(\mathbf{u}_i), \quad (6)$$

and G^i and G^j are *praxeologically independent* sub-societies.

The concept of conditioning, as defined earlier, is the property that underlies the probabilistic syntax (albeit with praxeological rather than epistemological semantics) of the utilities. Thus, there is a distinct analogy between the use of this syntax as an epistemological model of belief regarding the truth of propositions and a praxeological model of efficiency and effectiveness of taking action. It should be noted that it is also possible to define conditional probability (in the conventional epistemological sense) axiomatically, rather than as a derivative of the traditional Kolmogorov axioms [19, Chapter 2].

Social utilities differ from classical utilities in several ways. First, classical utilities are assumed to be invariant to scale and zero-level, but social utilities are not. Second, interpersonal comparisons of utility are not permitted with classical utilities, but such comparisons are both natural and desirable with social utilities. Third, classical utilities are functions of the actions of all agents, whereas social utilities are functions of *preferences for action*, and only within the sub-society over which they are defined. However, because sub-society utilities are derived from joint social utilities by marginalization, they encode all relationships from the global context that are important to the sub-society.

OPTIMALITY

For social utilities to be of practical value, two fundamental issues must be addressed: (a) how they can be used for modeling mixed-motive societies, and (b) how the performance of agents who use social utilities can be evaluated. In other words, some notion of optimization must be defined.

A classical way to achieve an optimal group decision requires the maximization of a social welfare function, typically defined as an aggregation of individual welfare functions [20]. However, this violates the autonomy of participating agents. If attempted, there would likely be a lack of consensus regarding what is best. Indeed, one may view a social welfare function as a mathematical "dictator" which imposes a single value system on potentially independent and possibly uncoordinated decision makers, and so creates a group, in the sense of a super-individual, by *fiat*.

An alternative approach is to broaden the applicability of the classical approach by including social considerations into the utility functions, such as inequity aversion, fairness, and reciprocal kindness. Researchers report more accurate modeling of human behavior in certain realistic settings [21–24]. However, as discussed earlier, accounting for social considerations via classical utility functions is tantamount to simulating social behavior with a mechanism designed for asocial behavior. Simply put, it is the wrong tool. The challenge is to develop a methodology that (a) possesses some form of logical and internal consistency between individual and group-level interests, and (b) admits a well-defined concept of optimality. This goal requires a new approach to the formulation of decision problems.

Decision Formulation

To motivate a re-formulation of the multi-agent decision problem, it will be helpful to change the context and consider first an epistemological decision problem. Suppose an agent is confronted with a number of propositions, all of which are possibly true, but only one of which is in fact true. The agent could insist on a single answer—"the truth and nothing but the truth"—but if the evidence is not sufficient to identify a single proposition as true, such behavior would be temerarious. A more circumspect investigator would, in the interest of avoiding error, eliminate from consideration only those possibilities that are unlikely to be true or that, even if true, are of little consequence (as Popper noted, "Yet we must also stress that *truth is not the only aim of science.* We want more than truth: what we look for is *interesting truth*" [25, p. 229, emphasis in original]), and continue to investigate until sufficient evidence is obtained to identify the true proposition.

The pragmatic philosopher William James referred to this formulation of a decision problem as *error-avoiding*: "There are two ways of looking at our duty in the matter of opinion,—ways entirely different, and yet ways about whose difference the theory of knowledge seems hitherto to have shown very little concern. *We must know the truth, and we must avoid error*,—these are our first and great commandments as would-be knowers; but they are not two ways of stating an identical commandment, they are two separable laws . . . Believe truth! Shun error!—these, we see, are two materially different laws; and by choosing between them we may end by coloring differently our whole intellectual life. We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may, on the other hand, treat the avoidance of error as more imperative, and let truth take its chance" [26, pp. 17,18, emphasis in original].

Simply put, an error-avoider is a cautious truth-seeker; one who does not insist on identifying a unique solution in all circumstances. There are various degrees of error-avoidance, depending on the amount and quality of information available. Ideally, one is able to eliminate all but one proposition, thereby exposing the true one. But in general, there may be several non-eliminated propositions, each with a claim to being true. An error-avoider refines its choices to the extent warranted by the information, but is not obligated to settle on a unique solution.

The philosopher Isaac Levi [27] has defined a rigorous approach to the error-avoidance formulation of the decision problem. As discussed above, the proximate aim of the agent is to avoid error, but this aim is tempered by a demand for information. Information, in this context, is obtained by eliminating propositions from serious consideration. Levi constructs two utilities to account for these competing criteria: an error-avoidance utility and an informational value utility.

Let $U = \{u_1, u_2, \ldots, u_N\}$ be a finite set of propositions, one of which is true, and let $\mathcal{B}$ denote a Boolean algebra of subsets of U. Then for any set $A \in \mathcal{B}$, A is true if and only if A contains the true element. The error-avoidance utility is then defined as

$$T(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where A is the set of propositions that will not be eliminated. Note that a conventional utility would have the same structure, but requires A to be a singleton set. It is straightforward to see that $T(U) = 1$ and that T is additive over disjoint sets (i.e., $T(A_1 \cup A_2) = T(A_1) + T(A_2)$ if $A_1 \cap A_2 = \varnothing$). Thus, T is a normalized measure over $\mathcal{B}$.

An agent can guarantee that $T(A) = 1$ only by setting $A = U$, but this results in no information gain. To temper the desire to avoid error with the demand for information, the agent must compute the informational value of each set A. For reasons that will become clear, this value will be determined by considering the informational value of rejecting A.

Whether or not it is true, A is assumed to have some intrinsic informational value. Such abductive inferences are hypothetical statements of informational importance while temporarily ignoring considerations of veracity. They may take many forms, including economic, political, moral, cognitive, aesthetic, or personal. The following assumptions constitute a reasonable model for informational value: (a) informational value is non-negative; (b) the informational value of rejecting the union of two disjoint sets of propositions is equal to the sum of the informational values of rejecting the individual sets; and (c) the informational value of rejecting U is unity. This structure implies that informational valuations comply with the rules of classical measure theory. Let P_R denote a measure that maps elements of $\mathcal{B}$ to the unit interval. $P_R(A)$ is a utility that represents the informational value that accrues to the agent if A is rejected. Defining informational value in this way corresponds to the error-avoiding view of rejecting propositions, rather than the conventional view of accepting one and only one proposition.

The utility P_R is not a classical utility because it is a mapping of sets of propositions—elements of the Boolean algebra $\mathcal{B}$ over the proposition space U—rather than a mapping of individual propositions in U. However, if the Boolean algebra contains all singleton sets, the resulting classical utility function $p_R(u) = P_R(\{u\})$. $p_R(u)$ is a mass function, that is, $p_R(u) \geq 0$ for all $u \in U$ and $\sum_{u \in U} p_R(u) = 1$, so $P_R(A) = \sum_{u \in A} p_R(u)$. The informational value of nonrejection of A is then defined as

$$I(A) = 1 - P_R(A). \quad (8)$$

The utility of both avoiding error and acquiring information is given by a convex combination of $T(A)$ and $I(A)$, yielding Levi's *epistemic utility function*

$$\phi(A) = \alpha T(A) + (1 - \alpha) I(A), \quad (9)$$

where $0 \leq \alpha \leq 1$. The parameter α represents the relative importance that is attached to avoiding error versus acquiring information. Setting $\alpha = 0$ puts a premium on avoiding error, and setting $\alpha = 1$ places a premium on the desire for information regardless of its veracity. As a practical issue, α should be restricted to the interval $[1/2, 1]$ to ensure that no erroneous answer is preferred to any correct answer.

By construction, ϕ is a von Neumann-Morgenstern utility, even though it is a function of *sets* of propositions, rather than of single propositions. Consequently, it is invariant to scale and zero level, and an equivalent utility is

$$\begin{aligned} \phi^\alpha(A) &= \frac{1}{\alpha}\phi(A) - \frac{1-\alpha}{\alpha} \\ &= T(A) - qP_R(A) \\ &= \begin{cases} 1 - qP_R(A) & \text{if } A \text{ is true} \\ -qP_R(A) & \text{if } A \text{ is false} \end{cases} \end{aligned} \quad (10)$$

where $q = (1 - \alpha)/\alpha$. Thus, $0 \leq q \leq 1$.

To complete the framework, consider the evidence regarding the propositions. Let P_S denote a probability measure over the Boolean algebra $\mathcal{B}$ such that $P_S(A)$ is the probability (e.g., belief or other evidential support) that A contains the true proposition, and let p_S denote the associated mass function. The expected epistemic utility is

$$\begin{aligned} E\phi^\alpha(A) &= [1 - qP_R(A)]P_S(A) - qP_R(A)[1 - P_S(A)] \\ &= P_S(A) - qP_R(A). \end{aligned} \quad (11)$$

It is now evident that expected epistemic utility is maximized by the largest element of $\mathcal{B}$ for which the probability of truth is as least as great as q times the informational value of rejection. Thus, expected epistemic utility is maximized by rejecting all and only those elements of U for which $p_S(u) < qp_R(u)$. Here, q may be interpreted as the *index of caution*. As q increases, the agent rejects more propositions and becomes more willing to risk error in the interest of obtaining more information.

To illustrate the application of Levi's theory, suppose X were to consider three hypotheses to explain some observed symptoms: $u_1 =$ "indigestion", $u_2 =$ "food poisoning", and $u_3 =$ "ulcer". Suppose X's belief regarding the truth of these three hypotheses are $p_S(u_1) = 0.8$, $p_S(u_2) = p_S(u_3) = 0.1$. Also, suppose X_2 considers u_2 to be twice as informationally valuable as u_1, and u_3 to be ten times as informationally valuable as u_1. A reasonable way to compute the informational value of rejection is to define $p_R(u_i)$ as the normalized reciprocal of the informational value of retention, yielding $p_R(u_1) = 0.625$, $p_R(u_2) = 0.3125$, and $p_R(u_3) = 0.0625$. Setting $q = 1$, only u_1 and u_3 are retained. Even though u_1 is by far the most likely cause of the symptoms, u_3 is not rejected, because the informational value of that hypothesis requires a great deal of evidence that it is not the correct hypothesis before it can be eliminated from consideration. Hypothesis u_2, on the other hand, is not sufficiently important, even if true, to be retained as a serious possibility.

In the interest of avoiding error while acquiring information, X is more conservative than an agent demanding a unique resolution in the interest of seeking only the truth. With the error-avoiding approach, the probability of error is 0.1, and with the truth-seeking approach, the probability of error is either 0.2 or 0.9. Of course, the price paid for avoiding error is that X does not achieve a unique resolution of the issue. Nevertheless, by rejecting the hypotheses for which the informational value of rejection exceeds the probability of being true, the retained hypotheses can be used to guide the search for additional evidence to narrow the decision to a single hypothesis.

Satisficing Theory

An epistemologist makes decisions to gain information, while a praxeologist makes decisions to achieve goals (success). The analogue to a set of epistemological propositions is a set of possible actions. The degree to which the objective is not achieved is the degree of failure, and the analogue to error-avoidance is failure-avoidance. While Levi's formalism for optimal error-avoidance is motivated by a lack of sufficient information to justify a unique answer, it is also well suited to the analogous praxeological situation of having several possible actions that can be justified as being adequate.

Transitioning from the epistemological to the praxeological requires an appropriate analogue to the informational value of rejection. Informational value is a resource that is conserved if the proposition is rejected and consumed if a proposition is retained. For example, if a proposition with little monetary value (and consequently a high informational value of rejection) were rejected, the agent would conserve its monetary resource. Similarly, if an expensive act (e.g., one that cost a great deal of money) were rejected, the agent would not expend the corresponding money—it would be conserved. By rejecting an act, the agent effectively conserves the associated resource (such as monetary costs, damage to needful equipment, exposure to hazards, loss of social status, consumption of energy, risk to personnel, etc.). Thus, whereas the intent of the epistemologist is to acquire information while avoiding error, the intent of the praxeologist is to conserve resources while avoiding failure.

