

# Satisficing, preferences, and social interaction: a new perspective

Wynn C. Stirling<sup>1</sup>  · Teppo Felin<sup>2</sup>

© Springer Science+Business Media New York 2015

**Abstract** Satisficing is a central concept in both individual and social multiagent decision making. In this paper we first extend the notion of satisficing by formally modeling the tradeoff between costs (the need to conserve resources) and decision failure. Second, we extend this notion of “neo”-satisficing into the context of social or multiagent decision making and interaction, and model the social conditioning of preferences in a satisficing framework.

**Keywords** Game theory · Satisficing · Conditional preferences · Multiple agents

## 1 Introduction

Rational decision making is among the most challenging of intellectual enterprises. The challenge is perhaps greatest in group settings where social influence exists among the individuals. In such cases, individuals may be motivated to extend concern beyond their own narrowly construed self-interest and consider the welfare of others as part of their own rational behavior. To accommodate this desire, individuals may wish to modulate their preferences as a function of the social influence that others exert on them. This flexibility is not easily accommodated, however, if individuals are constrained to express their preferences with categorical linear orderings.

We introduce two innovations that may help to mitigate this situation. First, we replace the classical linear preference ordering structure, which is designed to identify

---

✉ Wynn C. Stirling  
wynn\_stirling@byu.edu

<sup>1</sup> Electrical and Computer Engineering Department, Brigham Young University, Provo, UT 84602, USA

<sup>2</sup> Saïd Business School, University of Oxford, Oxford, UK

optimal choices, with a preference structure that is designed to identify solutions that are “good enough,” or satisficing, and which provides a natural framework for negotiation and compromise. Second, we propose the use of conditional preference orderings, which enable individuals to modulate their preferences to respond to the social influence exerted by others.

### 1.1 Satisficing concepts: individual and social

Satisficing, as introduced by Simon as a model of bounded rationality (Simon 1955, 1982), is motivated by the need to make decisions when informational or computational limitations render it impossible or impractical to apply conventional decision-theoretic formalisms. Simon’s essential idea is to test alternatives against a heuristic “aspiration.” Once an alternative is found that achieves the aspiration, it is deemed to be “good enough.”

Perhaps an equally important, but far less studied, motivation for satisficing is that being “good enough” is a natural perspective for decision makers to adopt when seeking a compromise in social decision-making situations: what is best for me may not be best for you and vice versa, but there may nonetheless be a solution that is good enough for both of us. This suggests that satisficing is a socially amenable rationale for multiagent decision making, especially in social environments that move beyond narrow self-interest and pure competition.

Simon is intentionally vague concerning the way aspirations are defined: “Aspirations are expectations—adjusted in the long run to realities—of the result that can reasonably be attained. They are not formed on the basis of detailed evaluation of alternative courses of action” (Simon 1982, p. 399). In other words, an aspiration generates a heuristic decision rule that corresponds to some fundamental notion of rational behavior. Perhaps the most widely applied concept of rational behavior is the heuristic notion that more is preferred to less, which underlies the aspiration that a solution is “good enough” if it is close enough to being optimal (i.e., it achieves the aspiration) to be acceptable. In other words, optimization is simply the mathematical formalization of the heuristic that more is preferred to less.

Another concept of rational behavior that has not received great attention in the satisficing context is the heuristic notion of getting one’s money’s worth. This perspective is fundamentally different from the perspective that underlies the aspiration to approximate optimization. Rather than comparing alternatives against an external standard, it involves an internal standard by making comparisons between the effectiveness and efficiency of each alternative. With this notion of performance, an appropriate aspiration would be for concern for the effectiveness of an alternative to dominate concern for its inefficiency. Essentially, the content of our paper is to explain how this alternative perspective can be formalized into a mathematical model that is as theoretically rigorous as is the mathematical model of optimization. Whereas Simon observed that a particular heuristic notion of what it means to be good enough can replace the corresponding mathematical formalism (optimization); we go the other way. We start with a different heuristic notion of what it means to be good enough and derive the corresponding mathematical formalism. Furthermore, we show how this formal-

ism naturally extends to multiagent decision scenarios. Our specific motivation is to develop a multiagent satisficing decision model that accommodates socially amenable concepts of behavior, including coordination, cooperation, compromise, altruism, and negotiation.

We develop our approach using the classical game theoretic framework utilized in extant scholarship in multiagent decision making. Building on the theory introduced by von Neumann and Morgenstern (1947) and further advanced by Nash (1950, 1951), noncooperative game theory is well established as a rational means of multiagent decision making in environments where each individual is motivated by narrowly construed self-interest and seeks to identify best (utility maximizing) responses to all other individuals' strategies. Under this scenario, each individual possesses a linear ordering over the outcomes. Given this utility structure, individual rationality dictates that individuals should adopt a constrained optimal solution under the assumption that other individuals behave similarly (e.g., Nash equilibrium).

Our approach, however, utilizes an entirely different utility structure. Rather than making inter-alternative comparisons (comparisons between alternatives) as is done with traditional linear preference orderings, we advocate making intra-alternative comparisons (comparisons within alternatives) by considering each alternative in terms two distinct attributes: its effectiveness in terms of achieving the objective without regard for the consumption of resources, and its efficiency in terms of conserving resources without regard for achieving the objective.

The concept of separating attributes into different utilities is not new. Other scholars (Elster 1985; Harsanyi 1955; Margolis 1990; Sen 1990) have argued that it is unwise to aggregate conflicting interests into a single preference ordering. Expanding this line of reasoning, Steedman and Krause (1985) maintain that an individual, although an indivisible unit, is nevertheless capable of considering its choices from different points of view, and that separate utilities may be defined to correspond to different 'facets' or 'selves' of an individual. Our approach is to follow (Steedman and Krause 1985) and represent each agent by two selves, which effectively become the atoms of the decision problem. We view an individual  $X$  as the composite of a selecting self, denoted  $S$ , that evaluates alternatives in terms of effectiveness without concern for efficiency, and a rejecting self, denoted  $R$ , that evaluates alternatives in terms of inefficiency without concern for effectiveness.

This approach was first introduced by Stirling (2003). However, that analysis was limited in that it did not take into account the fact that an individual's utility is a function of the actions of others, not just its own action—a significant conceptual limitation. We introduce the term neo-satisficing to distinguish this extended concept. We retain the "satisficing" terminology because our usage is consistent with the issue that motivated Simon's original usage: to identify options that are good enough. Our usage differs only in the operational definition of being good enough.

Another central contribution of this paper is that we extend satisficing, or neo-satisficing, into the mathematical structure of conditional game theory, as developed by Stirling (2012) and Stirling and Felin (2013).<sup>1</sup> When making decisions in a social

<sup>1</sup> For an in-depth review of Stirling (2012), see Ross (2014).

environment, it is natural for an individual to take into consideration the opinions of others when forming her own opinions and preferences. Deciding agents are likely to be influenced by others for a number of reasons: they may like (or dislike, for that matter) the others involved, they may value others' opinions, or they may have an existing relationship with others (e.g., familial, friendship or professional). In such situations it becomes essential to consider expressions of preference that account for interests beyond the self, narrowly conceived. Thus we develop a formal decision model for multiagent satisficing decision making that enables individuals to express their preferences as functions of the interests of those who socially influence them, thereby enabling socially amenable concepts of behavior such as coordination, cooperation, compromise, altruism, and negotiation to be explicitly modeled.

## 1.2 Related theories

To account for social relationships that exist among the members of a collective, several innovations have been applied to classical game theory. Behavioral game theory (Bolton and Ockenfels 2005; Fehr and Schmidt 1999; Sen 1990; Camerer 2003, 2004b; Camerer et al. 2004a; Henrich 2005) is a response to the need to introduce psychological realism and social influence into game theory by incorporating notions such as fairness and reciprocity into the utilities, in addition to considerations of material benefit. The closely related field of psychological game theory (Geanakoplos et al. 1989; Dufwenberg and Kirchsteiger 2004; Colman 2003; Battigalli and Dufwenberg 2009; Gilboa and Schmeidler 1988) also employs utilities that account for beliefs as well as actions and takes into consideration belief-dependent motivations such as guilt aversion, reciprocity, regret, and shame. Bacharach (1999, 2006) and Sugden (1993, 2000, 2003) have studied the concept of team reasoning, where individuals move back and forth between evaluating choices from the perspective of their individual utilities and from the perspective of group-level utility. Regardless of the issues used to define the preferences, however, these approaches to multiagent decision making differ from ours in that they use linear preference orderings over the outcomes and, therefore, the solution concepts used by classical game theory continue to apply.

Satisficing has largely received attention at the level of individual decision making, although a few approaches have extended the notion to the multiagent case. Pazgal (1997) studies mutual interest games and establishes conditions under which satisficing solutions can lead to Pareto dominant cooperative outcomes. Bendor et al. (2009)) address the multiagent case by designing aspiration criteria to distinguish between “good” and “bad” action profiles, and establish conditions for the existence of a set of “good” profiles that contain a Pareto optimal solution. Both of these approaches are significantly different from ours. They are fundamentally heuristic and are based on the standard Simonian notion that “good enough” is an approximation to optimality.