Many theorists (e.g., [28–31]) have argued that it is unwise to aggregate conflicting interests into a single preference ordering. Some have asserted that in a social setting individuals have two selves. These selves are similar to the "facets" or "aspects" of a self as defined by [32], who maintain that an agent, although an indivisible unit, nevertheless is capable of considering his or her choice from different points of view, and that separate utilities may be defined to correspond to each facet of an individual. A natural way to classify attributes is according to their effectiveness and efficiency. Each individual X_i

may be viewed as being composed of two selves: the *selecting self*, denoted S_i, who evaluates actions in terms of effectiveness without concern for efficiency, and the *rejecting self*, denoted R_i, who evaluates actions in terms of efficiency, without concern for effectiveness. When viewed simultaneously from both perspectives, the agent is denoted as the concatenation of these two selves, i.e., $X_i = S_i R_i$.

Praxeological interpretations may now be given to the utility functions p_{R_i} and p_{S_i}, which are mass functions and, hence, marginal social utilities. For each $u_i \in U_i$, $p_{S_i}(u_i)$ is a measure of the effectiveness of u_i, and we call it the *selectability* mass function. The relationship $p_{S_i}(u_i) > p_{S_i}(u_i')$ means that u_i is more effective than u_i' in terms of avoiding failure. Similarly, $p_{R_i}(u_i)$ is a measure of the inefficiency of u_i, and we refer to it as the *rejectability* mass function. The relationship $p_{R_i}(u_i) > p_{R_i}(u_i')$ means that u_i is less efficient than u_i' in terms of conserving resources. The set

$$\Sigma_q^i = \{u_i \in U_i \colon p_{S_i}(u_i) \geq q_i p_{R_i}(u_i)\} \qquad (12)$$

constitutes the set of actions for which effectiveness, as measured by the selectability function, is at least as great as q_i times the inefficiency, as measured by the rejectability function. The set Σ_q^i is called the *satisficing set* for X_i. The praxeological interpretation of q_i is a measure of caution, as before. As q_i increases, so does the number of actions that are rejected, indicating that the agent is increasingly willing to risk failure in the interest of conserving resources. As will become more apparent in the multi-agent case, an appropriate interpretation of q_i is as an *index of negotiation*, since lowering q_i enlarges the satisficing set, thereby increasing the opportunities for reaching a compromise. Of course, lowering q_i is tantamount to X_i lowering its standards of what is deemed to be acceptable.

The most fundamental way the error-avoidance formulation differs from the classical optimization formulation is that, whereas the classical formulation involves comparisons of a single attribute (utility) between multiple actions to identify the best one, the error-avoidance formulation involves comparisons between multiple attributes (effectiveness and efficiency) for each action to decide whether or not to reject.

To motivate this alternative concept of decision making, consider three separate notions: superlative, comparative, and positive. Much of human decision making employs one of these notions. Individual rationality is an example of the superlative, where decision makers make global comparisons of their options and choose the best one. In contrast, heuristic decision making is an example of the positive, where decision makers rely upon the belief that a rule that has worked in the past will also work well in the future (e.g., rule-based expert systems). Although economic, psychological, philosophical, engineering, and computer science literatures are replete with discussions of these two notions, they are relatively silent regarding decision making that is comparative, even though people often seem to work toward a decision by first eliminating bad choices before settling on acceptable ones. They compare the pros versus the cons, upsides versus downsides, benefits versus costs, etc. This way of making decisions is more primitive than a total rank-ordering of options (superlative), but is more sophisticated than simply following heuristic rules (positive). The idea of viewing an action from two perspectives—one focused on the positive consequences of adopting it and the second focused on the negative consequences—is a powerful concept, and one for which a mathematically rigorous formalization is long overdue.

Just as the colloquial notion of achieving "the best and only the best" is useful once it has been mathematically formalized as maximizing utility, the colloquial notion of "getting one's money's worth" is useful once it is formalized. The difference between these two concepts is significant: the former is intrinsically an individual enterprise, while, as shall be shown, the latter can be extended naturally to groups and individuals simultaneously.

The term *satisficing* was originally introduced by Simon [33] as a type of bounded rationality. He proposed to halt searching for the optimal solution when the expected improvement is insufficient to justify the costs of continuing to search. The halting criterion is a heuristically defined aspiration level. Thus satisficing, in Simon's sense, is firmly rooted in individual rationality and is a heuristic approximation to utility maximization.

The failure-avoidance formulation motivates a new and mathematically precise definition of satisficing. Reducing the set of non-rejected alternatives to the minimum eliminates as many options as possible; each of the remaining alternatives is "good enough" in the sense that its effectiveness outweighs its inefficiency. Satisficing decisions are optimal in that they eliminate the maximum number of failure-prone actions. *Thus, satisficing agents are optimal failure-avoiders*. Furthermore, if they succeed in eliminating all but one action, they will become optimizers in the classical sense (as Stirling [34] has shown, an optimal solution is also a satisficing solution). Thus, rather than a heuristic approximation to classical optimization, *satisficing is a generalization of classical optimization*.

To illustrate the satisficing way of making decisions, suppose that X is in the market for a new automobile and must choose from among five alternatives, denoted A through E. Three criteria are considered: performance, reliability, and affordability. Table 1 displays the utility of each of the vehicles for each of these attributes.

Table 1: Utility of vehicle attributes.

Vehicle u	Performance $p(u)$	Reliability $r(u)$	Affordability $a(u)$
A	3	1	5
B	5	3	1
C	2	4	4
D	1	5	3
E	4	2	2

The optimizer's formulation of this multi-attribute decision problem is to demand the best deal by defining a utility function to be maximized. Assuming that X weights the three attributes equally in importance, a global utility may be formed as the sum of the three attribute-level utilities, yielding

$$\phi(u) = p(u) + r(u) + a(u) \tag{13}$$

for each $u \in \{A, B, C, D, E\}$, as displayed in the second column of Table 2. Clearly, the unique optimal option is C. But demanding the best deal is not the only way to frame the problem. Another way is for X to demand to get its money's worth. This formulation does not involve making inter-vehicle comparisons; rather, it involves intra-vehicle comparisons of attributes for each alternative. To make these comparisons, X requires operational definitions of selectability and rejectability. Accordingly, X identifies performance and reliability as selecting attributes, and cost as a rejecting attribute. The values associated with these attributes are combined, normalized, and tabulated in the last two columns of Table 2 (the ordering on the affordability attribute has been reversed to convert it to cost). Setting $q = 1$, selectability exceeds rejectability for options A and C, selectability equals rejectability for D, and rejectability exceeds selectability for B and E.

Table 2: Global performance and selectability/rejectability functions.

Vehicle u	Global Utility $\phi(u)$	Selectability $p_S(u)$	Rejectability $p_R(u)$
A	9	0.133	0.067
B	9	0.267	0.333
C	10	0.200	0.133
D	9	0.200	0.200
E	8	0.200	0.267

Figure 1 provides a cross plot of selectability versus rejectability as u is varied over its domain, with p_R the abscissa and p_S the ordinate. The diagonal line corresponding to $q = 1$ constitutes a threshold dividing the satisficing and non-satisficing alternatives. Although both C and D are satisficing, D costs more than C without offering increased benefit. As will be discussed, C dominates D. Thus, options A and C are the non-dominated satisficing solutions.

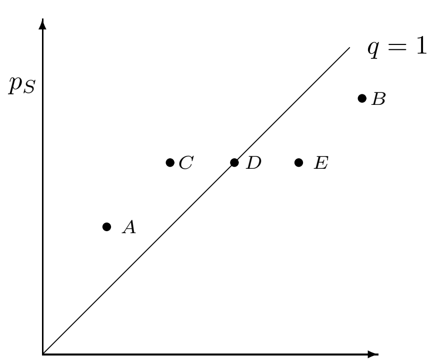

Figure 1: Cross-plot of selectability versus rejectability.

This example illustrates the fact that, when the same criteria are used to define both the optimal and satisficing solutions, the optimal solution is also satisficing. A key difference between these two methods is that the satisficing approach provides insight into the attributes of all alternatives, while the optimal approach focuses exclusively on identifying the best solution without distinguishing between non-optimal alternatives. For example, although A and B have the same global utility, they are not equal in terms of satisficing: one gives X its money's worth, while the other does not. Of course, at the moment of truth, X must decide between A and C. Since satisficing decision theory is not designed to provide a unique solution, ancillary criteria must be evoked to make a final choice. Ways to do this will be discussed subsequently.

MULTI-AGENT SATISFICING

In the single-agent case, satisficing theory sheds new light on the decision problem, but is otherwise of limited interest. Its real power becomes evident when extended to multi-agent systems. Classical theory interprets optimality in terms of unique solutions, but in the case of multi-agent decisions this insistence overreaches: the number of different perspectives makes it logically impossible to choose a single best solution from considerations internal to the group. In contrast, the satisficing approach preserves opportunities for negotiation and compromise by preserving a set of *adequate* choices rather than selecting a single "best" solution. As shall be established, it is possible for a group and all its members to obtain optimal error-avoiding solutions.

To ensure well-formedness, the concepts of effectiveness and efficiency must not be re-statements of the same attribute. Consequently, for a single-agent decision problem, it is reasonable to assume that the selectability of an attribute should not depend on its rejectability, and vice versa. Thus, S and R are praxeologically independent. In a multi-agent setting, however, the interaction between one agent's efficiency/effectiveness and another agent's efficiency/effectiveness can generate praxeological dependencies between the various selves of a society. Thus, in group settings, the selectability and rejectability measures associated with effectiveness and efficiency cannot be specified independently of each other. A critical aspect of modeling the behavior of a multi-agent society, therefore, is the representation of the interdependence of both effectiveness and efficiency of all possible joint actions that could be undertaken.

The Interdependence Function

Let $G = \{X_1, \ldots, X_n\}$ be a society of n agents with joint action space $\mathbf{U} = U_1 \times \cdots \times U_n$. Let $G_S = \{S_1, \ldots, S_n\}$ denote the collection of selecting selves, and let $G_R = \{R_1, \ldots, R_n\}$ denote the collection of rejecting selves. Then an equivalent representation of the society in terms of the selecting and rejecting selves is $G_{SR} = \{S_1 \cdots S_n R_1 \cdots R_n\}$. Let $G' = \{X_{i_1}, \ldots, X_{i_k}\}$ be a sub-society of G. Expressed in terms of the corresponding selves, $G'_{SR} = \{S_{i_1} \cdots S_{i_k} R_{i_1} \cdots R_{i_k}\}$. The *interdependence function* of G is a mass function of $2n$ variables of the form $p_{G_{SR}}(\mathbf{u}; \mathbf{v})$ where $\mathbf{u}, \mathbf{v} \in \mathbf{U}$. It is sometimes useful to employ the equivalent notation $p_{S_1 \cdots S_n R_1 \cdots R_n}(\mathbf{u}; \mathbf{v})$. Also, the interdependence function of the sub-society G' is of the form

$p_{G'_{SR}}(\mathbf{u}';\mathbf{v}')$, where $\mathbf{u}',\mathbf{v}' \in \mathbf{U}' = U_{i_1} \times \cdots \times U_{i_k}$.

The interdependence function is a social utility, as defined earlier, that accounts for all possible effectiveness and efficiency relationships that exist between the selves involved in a multi-agent decision problem. It does this in the same way that a multivariate probability mass function accounts for all statistical dependencies between multiple random phenomena. Thus, to formulate a multi-agent satisficing problem, the key task is to define the interdependence function.

Efficient Representations

The interdependence function is a mass function and may be most simply represented by factoring it into the product of conditional and marginal mass functions. Graph theory is a powerful way to express this factorization. In particular, the flow of influence between selves may be expressed by a *praxeic network*, that is, a directed acyclic graph (DAG) analogous to a Bayesian network (e.g., see [35, 36]). A praxeic network for n agents comprises $2n$ vertices (one for each self), with edges representing influence relationships (either effectiveness or efficiency) as modeled by conditional social utilities.

Consider the praxeic network of the three-agent system displayed in Figure 2. The *parents* of a vertex V, denoted $\mathrm{pa}\,(V)$, is the set of vertices that influence it. By inspection it is seen that $S_1 = \mathrm{pa}\,(S_2) = \mathrm{pa}\,(S_3) = \mathrm{pa}\,(R_2)$ and $S_2 = \mathrm{pa}\,(R_1) = \mathrm{pa}\,(R_3)$. If a vertex V has no parents, the $\mathrm{pa}\,(V) = \varnothing$. The *children* of a vertex V, denoted $\mathrm{ch}\,(V)$, is the set of vertices that are directly influenced by V. Thus, $\mathrm{ch}\,(S_1) = \{S_2, S_3, R_2\}$ and $\mathrm{ch}\,(S_2) = \{R_1, R_3\}$. The *descendants* $\mathrm{de}\,(V)$ of V is the set of all vertices that are influenced, directly or indirectly (via children, children's children, etc.) by the given vertex. Thus, $\mathrm{de}\,(S_1) = \{S_2, S_3, R_1, R_2, R_3\}$, and $\mathrm{de}\,(S_2) = \mathrm{ch}\,(S_2)$.

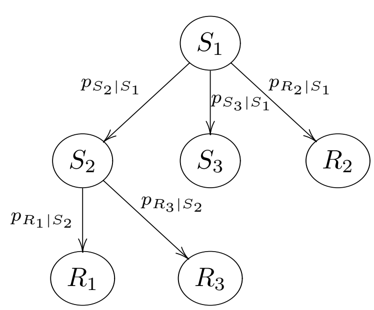

Figure 2: A praxeic network for a three-agent society.

The key property of DAGs is the *Markov property*: non-descendant non-parents of a vertex are conditionally independent of the vertex, given the state of its parent vertices. This property may be used to prove the equivalence of a DAG whose edges are conditional mass functions with a joint mass function for all of the vertices in the graph (for a proof, see [35, 37]). Thus, if local influence relationships can be expressed with a directed acyclic graph, then the influence relationships can be represented by conditional mass functions where the dependencies flow in only one way: from parents to children. The interdependence function thus has the form

$$p_{S_1 \ldots S_n R_1 \ldots R_n}(u_1, \ldots, u_n; v_1, \ldots, v_n) = \prod_{i=1}^{n} p_{S_i|\,\mathrm{pa}\,(S_i)}\big(u_i|\,\mathrm{pa}\,(u_i)\big) p_{R_i|\,\mathrm{pa}\,(R_i)}\big(v_i|\,\mathrm{pa}\,(v_i)\big), \quad (14)$$

where, if $\mathrm{pa}\,(S_i) = \varnothing$, then $p_{S_i|\,\mathrm{pa}\,(S_i)}(u_i|\,\mathrm{pa}\,(u_i)) = p_{S_i}(u_i)$, the unconditional marginal social utility. The interdependence function corresponding to the DAG illustrated in Figure 2 is

$$\begin{aligned} &p_{S_1 S_2 S_3 R_1 R_2 R_3}(u_1, u_2, u_3; v_1, v_2, v_3) = \\ &\quad p_{S_1}(u_1) \cdot p_{S_2|S_1}(u_2|u_1) \cdot p_{S_3|S_1}(u_3|u_1) \cdot \\ &\quad p_{R_1|S_2}(v_1|u_2) \cdot p_{R_2|S_1}(v_2|u_1) \cdot p_{R_3|S_2}(v_3|u_2). \end{aligned}$$

For societies that can be characterized by marginal and conditional social utilities defined over small clusters of individuals, a graphical representation provides a convenient way to construct a global society model from local relationships. An important advantage of viewing a multi-agent satisficing decision problem in terms of graph theory is that it leads to computationally efficient algorithms such as Pearl's Belief Propagation Algorithm [35] for computing the selectability and rejectability functions for the society, any sub-society, or any individual. Although Cooper [38] proved that the computational complexity of a general Bayesian network is NP-hard, many interesting networks will involve only sparsely linked vertices, in which case the published algorithms offer tractable performance.

Satisficing Games

A satisficing game, as defined by Stirling [34], is a triple $(G, \mathbf{U}, p_{G_{SR}})$. Given the game scenario, the first step is to identify operational definitions of selectability or rejectability for each of the $2n$ selves in an n player game. The next step is to define the relationships that exist between the various selves and to construct the praxeic network that consequently defines the interdependence function. Once the interdependence function is defined, the selectability and rejectability functions of the society and all sub-societies may be obtained by marginalization.

The *joint selectability and rejectability mass functions* of a society G are given as

$$p_{G_S}(\mathbf{u}) = \sum_{\{\mathbf{v}\}} p_{G_{SR}}(\mathbf{u};\mathbf{v}) \quad (15)$$

$$p_{G_R}(\mathbf{v}) = \sum_{\{\mathbf{u}\}} p_{G_{SR}}(\mathbf{u};\mathbf{v}) \quad (16)$$

and, for any sub-society G' of G, the corresponding marginal selectability and rejectability social utilities are, for $\mathbf{u}' \in \mathbf{U}'$ (in the not-sum notation introduced in Theorem 1),

$$p_{G'_S}(\mathbf{u}') = \sum_{\neg\{\mathbf{u}'\}} p_{G_S}(\mathbf{u}) \quad (17)$$

$$p_{G'_R}(\mathbf{u}') = \sum_{\neg\{\mathbf{u}'\}} p_{G_R}(\mathbf{u}). \quad (18)$$

In particular, the *individual marginal selectability and rejectability mass functions* are, for $i = 1, \ldots, n$,

$$p_{S_i}(u_i) = \sum_{\neg\{u_i\}} p_{G_S}(\mathbf{u}) \quad (19)$$

$$p_{R_i}(u_i) = \sum_{\neg\{u_i\}} p_{G_R}(\mathbf{u}). \quad (20)$$

Once the marginal selectability and rejectability functions have been computed, the individually satisficing sets are easily obtained for each agent as

$$\Sigma^i_{q_i} = \{u_i \in U_i \colon p_{S_i(u_i)} \geq q_i p_{R_i(u_i)}\} \quad (21)$$

for $i = 1, \ldots, n$. Notice that each agent may have its own q value, which controls its openness to negotiation. The Cartesian product of the individually satisficing sets is called the *satisficing rectangle*:

$$\mathcal{R}_G = \Sigma^1_{q_1} \times \cdots \times \Sigma^n_{q_n}. \quad (22)$$

The satisficing rectangle is the set of all option vectors that are simultaneously satisficing for all of the individuals. It does not, however, represent a group preference. The set of option vectors that are jointly satisficing for the group G is computed from the joint selectability and rejectability functions (15) and (16), and for any sub-society G^1, the corresponding sub-society selectability and rejectability functions are given by (17) and (18). For a society G the *jointly satisficing set* is

$$\mathbf{\Sigma}^G_{q_G} = \{\mathbf{u} \in \mathbf{U} \colon p_{G_S}(\mathbf{u}) \geq q_G p_{G_R}(\mathbf{u})\}, \quad (23)$$

where q_G is the q-value for the group. Furthermore, for any sub-society $G^1 = G^1_S G^1_R$, the sub-society satisficing set with corresponding action subspace $\mathbf{U}_1$ is

$$\mathbf{\Sigma}^{G^1}_{q_{G^1}} = \{\mathbf{u}_1 \in \mathbf{U}_1 \colon p_{G^1_S}(\mathbf{u}_1) \geq q_{G_1} p_{G^1_R}(\mathbf{u}_1)\}. \quad (24)$$

Endogeny and Social Coherence

The satisficing solution concept induces an emergent preference ordering for the society. Define the group preference relationship $\{\overset{s}{\succ}_G, \overset{s}{\sim}_G\}$, where

$$\begin{array}{l} \mathbf{u} \overset{s}{\succ}_G \mathbf{u}' \text{ if } \mathbf{u} \in \mathbf{\Sigma}_G \text{ and } \mathbf{u}' \notin \mathbf{\Sigma}_G \\ \mathbf{u} \overset{s}{\sim}_G \mathbf{u}' \text{ if } \mathbf{u}, \mathbf{u}' \in \mathbf{\Sigma}_G \text{ or } \mathbf{u}, \mathbf{u}' \notin \mathbf{\Sigma}_G \end{array}. \quad (25)$$

It is important to appreciate that this group-level preference ordering is determined by the endogenous relationships that exist among the individuals, and need not correspond to an externally conceived notion of group functionality. It is an emergent manifestation of the social welfare of the group as a function of the way the unconditional and conditional preferences of its members combine. Social welfare, in this sense, thus accounts for all tendencies for cooperation and competition that exist among the individuals, but is not an aggregation of individual welfares. Since it is emergent, its exact nature will generally not be predictable in advance, even for a cooperatively disposed group.

The individually satisficing sets also induce agent-level preference relationships $\{\overset{s}{\succ}_i, \overset{s}{\sim}_i\}$, where

$$\begin{array}{l} u_i \overset{s}{\succ}_i u'_i \text{ if } u_i \in \Sigma^i_{q_i} \text{ and } u'_i \notin \Sigma^i_{q_i} \\ u_i \overset{s}{\sim}_i u'_i \text{ if } u_i, u'_i \in \Sigma^i_{q_i} \text{ or } u_i, u'_i \notin \Sigma^i_{q_i} \end{array}. \quad (26)$$

These individual and group preference orderings provide a means for reconciling group and individual preferences. The *compromise set* $\mathcal{C}_G$ consists of all joint actions that are simultaneously satisficing for the group and for each of its constituent members, and is defined by the intersection of the jointly satisficing set and the satisficing rectangle:

$$\mathcal{C}_G = \mathcal{R}_G \cap \mathbf{\Sigma}^G_{q_G}. \quad (27)$$

This set may be empty, but that is not a weakness of the theory. Rather, it is the recognition that societies can be populated by individuals who are so diametrically opposed to each other that they reach an impasse and cannot agree to do anything jointly satisficing. However, the following weak relationship always exists between $\mathcal{R}_G$ and $\mathbf{\Sigma}^G_{q_G}$.

Theorem 2 (The negotiation theorem) *Let G be a society and let G^1 and G^2 be arbitrary disjoint sub-societies with action subspaces $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively. If $\mathbf{u}_1 \in \Sigma_{G^1}$ and $q_G \leq q_{G^1}$, then there exists $\mathbf{u}^*_2 \in \mathbf{U}_2$ such that $(\mathbf{u}_1, \mathbf{u}^*_2) \in \Sigma_{G^1 \cup G^2}$.*

Proof This result is established by the contra-positive, namely, that if $(\mathbf{u}_1, \mathbf{u}_2) \notin \Sigma_{G^1 \cup G^2}$ for all $\mathbf{u}_2 \in \mathbf{U}_2$, then $\mathbf{u}_1 \notin \Sigma_{G^1}$. Suppose $p_{G_S}(\mathbf{u}_1, \mathbf{u}_2) < q_G p_{G_R}(\mathbf{u}_1, \mathbf{u}_2)$ for all $\mathbf{u}_2 \in \mathbf{U}_2$. Then $p_{G^1_S}(\mathbf{u}_1) = \sum_{\{\mathbf{u}_2\}} p_{G_S}(\mathbf{u}_1, \mathbf{u}_2) < q_G \sum_{\{\mathbf{u}_2\}} p_{G_R}(\mathbf{u}_1, \mathbf{u}_2) = q_G p_{G^1_R}(\mathbf{u}_1) \leq q_{G_1} p_{G^1_R}(\mathbf{u}_1)$, hence $\mathbf{u}_1 \notin \Sigma_{G^1}$. □

Although this theorem is simple, it is important: it establishes that G^1 need not be subjugated in order to accommodate the interests of the society. In particular, for $G^1 = \{X_i\}$, every individual has a seat at the table in the sense that, if an action is individually satisficing for it, then that action is an element of at least one jointly satisficing solution. This condition is perhaps the weakest possible for meaningful negotiations to occur.