We note that our approach has some links with extant work on multiattribute decision making (Dyer and Sarin 1979; Keeney and Raiffa 1976; Lam et al. 2013). For example, in this domain Busemyer and Townsend (1993) have emphasized how trade-offs and uncertainty impact preferences and emergent choices. More recently Bhatia (2013) has studied how highly accessible attributes in a multi-alternative setting impact

decision making over time. Though our work features some natural links to this work on multiattribute decision making, there are also some differences. Specifically, multiattribute decision theory involves making value tradeoffs between the achievement of one objective at the expense of another objective. A common way for a single decision maker to approach this problem is to view each objective as a distinct agent with a single attribute. The agents must collectively choose an alternative that achieves Pareto efficiency by each making its own evaluation over the set of alternatives. However, there are no direct comparisons made among the agents. With a neo-satisficing multiagent decision making scenario, each agent is separated into two selves and intra-agent comparisons are made for each alternative. Those alternatives that survive the comparison test are deemed to be satisficing. However, there are no direct comparisons made among the alternatives. Thus, although the two concepts have conceptual similarities, they also have significant differences, and applying our concept of satisficing to multiattribute decision making is not the focus of this paper.

The decision-making concept that is perhaps closest to our neo-satisficing approach is some form of cost-benefit analysis (CBA), which might be used to, say, justify investing funds for projects under consideration by government agencies or private organizations. However, there are some important differences between CBA and our neo-satisficing approach. First, CBA requires a common unit of measurement, which requires all human economic, moral, health, and other such issues to be expressed in terms of a common transferable medium, such as money. Second, even though multiple individuals may be involved in the analysis, they are not autonomous decision makers with their own alternative sets. Rather, there is one alternative set—the collective must act as a unit to make a single choice. Consequently, the CBA decision problem is a social choice problem rather than a game in the usual sense. Third, CBA is, at the end of the day, a heuristic-based form of analysis. There is no implication that a CBA decision is optimal in any sense. In fact, it may not comply with Pareto optimality since the welfare of some participants may be improved while the welfare of others may be reduced (see [Kaldor 1939](#); [Hicks 1939](#)). It can also be shown that attempts to remedy this problem via side payments can actually reverse the decision (see [Scitovsky 1941](#)).

## 2 An alternate perspective

When making performance evaluations, there are three distinct degrees of comparison: superlative, comparative, and positive. By definition, optimization-based decision theory is of the superlative degree: only the best is satisfactory. Decisions based on heuristics are of the positive degree: the choice meets a criterion that defines what it means to be ‘good.’ Arguing that a choice is ‘good enough,’ according to Simonian satisficing, is a coarsening of the superlative degree on the one hand (settling for being approximately best), and a refinement of the positive degree on the other (good enough, not just good). What might be of real interest, however, is a non-heuristic definition of performance that is distinct from the superlative and positive degrees and qualifies as a true instantiation of the comparative degree. We might ask the question: Does there exist a means of evaluation that does not derive from either the superlative or the positive notions of quality? If so, then it must be formulated in a way that avoids

linear preference orderings over alternatives and is not based on heuristics. It must have a mathematical foundation that is as rigorous (albeit different) as the mathematical foundation of optimization. Identifying such a methodology requires a perspective different from the traditional optimization-based point of view.

Generally, the interests of all agents in a multiagent decision problem will not coincide, and a compromise must be achieved. When seeking a compromise, attention shifts, at least to some degree, from insisting on a solution that maximizes success to settling for a solution that at least avoids failure. Indeed, even from Simon's perspective on satisficing, the aspiration level can be viewed as establishing minimum requirements, which is tantamount to establishing criteria for avoiding failure.

Addressing a decision problem from the perspective of avoiding failure provides flexibility that is difficult to achieve under the perspective of optimization. In particular, it allows us to distinguish between what (Unger 1975, p. 55) calls absolute (limit) and relative (degree) terms. "Semantically, we may say that our absolute terms indicate, or purport to denote, an absolute limit. This limit is approached to the extent that the relevant relative property or properties are absent in the thing to which one might sensibly apply the absolute term, or its correlative relatives [emphasis in original]." Superlative concepts such as "best" and "worst" are clearly absolute terms. Terms such as "success" and "failure," on the other hand, can be seen as degree terms, and admit different levels of compliance. If one does not achieve the best possible result, then one fails, at least from the perspective of optimization. But, the degree of failure—if the chosen solution were second best—is not as severe as it would be if the chosen solution were worst. Thus, if we wish to view a decision problem from the failure avoidance perspective, we would be wise to avoid using absolute terms.