The ability of an agent to adjust its index of negotiation q_i provides a mechanism for autonomously exploring the effects of constraints on the decision problem. This is an important new capability. If a given set of constraints leads to a solution judged to be inadequate, conventional methodologies require the constraints to be revised by trial-and-error—their effects cannot be judged without generating a new solution. By representing explicitly the effects of social constraints on group decisions, satisficing game theory makes those constraints available for dynamic modification by the agents themselves, thus increasing the environmental variability with which the group can cope. For example, agents may resolve an impasse by relaxing their standards of performance. This may be done by each player incrementally reducing its q_i and re-computing the compromise set until it becomes non-empty. An important feature of this procedure is that each agent can control its own standards of performance. Once an agent has reached the limit of its willingness to reduce its standards, it may hold its q_i value constant. If the compromise set remains empty after some agent has reduced its q_i to its minimum acceptable level, then an impasse cannot be avoided. Such a society is dysfunctional. However, it may still be possible for some sub-society to break away from the larger group and continue to negotiate.

The compromise set represents the set of all decision vectors that are simultaneously acceptable to the group and to each member. Assuming that, perhaps as a result of negotiation, it is non-empty, there is no guarantee that it is a singleton; there may be multiple compromise decisions. The compromise set can be reduced by first eliminating any satisficing solutions that are dominated by superior solutions. For every $\mathbf{u} \in \mathbf{U}$ let

$$\begin{aligned} B_S(\mathbf{u}) &= \{\mathbf{v} \in \mathbf{U} : p_{\mathbf{R}}(\mathbf{v}) < p_{\mathbf{R}}(\mathbf{u}) \text{ and } p_{\mathbf{S}}(\mathbf{v}) \geq p_{\mathbf{S}}(\mathbf{u})\} \\ B_R(\mathbf{u}) &= \{\mathbf{v} \in \mathbf{U} : p_{\mathbf{R}}(\mathbf{v}) \leq p_{\mathbf{R}}(\mathbf{u}) \text{ and } p_{\mathbf{S}}(\mathbf{v}) > p_{\mathbf{S}}(\mathbf{u})\}, \end{aligned} \tag{28}$$

and define the set of alternatives that are *strictly better* than $\mathbf{u}$: $B(\mathbf{u}) = B_S(\mathbf{u}) \cup B_R(\mathbf{u})$; that is, $B(\mathbf{u})$ consists of all possible alternatives that are either less rejectable and not less selectable than $\mathbf{u}$, or more selectable and not more rejectable than $\mathbf{u}$. If $B(\mathbf{u}) = \varnothing$, then no alternative can be preferred to $\mathbf{u}$ in both selectability and rejectability. The *non-dominated* set is $\mathfrak{N}_G = \{\mathbf{u} \in \mathbf{U} : B(\mathbf{u}) = \varnothing\}$.

The *optimal compromise set* is

$$\Omega_G = \mathfrak{N}_G \cap \mathcal{C}_G. \tag{29}$$

All elements of the optimal compromise set lay claim to some notion of optimality. For example, $\mathbf{u}_M = \arg\max_{\mathbf{u}\in\Omega_G}\{p_{\mathbf{S}}(\mathbf{u}) - q_G p_{\mathbf{R}}(\mathbf{u})\}$ maximizes the difference between group selectability and group rejectability; $\mathbf{u}_R = \arg\min_{\mathbf{u}\in\Omega_G} p_{\mathbf{R}}(\mathbf{u})$ minimizes group rejectability, and $\mathbf{u}_S = \arg\max_{\mathbf{u}\in\Omega_G} p_{\mathbf{S}}(\mathbf{u})$ maximizes group selectability. The remaining elements of Ω_G represent other optimal tradeoffs between effectiveness and efficiency. If the indices of negotiation for the agents are sufficiently reduced, an optimal compromise will eventually exist.

Altruism

Social utilities and the satisficing solution concept provide a rigorous approach to the longstanding problem of altruism. While usually understood to mean that one is willing to sacrifice to benefit another, altruism could also take a malevolent form, in which an agent sacrifices to injure another. In either case, an altruistic agent by definition takes into consideration the preferences of others when defining its own preferences. By design, classical utilities accommodate only self-interest, and in this framework altruism can be accommodated only by redefining self-interest. This has proven highly problematic, not only philosophically but also in practice. It is possible to simulate benevolence or malevolence in particular situations, but the redefinition tends to be too specific and context dependent. In particular, it can not distinguish between *categorical altruism*, the willingness to always relinquish one's own self-interest, and *conditional altruism*, a willingness to relinquish one's own self-interest if, and only if, (a) the other wishes to take advantage of the offered largesse (for benevolent altruism), or (b) the other wishes to act in a way that elicits punishment (for malevolent altruism). This more sophisticated expression of altruism is simply not possible with classical utilities, since they are functions of possible player actions, not preferences for action.

In contrast, a socially rational agent may dynamically adjust its preferences as a function of the preferences of others. For example, suppose X_1 were willing to defer to X_2 by preferring u_1' if X_2 were to prefer u_2', otherwise, X_1 would prefer u_1''. This accommodation can be implemented by setting

$$p_{S_1|S_2}(u_1|u_2) = \left\{ \begin{array}{l} \begin{cases} 1 & u_1 = u_1',\ u_2 = u_2' \\ 0 & u_1 \neq u_1',\ u_2 = u_2' \end{cases} \\ \\ \begin{cases} 0 & u_1 = u_1',\ u_2 \neq u_2' \\ 1 & u_1 = u_1'',\ u_2 \neq u_2' \end{cases} \end{array} \right. . \tag{30}$$

Conditional social utilities permit the agent to examine each possible hypothetical situation and adjust its preferences as if the other agent actually most preferred to select or reject each of its possible alternatives. These specifications can be determined before the actual preferences of X_2 become available to X_1. Once p_{S_2} is given, X_1's marginal selectability becomes $p_{S_1}(u_1) = \sum_{u_2} p_{S_1|S_2}(u_1|u_2)p_{S_2}(u_2)$, which takes into account both its own and X_2's preferences. From this construction it is clear that, if X_2 does not strongly prefer u_2', then X_1's preferences are essentially unaltered. In this way, X_1 considers X_2's preferences but does not need to "throw the game" categorically in order to demonstrate a willingness to give deference to X_2.

THE ULTIMATUM GAME

The Ultimatum game is a much-studied example of a simple social relationship where it is difficult to reconcile observed behavior with the classical game-theoretic solution [5, 39–42]. In this two-player game, a proposer X_1 offers a responder X_2 a fraction of a sum of money, and X_2 chooses whether to accept (in which case the two divide the money as proposed) or to refuse (in which case neither player receives anything). In either case, the game is over. Within the framework of classical game theory, the unique subgame perfect equilibrium solution is for X_1 to offer the smallest possible non-zero amount, and for X_2 to accept what is offered. Interestingly, such a strategy is rarely adopted by human decision makers. Proposers are inclined to offer fair deals and responders are inclined to reject unfair deals.

Ultimatum is relevant to the multi-agent systems community as, for example, a model of bargaining and negotiation for various applications including electronic commerce [43]. Because of its simplicity, the game has become a prototype for decision problems where social behavior is not adequately explained by the hypothesis of individual rationality.

In an attempt to bring the classical game-theoretic results into line with experimental results, researchers in behavioral economics have proposed to alter the payoff for the Ultimatum game by modifying the utility functions. For example, Fehr and Schmidt [22] augment utility by non-pecuniary terms that account for both disadvantageous and advantageous inequity, and show that, with this modified utility, it is possible to achieve equilibria that are "fair" from the perspectives of both players. Their approach fundamentally changes the game, however, because the players adopt new utility functions that categorically re-define their preferences. As Sen observed: "It is possible to define a person's interests in such a way that no matter what he does he can be seen to be furthering his own interests in every act of choice . . . no matter whether you are a single-minded egoist or a raving altruist or a class-conscious militant, you will appear to be maximizing your own utility in this enchanted world of definitions" [44, p. 29].

The following analysis demonstrates that a socially rational formulation of the game provides a natural way to incorporate social attributes directly into the game description. It does not rely on an *ad hoc* redefinition of utility functions that, by their structure, are designed only for individually rational agents. Although the model makes specific predictions about behavior, we do not claim that this represents actual human behavior.

Classical Formulation of the Ultimatum Game

The action set for the proposer of the original Ultimatum game is a continuum (the unit interval). Fortunately, the minigame of Gale et al. [45], in which the proposer can make only one of two offers, preserves Ultimatum's essence while simplifying the analysis. Let X_i have the action set $U_i = \{u_i, u_i'\}$, and let S_i and R_i denote the respective selecting and rejecting selves for $i = 1, 2$. Let X_1's two offers be h and ℓ (high and low), with $0 < \ell \ll h \leq \frac{1}{2}$, corresponding to the fraction of the fortune offered to X_2. The responder's options are a (accept) and r (refuse). The standard payoff matrix for this minigame is displayed in Table 3. The unique Nash equilibrium is for X_1 to play ℓ and X_2 to play a.

Table 3: Payoff matrix for the Ultimatum minigame.

	X_2	
X_1	a	r
h	$(1-h,\ h)$	$(0,\ 0)$
ℓ	$(1-\ell,\ \ell)$	$(0,\ 0)$

In the classical formulation, social coherence can be violated by the exogenous imposition of a social preference. For example, there is strong empirical evidence that groups reveal a preference for fair treatment [46, 47]. In the minigame, individual rationality dictates that $\ell \succ_1 h$. If fair treatment is imposed on the minigame then $(h, a) \succ_G (\ell, a)$ and $(h, r) \succeq_G (\ell, r)$. Since X_1's preference is never preferred by the group, social coherence is violated. In this case at least, empirical behavior appears to be more consistent with social coherence than with individual rationality.

Satisficing Formulation of the Ultimatum Game

To frame Ultimatum as a satisficing game, the payoff matrix must be replaced with a social utility function that accounts for the preferences of the four selves involved: S_1, S_2, R_1, and R_2. In this formulation the *intemperance* of X_1 and the *indignation* of X_2 are the dominant social attributes of the players. These attributes are denoted by an intemperance index τ and an indignation index δ, respectively, where $0 \leq \tau \leq 1$ and $0 \leq \delta \leq 1$. The condition $\tau \approx 1$ means that X_1 is extremely avaricious, while $\tau \approx \frac{1}{2}$ means that X_1 is willing to restrain its desire for wealth. The condition $\delta \approx 1$ means that X_2 is easily offended, while $\delta \approx 0$ means that X_2 is easily pleased. For the present purpose, assume that these parameters are fixed properties of the players. X_1 may temper its avarice because of benevolence toward X_2, because of an aversion to inequity, or because of suspicion that X_2 may refuse an unfair offer—the precise motive is not important here. The key point is that the parameters are treated as endogenous attributes, not the result of exogenous forces that cause the players to change their utilities.

There is not a unique way to define the selecting and rejecting selves S_1, S_2, R_1, and R_2 of the players, but it is reasonable to associate the selecting self with the goal of the game, which is to receive as much of the fortune as possible. The rejecting self is associated with the efficiency with which the goal is pursued. This attribute, however, must be independent of effectiveness, and hence cannot be a function of the ratio of the fortune one receives. It must therefore be a function of whether or not *any* reward is received. Thus, both players are inefficient if, and only if, the responder refuses the offer.

Since X_1 plays first, X_1's utility structure need not be conditioned on X_2's response (although this remains a possibility). Thus, X_1's social utilities are unconditional. S_1's selectability (as a function of intemperance) is expressed as

$$p_{S_1}(h) = 1 - \tau \quad \text{and} \quad p_{S_1}(\ell) = \tau. \tag{31}$$

S_1's rejectability is a function of X_2's indignation as well as its own intemperance, and is expressed as

$$p_{R_1}(h) = \tau(1-\delta) \quad \text{and} \quad p_{R_1}(\ell) = 1 - \tau(1-\delta); \tag{32}$$

that is, if X_2 were highly indignant ($\delta \approx 1$), offering the high fraction to the responder would have low rejectability.

Since X_2 makes the second move, the preferences of X_2's selves will be conditioned on X_1's choice. Define $p_{S_2|S_1}$ as

$$p_{S_2|S_1}(a|h) = 1 \quad \text{and} \quad p_{S_2|S_1}(r|h) = 0. \tag{33}$$

If, however, S_1 were to select ℓ, then S_2 would be indignant and would prefer to select r with weight δ and a with weight $1 - \delta$. Thus,

$$p_{S_2|S_1}(a|\ell) = 1 - \delta \quad \text{and} \quad p_{S_2|S_1}(r|\ell) = \delta. \tag{34}$$

Next consider X_2's rejecting self, R_2. If S_1 were to select h, then R_2 would prefer to reject r. Thus,

$$p_{R_2|S_1}(a|h) = 0 \quad \text{and} \quad p_{R_2|S_1}(r|h) = 1.$$

If S_1 were to select ℓ, then R_2 would be indignant and would prefer to reject a with weight δ and r with weight $1-\delta$. Thus,

$$p_{R_2|S_1}(a|\ell) = \delta \quad \text{and} \quad p_{R_2|S_1}(r|\ell) = 1-\delta.$$

The resulting two-agent interdependence function can be factored according to the chain rule of probability. In general, there are many ways to apply this rule, but for this application, an obvious factorization is

$$p_{S_1S_2R_1R_2}(u_1,u_2;v_1,v_2) = p_{S_2|S_1R_1R_2}(u_2|u_1;v_1,v_2) \cdot p_{R_1|S_1R_2}(v_1|u_1;v_2)\cdot p_{R_2|S_1}(v_2|u_1)\cdot p_{S_1}(u_1). \quad (35)$$

Consider each of the factors on the right hand side of (35) in turn. In the Ultimatum game, S_2's conditional preferences depend only on the choice S_1 makes and are not influenced by R_1 or R_2, so $p_{S_2|S_1R_1R_2}$ reduces to $p_{S_2|S_1}$. Because R_1 is not influenced by any other self, the term $p_{R_1|S_1R_2}$ may be replaced by p_{R_1}. The final terms are already in a form defined by the game model. Substituting the simplified terms into (35) yields

$$p_{S_1S_2R_1R_2}(u_1,u_2;v_1,v_2) = p_{S_1}(u_1)\cdot p_{S_2|S_1}(u_2|u_1)\cdot p_{R_1}(v_1)\cdot p_{R_2|S_1}(v_2|u_1). \quad (36)$$

The Ultimatum game provides an opportunity for categorical altruism by the proposer, and conditional altruism by the responder [see (31)]. If the proposer is purely selfish or completely intemperate ($\tau = 1$), no deference would be shown for the responder's welfare (or no concern for retaliation). However, if $\tau < 1$, the proposer accommodates the preferences of the responder (or acts to reduce the potential for retaliation) at its own expense. The responder exhibits malevolent conditional altruism (to the degree defined by δ) by sacrificing its own welfare to punish the proposer for an unfair offer [see (34)]. But if the proposer makes a fair offer, then the responder's utility function reflects acceptance [see (33)].

The Satisficing Rectangle

The Ultimatum minigame is such that its marginals are easily computed without resorting to a formal algorithm. Applying (15) and (16), then (19) and (20) in sequence, the selectability and rejectability marginals for the responder are

$$\begin{aligned} p_{S_2}(u_2) &= \sum_{\{u_1\}}\sum_{\{v_1\}}\sum_{\{v_2\}} p_{S_2|S_1}(u_2|u_1)\cdot p_{R_1}(v_1)\cdot p_{R_2|S_1}(v_2|u_1)\cdot p_{S_1}(u_1) \\ &= \sum_{\{u_1\}} p_{S_2|S_1}(u_2|u_1)\cdot p_{S_1}(u_1) \\ &= \begin{cases} 1-\tau\delta & \text{for } u_2 = a \\ \tau\delta & \text{for } u_2 = r \end{cases} \end{aligned} \quad (37)$$

and

$$\begin{aligned} p_{R_2}(v_2) &= \sum_{\{u_1\}}\sum_{\{u_2\}}\sum_{\{v_1\}} p_{S_2|S_1}(u_2|u_1)\cdot p_{R_1}(v_1)\cdot p_{R_2|S_1}(v_2|u_1)\cdot p_{S_1}(u_1) \\ &= \sum_{\{u_1\}} p_{R_2|S_1}(v_2|u_1)\cdot p_{S_1}(u_1) \\ &= \begin{cases} \tau\delta & \text{for } v_2 = a \\ 1-\tau\delta & \text{for } v_2 = r \end{cases}. \end{aligned} \quad (38)$$

Now recall that the satisficing set for a decision maker is the set of actions for which the selectability is at least as great as the product of q, the index of caution, and the rejectability. Thus, setting $q = 1$ and comparing (31) with (32) and (37) with (38), the satisficing sets for the proposer and responder are

$$\begin{aligned} \Sigma_1(\tau,\delta) &= \{u_1 \in U_1 \colon p_{S_1}(u) \ge p_{R_1}(u_1)\} \\ &= \begin{cases} \{h\} & \text{if } \tau < \frac{1}{2-\delta} \\ \{\ell\} & \text{if } \tau > \frac{1}{2-\delta} \\ \{h,\ell\} & \text{if } \tau = \frac{1}{2-\delta} \end{cases} \\ \Sigma_2(\tau,\delta) &= \{u_2 \in U_2 \colon p_{S_2}(u) \ge p_{R_2}(u_1)\} \\ &= \begin{cases} \{a\} & \text{if } \tau < \frac{1}{2\delta} \\ \{r\} & \text{if } \tau > \frac{1}{2\delta} \\ \{a,r\} & \text{if } \tau = \frac{1}{2\delta} \end{cases}. \end{aligned}$$

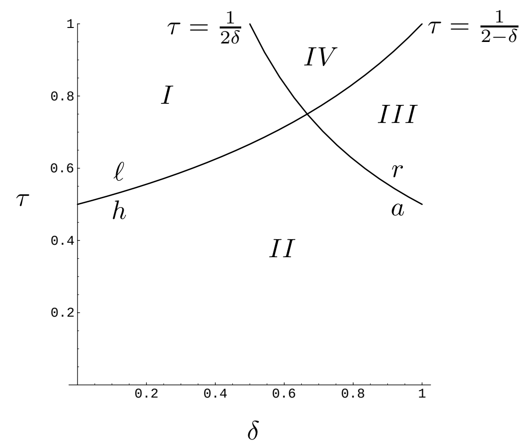

Figure 3: (τ,δ) regions for the satisficing rectangles.

Figure 3 displays the satisficing rectangle for the Ultimatum game as a function of τ and δ. Values of τ and δ that lie above the curve labeled $\tau = \frac{1}{2-\delta}$ (regions I and IV) result in the proposer offering a low fraction, and values that lie below this curve result in a high fraction. For (τ,δ) pairs that lie on the line, both h and ℓ are satisficing for the proposer. Next, consider the responder. Values of τ and δ that lie above the curve labeled $\tau = \frac{1}{2\delta}$ (regions III and IV) result in the responder refusing the offer, and values that lie below this curve result in accepting the offer. For (τ,δ) pairs that lie on the line, both a and r are satisficing for the responder. The two curves divide the (τ,δ) square into four regions, corresponding to four different satisficing rectangles (ignoring boundaries):

$$\begin{aligned} \mathcal{R}_G(\tau,\delta) &= \Sigma_1(\tau,\delta)\times\Sigma_2(\tau,\delta) \\ &= \begin{cases} \{(\ell,a)\} & \text{for } (\tau,\delta)\in I \\ \{(h,a)\} & \text{for } (\tau,\delta)\in II \\ \{(h,r)\} & \text{for } (\tau,\delta)\in III \\ \{(\ell,r)\} & \text{for } (\tau,\delta)\in IV \end{cases}. \end{aligned}$$

In region I, a low fraction is offered and accepted, which is the Nash solution. It obtains when the proposer is intemperate and the responder is not readily indignant. In region II, a high fraction is offered and accepted. This solution obtains when the proposer is temperate and the responder reasonable. In region III, a high fraction is offered and refused, revealing an unreasonable indignation on the part of the responder. In region IV, a low offer is refused, since the responder is indignant in the face of an intemperate proposer.

Group Satisficing Solutions

The satisficing set for a society constitutes all joint actions that are good enough for the society collectively. For the Ultimatum game with $q = 1$,

$$\boldsymbol{\Sigma}_G(\tau, \delta) = \{(u_1, u_2) \in U_1 \times U_2 \colon p_{S_1S_2}(u_1, u_2) \geq p_{R_1R_2}(u_1, u_2)\},$$

where the joint selectability and joint rejectability mass functions are computed as

$$\begin{aligned} p_{S_1S_2}(u_1, u_2) &= \sum_{\{v_1\}} \sum_{\{v_2\}} p_{S_2|S_1}(u_2|u_1) \cdot p_{R_1}(v_1) \\ &\quad \cdot p_{R_2|S_1}(v_2|u_1) \cdot p_{S_1}(u_1) \\ p_{R_1R_2}(v_1, v_2) &= \sum_{\{u_1\}} \sum_{\{u_2\}} p_{S_2|S_1}(u_2|u_1) \cdot p_{R_1}(v_1) \\ &\quad \cdot p_{R_2|S_1}(v_2|u_1) \cdot p_{S_1}(u_1). \end{aligned}$$

The resulting joint selectability and joint rejectability functions are

$$\begin{aligned} p_{S_1S_2}(h, a) &= 1 - \tau \\ p_{S_1S_2}(h, r) &= 0 \\ p_{S_1S_2}(\ell, a) &= \tau - \tau\delta \\ p_{S_1S_2}(\ell, r) &= \tau\delta \end{aligned}$$

and

$$\begin{aligned} p_{R_1R_2}(h, a) &= \tau^2\delta - \tau^2\delta^2 \\ p_{R_1R_2}(h, r) &= \tau - \tau\delta - \tau^2\delta + \tau^2\delta^2 \\ p_{R_1R_2}(\ell, a) &= \tau\delta - \tau^2\delta + \tau^2\delta^2 \\ p_{R_1R_2}(\ell, r) &= 1 - \tau + \tau^2\delta - \tau^2\delta^2. \end{aligned}$$

Setting $q = 1$, the jointly satisficing set for the group is obtained by comparing the above functions for each joint action. Figure 4 displays this set as a function of τ and δ as defined by

$$\boldsymbol{\Sigma}_G(\tau, \delta) = \begin{cases} \{(\ell, a)\} & \text{for } (\tau, \delta) \in A \\ \{(\ell, a), (\ell, r)\} & \text{for } (\tau, \delta) \in B \\ \{(\ell, r)\} & \text{for } (\tau, \delta) \in C \\ \{(\ell, a), (\ell, r), (h, a)\} & \text{for } (\tau, \delta) \in D \\ \{(\ell, r), (h, a)\} & \text{for } (\tau, \delta) \in E \\ \{(h, a), (\ell, a)\} & \text{for } (\tau, \delta) \in F \\ \{(h, a)\} & \text{for } (\tau, \delta) \in G \end{cases}.$$

Notice that regions B, D, E, and F do not have unique solutions. In region B, for example, both joint actions (ℓ, a)

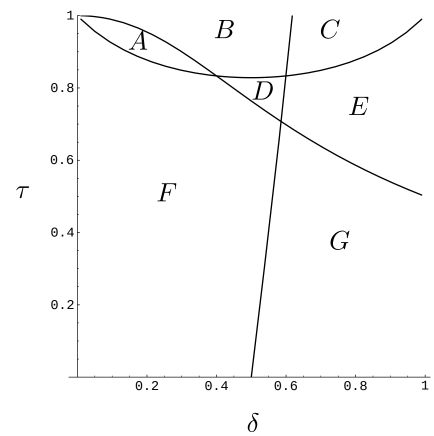

Figure 4: (τ, δ) regions for the jointly satisficing sets.

and (ℓ, r) are jointly satisficing for the (τ, δ) values in that region (high intemperance, low to moderate indignation). The society, if it were to act as a single entity, would not reject either of these joint actions—either would be good enough.

It is important to understand that a jointly satisficing set is a purely context-free mathematical result that need not have an obvious operational interpretation, but it is certainly possible to impose one. For regions A, F, and G, a group preference to obtain the fortune can be deduced. For region E, there is a group preference to be fair. For region D a possible group preference is to at least do something that is logical (such as not punishing without cause). For region C, the only possible group preference is to be dysfunctional. Region B amounts to indifference.

Compromise

By inspection, the compromise set for the Ultimatum game is

$$\mathcal{C}_G(\tau, \delta) = \mathcal{R}_G(\tau, \delta) \cap \boldsymbol{\Sigma}_G(\tau, \delta) = \begin{cases} (\ell, a) & \text{for } (\tau, \delta) \in I \cap (A \cup B \cup D \cup F) = I \setminus E \\ (h, a) & \text{for } (\tau, \delta) \in II \cap (E \cup F \cup G) = II \\ \varnothing & \text{for } (\tau, \delta) \in (III \cup I) \cap E \\ (\ell, r) & \text{for } (\tau, \delta) \in IV \cap (B \cup C \cup E) = IV \end{cases}.$$

Figure 5 displays the (τ, δ) regions that correspond to joint actions in the compromise set. For all $(\tau, \delta) \notin (III \cup I) \cap E$, there is a unique pair of individual choices consistent with the society's choice—an optimal compromise. It is possible to operationalize society/individual satisficing for each of these joint actions. For joint actions (h, a) and (ℓ, a), the optimal compromise is to share the fortune. For the joint action (ℓ, r), it is to be dysfunctional—players are so mismatched in temperament that failure cannot be avoided. The joint action (h, r) is never a compromise solution.

COORDINATED GRAPH SEARCH

A team of mobile autonomous agents is tasked to visit all of the nodes of an undirected graph whose edges represent paths between the nodes. Each agent possesses a copy of the graph and is able to communicate with other agents within a speci-

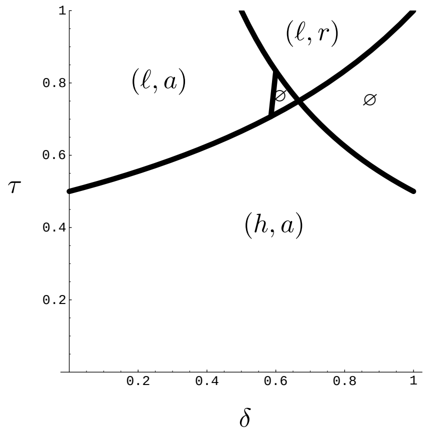

Figure 5: (τ, δ) regions for the compromise set.

fied radius. As long as the group maintains full communication connectivity (relaying messages if necessary), every agent will be continuously updated regarding which nodes have been visited. The performance criteria for this society is that (a) all nodes are visited by at least one agent such that duplication is kept small, and (b) full communication connectivity is maintained. It is assumed that the agents are homogeneous, the nodes are of equal value to the search effort, and the communication range is identical for all agents. When at a node, an agent's option set consists of the node it is at and all adjacent nodes. While on an edge, its option set will consist of the nodes that define the edge.

This scenario can be viewed as a generalized n-agent version of the classic traveling salesman problem, the complexity of which grows combinatorically with the number of nodes. The computational complexity of an optimal solution to this n-agent problem grows even faster, since it depends on the number of nodes in the graph, the number of agents, and the need to satisfy the communication constraint. Although it is impractical to compute the optimal solution, there are many ways to formulate a satisficing solution.

The Social Model

Since the search scenario is dynamic, the associated social utilities must evolve in time as agents make decisions and traverse the graph. It is assumed that time flows in discrete steps at each time $t = 0, 1, 2, \ldots$, and that each agent has knowledge of the positions of all agents as well as the status of all nodes (visited or not visited) up to time $t-1$. Thus, the social utilities for this problem will be functions of time as well as of the options.

Many social models could be defined to characterize this application. In contrast with the Ultimatum game, in which behavior is conditioned on the preferences of others, we employ *two* social models—one for the group and one for the individuals—and identify solutions that are consistent with both. This approach demonstrates the versatility of the satisficing approach to group and individual decision making.

Although two separate models are employed, both use the same operational definitions of selectability and rejectability. Recall that the aim of the decision maker (individual or group) is to conserve resources (associated with rejectability) while avoiding failure (associated with selectability). In the context of this game, failure occurs if one or more nodes are not visited, and resources are consumed if connectivity is lost.

The Individual Model

The individual model requires each agent to calculate its selectability and rejectability marginals at each time t using only knowledge of the state of the system (i.e., agent positions and node status) but without explicitly accounting for the future preferences of other agents. Because the calculations are simple, each agent can determine the satisficing sets for all agents and hence the entire satisficing rectangle at each time t.

Individual selectability p_{S_i} for each agent is determined as follows. If the option set contains unvisited nodes, then

$$p_{S_i}(u;t) \propto \begin{cases} 1 & \text{if } u \text{ has been visited prior to time } t \\ 5 & \text{if } u \text{ has not been visited prior to time } t \end{cases},$$

where the symbol $\propto$ implies that the values are normalized to become mass functions. If there are no unvisited nodes in the agent's option set, the agent performs a breadth-first search of the nodes connected to each of the nodes in its action set. Each such action u is then weighted based on the depth, $D(u)$, of the first unvisited node encountered if u is adopted. A smaller depth will receive a larger weight. These weights are then normalized across the action space, yielding

$$p_{S_i}(u;t) \propto \frac{1}{D(u)}. \tag{39}$$

Since it is associated with the loss of connectivity, the rejectability of an action for agent X_i should be a function of the current distance d_{ic} between X_i and its nearest neighbor. If taking action u causes d_{ic} to approach d_{max}, the maximum communication range, the rejectability must increase sharply in order to avoid communication failure. Let $d_i(u)$ denote the distance between X_i and its nearest neighbor that will result if action u is taken. A simple mechanism to increase rejectability in proportion to the propensity of an option to result in communication failure is

$$p_{R_i}(u;t) \propto 2^{\max\{1, \frac{d_i(u)}{d_{ic}}, \frac{2d_i(u)}{d_{max}}\}}. \tag{40}$$

Thus, if no action increases the distance to the nearest neighbor, or moves the agent to more than half of the maximum communication distance from its nearest neighbor, the rejectability of all options is uniform. But if an action causes the agent's position to exceed either of those limits, then the rejectability of that action increases exponentially with the amount the limitation is exceeded.

Each agent then forms its individually satisficing set as

$$\Sigma_i(q_i, t) = \{u_i \in U_i\colon p_{S_i}(u_i;t) \geq q_i p_{R_i}(u_i;t)\}. \tag{41}$$

The satisficing rectangle is then

$$\Re(t) = \Sigma_1(q_1, t) \times \Sigma_2(q_2, t) \times \cdots \times \Sigma_n(q_n, t). \tag{42}$$

The Group Model

Each agent can also compute its version of a jointly satisficing set. To do this, it needs operational definitions of what is selectable and rejectable for the group. Since this society is cooperatively disposed, a simple group social model is immediate: group selectability is proportional to the number of unvisited and unique nodes directly accessible by the group, and group rejectability is determined by the loss of connectivity.

The selectability at time t of a group action vector $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbf{U} = U_1 \times U_2 \times \cdots \times U_n$ is proportional to: (a) the number of unique unvisited nodes $\mathbf{u}$ reaches and (b) the total number of unique nodes accessible by $\mathbf{u}$ (whether or not visited). The first criterion encourages the group to visit new nodes, and the second encourages the group to spread out and seek new territory. For example, in a five-agent system, if $\mathbf{u}$ causes the group to move to four unique unvisited nodes and one visited node, it will receive a larger selectability than $\mathbf{u}'$, which causes the group to move to three unvisited nodes and two visited nodes. If, however, there are no currently available options that will take the group to unvisited nodes, then the group will at least be encouraged to move to as many different nodes as possible, thereby spreading out in an attempt to seek the remaining unvisited nodes. To formalize this structure, let $N_{nv}(\mathbf{u}, t)$ denote the number of unique unvisited nodes accessible by joint action $\mathbf{u}$ at time t, and let $N_{tot}(\mathbf{u}, t)$ denote the total number of unique nodes (whether or not visited) accessible by $\mathbf{u}$ at time t. Then the group selectability at time t is given by

$$p_{S_G}(\mathbf{u}, t) \propto N_{nv}(\mathbf{u}, t) + N_{tot}(\mathbf{u}, t). \quad (43)$$

Since the group as a whole shares the requirement to maintain connectivity, group rejectability is concerned with, and only with, maintaining group connectivity. This criterion results in an extremely simple group rejectability function:

$$p_{R_G}(\mathbf{u}, t) = \begin{cases} 0 & \text{if } \mathbf{u} \text{ maintains comm. at time } t \\ 1 & \text{otherwise} \end{cases}. \quad (44)$$

All joint options that do not disrupt connectivity are jointly satisficing, and the joint satisficing set has the form

$$\mathbf{\Sigma}_G(t) = \{\mathbf{u} \colon p_{R_G}(\mathbf{u}, t) = 0\}. \quad (45)$$

The Negotiation Process

Since each agent is able to compute the satisficing rectangle and the jointly satisficing set, it may also form the compromise set

$$\mathcal{C}_G(t) = \mathcal{R}_G(t) \cap \mathbf{\Sigma}_G(t). \quad (46)$$

If $\mathcal{C}_q$ is empty, then the negotiation mechanism of satisficing game theory is used: agents incrementally lower their q_i values until the compromise set is no longer empty. This models a form of autonomous negotiation in which each agent compromises and gives up a little performance to find a solution that is satisfactory to the entire group.

If the compromise set contains more than one vector, ties are broken by a simple lexicographical ordering. At the first level, only those jointly satisficing options most selectable to the group (visiting the greatest number of new nodes) are retained. If this set contains multiple elements, the second level of ordering is to rank the agents dynamically based on the number of unvisited nodes in their option sets. The agents then select the vector which is most *individually* selectable to the highest-ranking agent. Further ties are broken by deferring to the individual preferences of lower-ranking agents until a unique solution results.

A Simple Example: 5x5 Grid

As a simple example, consider a square grid of 25 nodes with nearest neighbor connections. Five agents start in the lower left corner and are assigned to traverse the graph with a communication radius of 1.6 times the distance between nodes. The agents, denoted a_1 through a_5, travel along paths depicted in Figure 6.

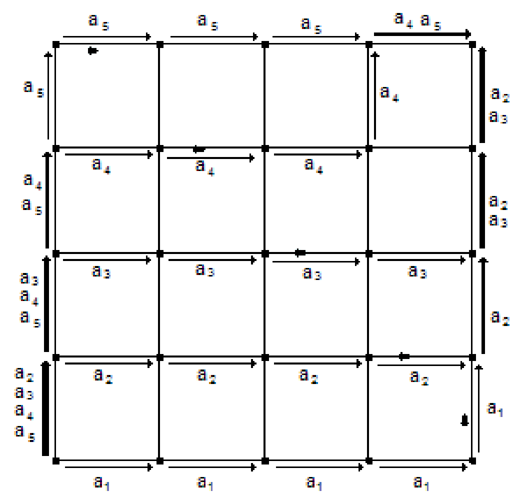

Figure 6: 5x5 Traversal Paths

The group selectability p_{S_G} encourages the agents to fan out as they traverse the graph. After simultaneously visiting the middle nodes, they converge as they approach the top right corner of the graph. A minimum number of nodes are visited by more than one agent, and the graph is searched quickly and efficiently. Connectivity is maintained throughout, allowing information to be relayed to all agents.

Random Graph Simulations

More complicated graphs were created by randomly generating 100 nodes within a 25×25 region using a uniform distribution. Edges were added to connect each node with its four nearest neighbors. An example graph is shown in Figure 7. Three search scenarios were conducted for each graph: (a) a five-agent search with a communications range of two units; (b) a three-agent search with a communication range of three units; and (c) a single-agent search. For consistency, agents begin their search from a randomly selected node near the center of the graph.

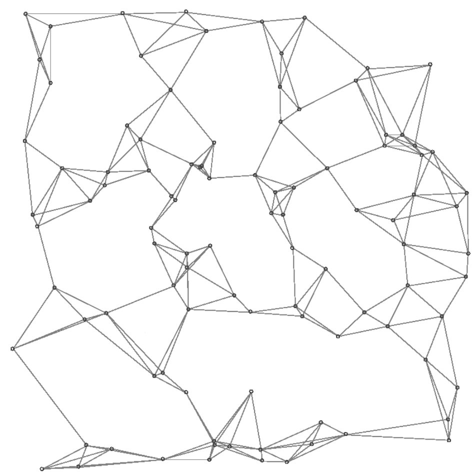

Figure 7: Sample Random Graph with 100 Nodes

Ideally, simulation results should be compared with those of an optimal search, but this is computationally infeasible for graphs of this complexity. To create a baseline for comparison, we implemented a boundedly-rational single agent algorithm that searches until the search cost exceeds the anticipated improvement of further search. The algorithm systematically conducts a depth-first search of possible tours starting from each node in the graph. After 30 minutes of CPU time, the algorithm switches to the next node in the graph as the starting point. Each level in the search tree corresponds to the choice of next unvisited node to visit; a greedy algorithm is employed, and nodes are considered in the order of their proximity to the previous node.

The boundedly-rational search algorithm differs from the satisficing search algorithm in three significant ways. First, it is free to start from any node in the graph. Second, its decisions are based on an extensive global search rather than conditions local to the agents and limited look-ahead, so it is poorly suited to real-time operation. Finally, since it involves a single agent, it has no movement constraints to ensure communication is maintained. Thus, although the boundedly-rational results provide bounds on the efficiency of satisficing multi-agent search algorithms, the bounds are not particularly tight.

Table 4 displays results averaged over 10 graphs for the three satisficing scenarios (top three rows) along with the boundedly-rational baseline (bottom row). From left to right, the columns list the number of agents, the communication radius, the average number of nodes visited per agent during the search, a measure of search efficiency referred to as *node efficiency*, the average distance traveled per agent during the search, and a second measure of efficiency called *distance efficiency*. Node efficiency is a measure of the extent to which duplication of visited nodes is avoided; the repetition exhibited by the boundedly-rational search is considered a lower-bound (total nodes visited prorated per agent). The measure is given by the ratio of this lower bound to the actual average number visited by each agent. Distance efficiency is determined in a similar manner, as the ratio of the lower bound of distance traveled in the baseline (total distance traveled prorated per agent) to the actual average distance traveled per agent.

Table 4: Results averaged over 10 randomly generated graphs. The bottom row is the boundedly-rational baseline.

Agts	Com Rad	Nds/Agt	Nd Eff	Dst/Agt	Dst Eff
5	2	40	56%	17.3	50%
3	3	52	71%	21.5	67%
1	n/a	121	92%	46.3	94%
1^*	n/a	111	100%	43.5	100%

As the number of agents increases, each agent visits fewer nodes and travels a shorter distance, but both measures of efficiency also decrease. There are three principal causes of this behavior. First, all agents start at the same point, and it takes time to spread out and search the graph. Second, agents must remain near each other to remain in communication. Third, certain nodes in the sparse graph become natural hubs for agents to remain and serve as communication relays.

The simulations provide numerous examples of emergent behavior that is not anticipated by the modeling assumptions. For example, as the group spreads out, the more central agents tend to sacrifice individual preferences (searching new territory) to maintain the connectivity of the group by remaining stationary while outlying agents continue to search, and this continues until the outlying agents return to the group. The willingness to act as a relay is not explicitly programmed into the agents, and it is a significant example of emergent coordination as agents balance the interests of the group with their individual interests.

Figure 8 illustrates typical emergent behavior. The small circle representing each node is filled if visited, and the large circle centered on each agent represents its communication range. In the first frame, all three agents are heading towards unique, unvisited nodes. In the second frame, a_1 and a_2 are moving to the left while a_3 moves to the right. In the third frame, the agents are approaching the limits of their communication range. To avoid communication loss, a_2 is returning to an area with no unvisited nodes so that it can relay messages between a_1 and a_3, as they cannot directly communicate. In the fourth frame, a_1 and a_3 continue their search, facilitated by a_2's sacrifice.

DISCUSSION

The Ultimatum game and the random graph search illustrate the wide range of applications for which satisficing game theory is applicable. In the Ultimatum game, the behavior of the players is governed by their social attributes, and the emergent group-level preferences can range from fair and well-coordinated to dysfunctional and ill-coordinated. The Ultimatum game is a static (one-move per player) game, whereas the graph search is a dynamic decision problem. In the graph-search problem, the goal of the agents is to find functional

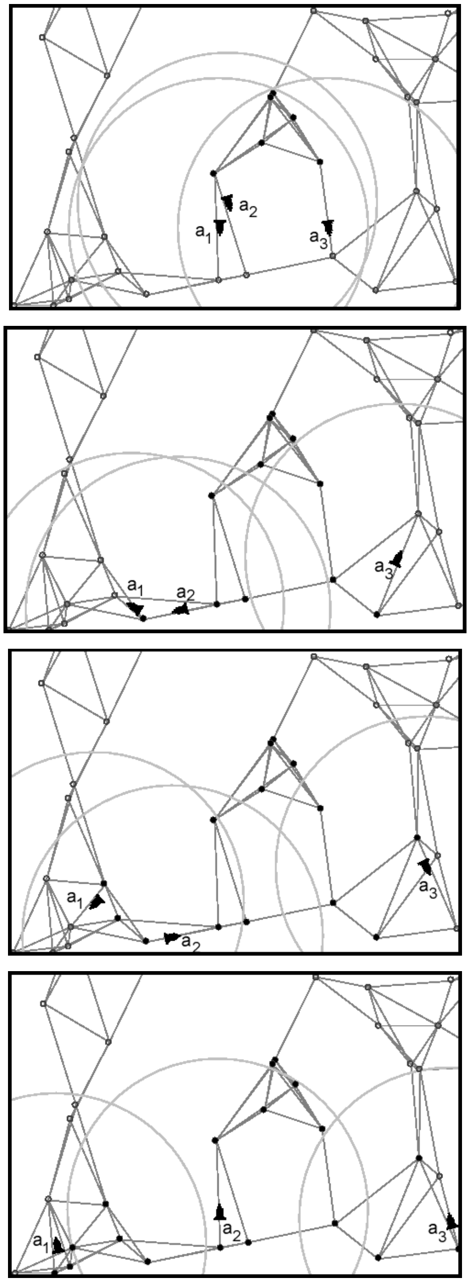

Figure 8: A sequence of frames showing emergent coordinated behavior.

compromises as the agents function in a multi-move dynamic environment. Emergent behavior is manifest by the agents as they self-select the different roles of searcher and communication relay.

The two applications demonstrate the robustness of the failure-avoidance formulation of a decision problem. With the Ultimatum game, the classical formulation as utility maximization is a poor model of actual human behavior. Formulated in terms of failure avoidance, the social attributes of the agents can be explicitly included in the model as parameters, and their effect on the decisions can be observed. A traditional utility-maximization formulation of the graph search problem would require the identification of an optimal strategy (most likely in the sense of bounded rationality), which would require extensive computation. Even if an optimal solution for the group were known, there would be no way to determine whether it would be acceptable to each of the individual agents. With the failure-avoidance formulation, however, it is possible to identify in real-time potential solutions that are simultaneously acceptable to all agents and to the group. The key to this success is that the agents, and the group as a whole, each identify multiple choices that are good enough according to a mathematically precise criterion, thus providing them with negotiating room to settle on a compromise.

Conflict can be resolved either by competition or by cooperation. Because it is based on the premise of individual rationality, classical game theory virtually forces competition, even if cooperation would be more natural and beneficial to the players individually and to the group as a whole. On the other hand, satisficing game theory permits conflict to be treated either as competition or cooperation with equal facility. It therefore provides a neutral foundation upon which to frame decision problems.

Satisficing game theory provides a framework within which procedures for negotiation and compromise can be implemented. By contrast, solution concepts associated with individual rationality provide existence proofs only; they are not constructive. For example, although a Nash equilibrium always exists, the theory does not provide a procedure for the agents to follow to arrive at the prescribed solution. Also, rather than simulating social relationships with an intrinsically asocial model, conditional altruism—a willingness either to defer to or to punish others according to the situation—can be modeled without completely redefining the game.

Social utilities and satisficing game theory together provide a systematic methodology for the design of socially rational multi-agent systems. Domain knowledge and the rules of the game are encoded into the social utility functions that compose the interdependence function. Because they possess the mathematical properties of multivariate probability mass functions, social utilities account for all possible social interactions, take advantage of whatever independence and conditioning properties are relevant, and make it impossible to assign high selectability (or rejectability) to an action set and its complement. Satisficing game theory makes explicit the conditions under which group and individual preferences can be reconciled, and provides a mechanism for altruism, negotiation and compromise.

Social rationality and satisficing game theory generalize the notions of individual rationality and competitive optimality, but at a price. The computational burden of the satisficing approach grows combinatorically if all possible interconnections between agents are considered. Fortunately, as with Bayesian networks, it is reasonable to expect that in practical situations and especially for large groups, the connectivity of praxeic networks will be somewhat sparse. The use of hierarchical and Markov structures can simplify the construction of the interdependence function and further reduce computation. Multi-agent decision making is, by its very nature, a complex enterprise. Moreover, the enterprise of analyzing complex natural systems or synthesizing complex artificial systems is intrinsically difficult and challenging. But, as Palmer observed, "Complexity is no argument against a theoretical approach if the complexity arises not out of the theory itself but out of the material which any theory ought to handle" [48]. While it is desirable in the interest of tractability to simplify a social model as much as possible, eliminating important social relationships in order to comply with individual rationality may provide an inadequate model of the society under investigation.

A key contribution of this paper is the demonstration that the syntax of probability theory, which is of immense value as a means of modeling epistemological phenomena, can also be of value as a model of praxeological phenomena. In a personal communication to Judea Pearl, Glenn Shafer observed that "probability is not really about numbers; it is about the structure of reasoning" [35, p 15]. The thesis of this paper is that the mathematical structure of probability theory *is also about the structure of coherent social interaction.* The combination of social utilities and satisficing game theory forms the basis of a unified treatment of group and individual decision making.

APPENDIX: PROOF OF THE SOCIALITY THEOREM

Sufficiency is established by setting

$$F[p_{G^j}(\mathbf{u}_j), p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)] = p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)p_{G^j}(\mathbf{u}_j). \quad (47)$$

The non-negativity condition ensures that F is non-decreasing in both arguments.

To prove necessity, let G^i, G^j and G^k be arbitrary pairwise disjoint sub-societies of G, and let $p_{G^i\cup G^j\cup G^k}$, $p_{G^i|G^j\cup G^k}$, $p_{G^i\cup G^j|G^k}$, $p_{G^i\cup G^j}$, $p_{G^i|G^j}$, and p_{G^i} be endogenously aggregated non-negative functions; that is, there exists a function F such that $p_{G^i\cup G^j}(\mathbf{u}_i, \mathbf{u}_j) = F[p_{G^j}(\mathbf{u}_j), p_{G^i|G^j}(\mathbf{u}_i|\mathbf{u}_j)]$ $\forall(\mathbf{u}_i, \mathbf{u}_j) \in \mathbf{U}_i \times \mathbf{U}_j$.

By framing invariance,

$$\begin{aligned} F\big[p_{G^j\cup G^k}(\mathbf{u}_j, \mathbf{u}_k), p_{G^i|G^j\cup G^k}(\mathbf{u}_i|\mathbf{u}_j, \mathbf{u}_k)\big] \\ = F\big[p_{G^k}(\mathbf{u}_k), p_{G^i\cup G^j|G^k}(\mathbf{u}_i, \mathbf{u}_j|\mathbf{u}_k)\big]. \quad (48) \end{aligned}$$

But

$$p_{G^j\cup G^k}(\mathbf{u}_j, \mathbf{u}_k) = F\big[p_{G^k}(\mathbf{u}_k), p_{G^j|G^k}(\mathbf{u}_j|\mathbf{u}_k)\big] \quad (49)$$

and

$$\begin{aligned} p_{G^i\cup G^j|G^k}(\mathbf{u}_i, \mathbf{u}_j|\mathbf{u}_k) \\ = F\big[p_{G^j|G^k}(\mathbf{u}_j|\mathbf{u}_k), p_{G^i|G^j\cup G^k}(\mathbf{u}_i|\mathbf{u}_j, \mathbf{u}_k)\big]. \quad (50) \end{aligned}$$

Thus,

$$\begin{aligned} F\left[F\big[p_{G^k}(\mathbf{u}_k), p_{G^j|G^k}(\mathbf{u}_j|\mathbf{u}_k)\big], p_{G^i|G^j\cup G^k}(\mathbf{u}_i|\mathbf{u}_j, \mathbf{u}_k)\right] \\ = F\left[p_{G^k}(\mathbf{u}_k), F\big[p_{G^j|G^k}(\mathbf{u}_j|\mathbf{u}_k), p_{G^i|G^j\cup G^k}(\mathbf{u}_i|\mathbf{u}_j, \mathbf{u}_k)\big]\right]. \end{aligned} \quad (51)$$

In terms of general arguments, this equation becomes

$$F\left[F(x, y), z\right] = F\left[x, F(y, z)\right], \quad (52)$$

called the *associativity equation*, which has been studied extensively [49, 50]. It has been shown by [51] (see also [52]) that if F is differentiable in both arguments, then the general solution of (52) is

$$f[F(x, y)] = f(x)f(y) \quad (53)$$

for some positive continuous monotonic increasing function f, which is otherwise arbitrary. Taking f as the identity function, we obtain

$$\begin{aligned} p_{G^i\cup G^j}(\mathbf{u}_i, \mathbf{u}_j) &= F\big[p_{G^i}(\mathbf{u}_i), p_{G^j|G^i}(\mathbf{u}_j|\mathbf{u}_i)\big] \\ &= p_{G^i}(\mathbf{u}_i)p_{G^j|G^i}(\mathbf{u}_j|\mathbf{u}_i), \end{aligned}$$

known as the *product rule*. It can also be shown, following Cox [51], that

$$\sum_{\mathbf{u}_i\in\mathbf{U}_i} p_{G^i}(\mathbf{u}_i) + \sum_{\mathbf{u}_{\neg i}\in\mathbf{U}_{\neg i}} p_{G^{\neg i}}(\mathbf{u}_{\neg i}) = 1 \quad (54)$$

for all sub-societies G^i, which is known as the *sum rule*. Finally, the non-decreasing property requires that all preference functions be non-negative, thus the preference functions must be mass functions.

We note that, regardless of the function f, so long as it is positive, continuous, monotonic, and increasing, the product and sum rules apply, and the preference functions are mass functions. Thus, we may take f to be identity without loss of generality.

REFERENCES

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

[2] J. F. Nash, "Non-cooperative games", *Ann. Of Math.*, vol. 54, pp. 289-295, 1951.

[3] R. M. Hogarth and M. W. Reder, Ed., Rational Choice. Chicago: Univ. Chicago Press, 1986.

[4] J. J. Mansbridge, Ed., Beyond Self-Interest. Chicago: Univ. of Chicago Press, 1990.

[5] A. M. Colman, "Cooperation, psychological game theory, and limitations of rationality in social interaction", *Behav. and Brain Sci.*, vol. 26, pp. 139-198, 2003.

[6] K. J. Arrow, "Rationality of self and others", in Rational Choice, R. M. Hogarth and M. W. Reder, Ed. Chicago: Univ. of Chicago Press, 1986, pp. 201-215.

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

[8] E. Sober and D. S. Wilson, Unto Others: The Evolution and Psychology of Unselfish Behavior. Cambridge, MA: Harvard Univ. Press, 1998.

[9] C. Camerer, Behavioral Game Theory: Experiments in Strategic Interaction. Princeton: Princeton Univ. Press, 2003.

[10] C. Camerer, G. Lowenstein, and M. Rabin, Ed., Advances in Behavioral Economics. Princeton: Princeton Univ. Press, 2004.

[11] J. Henrich, R. Boyd, S. Bowles, *et al.*, "'Economic Man'in Cross-culturative Perspective: Behavioral Experiments in 15 Small-scale Societies". *Behav. and Brain Sci.*, vol. 28, no. 6, pp. 795-855, 2005.

[12] T. C. Schelling, The Strategy of Conflict. Oxford: Oxford Univ. Press, 1960.
[13] L. Brown, Ed., The New Shorder Oxford English Dictionary. 4th ed. Oxford: Clarendon Press, 1993.
[14] W. C. Stirling, "Social Utility Functions, Part 1 — Theory", *IEEE Trans. on Syst., Man, and Cybern. (Part C)*, vol. 35, no. 4, pp. 522-532, 2005.
[15] W. C. Stirling and R. L. Frost, "Social Utility Functions, Part 2 — Applications", *IEEE Trans. on Syst., Man, and Cybern. (Part C)*, vol. 35, no. 4, pp. 533-543, 2005.
[16] F. P. Ramsey, "Truth and probability", in The Foundations of Mathematics and Other Logical Essays, R. B. Braithwaite, Ed. New York: The Humanities Press, 1950.
[17] B. de Finetti, "La prévision: ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*, vol. 7, pp. 1-68, 1937. Translated as "Forsight. Its Logical Laws, Its Subjective Sources", in Studies in Subjective Probability, H. E. Kyburg Jr. and H. E. Smokler, Ed. New York: Wiley, 1964, pp. 93-158.
[18] J. Kemeny, "Fair bets and inductive probabilities", *J. of Symb. Logic*, vol. 20, no. 1, pp. 263-273, 1955.
[19] A. Renyi, Foundations of Probability. San Francisco: Holden-Day, 1970.
[20] A. K. Sen, Collective Choice and Social Welfare. Amsterdam: North-Holland, 1979.
[21] G. Loewenstein, M. Bazerman, and L. Thompson, "Social utility and decision making in interpersonal contexts", *J. of Pers. and Soc. Psychol.*, vol. 57, pp. 426-441, 1989.
[22] E. Fehr and K. Schmidt, "A theory of fairness, competition, and cooperation", *Q. J. of Econ.*, vol. 114, pp. 817-868, 1999.
[23] G. Charness and M. Rabin, "Understanding social preferences with simple tests", *Q. J. of Econ.*, vol. 117, pp. 817-869, 2002.
[24] G. E. Bolton and A. Ockenfels, "A stress test of fairness measures in models of social utility", *Econ. Theory*, vol. 24, no. 4, pp. 957-982, 2005.
[25] K. R. Popper, Conjectures and Refutations: The Growth of Scientific Knowledge. New York: Harper & Row, 1963.
[26] W. James, The Will to Believe and Other Essays. New York: Dover, 1956.
[27] I. Levi, The Enterprise of Knowledge. Cambridge, MA: MIT Press, 1980.
[28] K. J. Arrow, Social Choice and Individual Values. 2nd ed. New York: John Wiley, 1963.
[29] J. Elster, Ed., The Multiple Self. Cambridge, UK: Cambridge Univ. Press, 1985.
[30] J. Harsanyi, "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility", *J. of Polit. Econ.*, vol. 63, pp. 309-321, 1955.
[31] H. Margolis, "Dual utilities and rational choice", in Beyond Self-Interest, J. J. Mansbridge, Ed. Chicago: Univ. of Chicago Press, 1990, pp. 239–253.
[32] I. Steedman and U. Krause, "Goethe's *Faust*, Arrow's Possibility Theorem and the individual decision maker", in The Multiple Self, J. Elster, Ed. Cambridge, UK: Cambridge Univ. Press, 1985, pp. 197-231.
[33] H. A. Simon, "A behavioral model of rational choice", *Q. J. of Econ.*, vol. 59, pp. 99-118, 1955.
[34] W. C. Stirling, Satisficing Games and Decision Making: With Applications to Engineering and Computer Science. Cambridge, UK: Cambridge Univ. Press, 2003.
[35] J. Pearl, Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann, 1988.
[36] S. L. Lauritzen, Graphical Models. New York: Springer Verlag, 1996.
[37] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, Probabilistic Networks and Expert Systems. New York: Springer Verlag, 1999.
[38] G. F. Cooper, "The computational complexity of probabilistic inference using belief networks", *Artif. Intell.*, vol. 42, pp. 393-405, 1990.
[39] W. Güth, R. Schmittberger, and B. Schwarze, "An experimental analysis of ultimatum bargaining", *J. of Econ. Behav. and Org.*, vol. 3, pp. 367-388, 1982.
[40] K. Binmore, A. Shaked, and J. Sutton, "Testing noncooperative bargaining theory: A preliminary study", *Amer. Econ. Rev.*, vol. 75, pp. 1178-1180, 1985.
[41] W. Güth and R. Tietz, "Ultimatum bargaining behavior: A survey and comparison of experimental results", *J. of Econ. Psych.*, vol. 11, pp. 417-449, 1990.
[42] R. H. Thaler, "The Ultimatum game", *J. of Econ. Persp.*, vol. 2, pp. 195-206, 1988.
[43] F. Zhong, D. J. Wu, and S. O. Kimbrough, "Cooperative Agent Systems: Artificial Agents Play the Ultimatum Game", *Group Decis. and Negot.*, vol. 11, pp. 433-447, 2002.
[44] A. K. Sen, "Rational fools: A critique of the behavioral foundations of economic theory", in Beyond Self-Interest, J. J. Mansbridge, Ed. Chicago: Univ. of Chicago Press, 1990.
[45] J. Gale, K. Binmore, and L Samuelson, "Learning to be imperfect: the ultimatum game", *Games and Econ. Behav.*, vol. 8, pp. 56-90, 1995.
[46] M. A. Nowak, K. M. Page, and K. Sigmund, "Fairness versus reason in the ultimatum game", *Science*, vol. 289, pp. 1773-1775, 2000.
[47] K. Sigmund, E. Fehr, and M. A. Nowak, "The economics of fair play", *Sci. Amer.*, vol. 286, pp. 83-87, 2002.
[48] F. R. Palmer, Grammar. Harmondsworth, UK: Penguin Books, 1971.
[49] N. H. Abel, "Untersuchung der Functionen zweier unabhängigen veränderlichen Grössen x und y, wie $f(x, y)$, welche die Eigenschaft haben, daß $f[z, f(x, y)]$ eine symmetrische Function von x, y, und z ist", *J. f. d. reine und angew. Math.*, vol. 1, pp. 11-15, 1826.
[50] J. Aczél, Lectures on Functional Equations and Their Applications. New York: Academic Press, 1966.
[51] R. T. Cox, "Probability, frequency, and reasonable expectation", *Amer. J. of Phys.*, vol. 14, pp. 1-13, 1946.
[52] E. T. Jaynes, Probability Theory: The Logic of Science. Cambridge, UK: Cambridge Univ. Press, 2003.

Received: March 27, 2008 Revised: April 16, 2008 Accepted: April 25, 2008