The practical perspective of avoiding failure is analogous to the epistemological perspective of avoiding error, and it may be instructive to learn from epistemologists who have addressed this issue. In particular, the views of William James and Isaac Levi are relevant. James (1956) argues that the quest for truth and the avoidance of error are fundamentally different ways of seeking knowledge.<sup>2</sup> Levi (1980) builds James's view and develops a pragmatic approach to knowledge acquisition as a decision-theoretic problem. He proposes two criteria with which to characterize propositions under consideration: the credal probability that quantifies the strength of belief that the proposition is true, and an information determining probability that quantifies the informational value that one would gain if the proposition were rejected, thereby permitting one to focus on propositions that are of more intrinsic value.<sup>3</sup> These two criteria

<sup>2</sup> "There are two ways of looking at our duty in the matter of opinion ... 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. ... 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" (James 1956, pp. 17,18). Landau and Chisholm (1995) argue that this perspective is also relevant to organizational management: "The difference is not simply semantic, it is that which distinguishes a success-oriented form a failure-avoidance management [emphasis in original]."

<sup>3</sup> Levi's use of the idea of "probability" here is in the sense of a propensity; that is, the tendency or disposition to yield an outcome of a certain kind.

**Table 1** Optimization vis à vis satisficing

	Epistemological context	Practical context
Optimization	Demand the truth and nothing but the truth	Demand the best and nothing but the best
Neo-satisficing	Acquire information while avoiding error	Conserve resources while avoiding failure

provide quantitative measures of the relative importance of truth and informational value.<sup>4</sup> Those propositions for which the informational value of rejecting them dominates their credal value are eliminated from consideration, thereby permitting one to focus interest on those propositions that are either likely to be true (an inductive valuation) or which provide significant information (an abductive valuation). Levi replaces the conventional superlative epistemological mantra of “seeking the truth and only the truth” by the more flexible and conservative notion of “acquiring useful information while avoiding error.” Thus, rather than obstinately pursuing the bold, and perhaps temerarious demand for the truth and only the truth in the absence of compelling information, Levi’s approach offers a more cautious and prudent approach when there is insufficient evidence to definitively ascertain which of a set of propositions is the true one. One can at least eliminate some (and perhaps many) of them from serious consideration. But if the evidence is sufficiently compelling to reject all but one alternative, then one has, by default, identified the true proposition. In this way, Levi’s mechanism conforms to the perspective of favoring the avoidance of error over an intransigent demand for truth.

Our neo-satisficing approach builds on Levi’s methodology by exploiting the analogy between the epistemological notion of avoiding error and the practical notion of avoiding failure. To make the analogy complete, however, we require a practical equivalent to the epistemological concept of informational value. In the epistemological context, information may be viewed as a resource that is consumed by not rejecting it. Accordingly, a natural practical equivalent is the consumption of practical resources, such as money, energy, time, exposure to hazard, the expenditure of economic, political, social, or moral capital, or any other quantities, material or otherwise, that are expended by the agents in the pursuit of their objectives. Rational behavior is thus expressed in terms of both effectiveness (i.e., avoiding failure) and efficiency (i.e., conserving resources). Much of our contribution lies in migrating Levi’s theory from the epistemological domain to the practical domain. We also extend this concept to the multiple agent case (which Levi does not do). Table 1 summarizes the relationships between epistemological and practical concepts.

<sup>4</sup> The importance of interest relative to truth is emphasized by both Whitehead and Popper. “It is more important that a proposition be interesting than that it be true. This statement is almost a tautology. For the energy of operation of a proposition in an occasion of experience is its interest, and is its importance” (Whitehead 1937, Part IV, Chapter XVI). “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 [emphasis in original]” (Popper 1963, p. 229).

### 3 The neo-satisficing model

#### 3.1 Mathematical formulation

We now formalize our notion of neo-satisficing. We expand the notion of utility from a mapping defined over a finite set  $\mathcal{A}$  (the set of alternatives) of points, to a mapping over a set of sets—a Boolean algebra of subsets of the alternative set. Our goal is to identify a set of alternatives such that each is an acceptable solution. In multiagent settings, this set may then be subjected to further analysis and negotiation to arrive at a compromise. In the interest of clarity, however, we first develop the neo-satisficing model for single agents and then extend to multiagent scenarios.

For a single agent  $X$ , let  $\mathcal{F}$  denote a Boolean algebra of subsets of  $\mathcal{A}$  and let  $U_S$  denote a failure avoidance (i.e., success achieving) utility such that, for any  $B \in \mathcal{F}$ ,  $U_S(B)$  quantifies the degree to which focusing on the set  $B$  avoids failure.<sup>5</sup> Also, let  $U_R$  denote a resource consumption utility such that, for any  $B \in \mathcal{F}$ ,  $U_R(B)$  quantifies the degree to which focusing on  $B$  consumes resources.

An optimization-based approach would necessarily focus on singleton results, but a failure avoiding approach would relax that constraint and seek to identify alternatives that have good failure avoidance characteristics and low resource consuming characteristics. To proceed, we must define the structure of these utilities. We propose the following attributes of these functions:

- Measures of both failure avoidance and resource consumption should be nonnegative and finite. Without loss of generality, we may restrict these utilities to range over the unit interval, that is,  $U_S: \mathcal{F} \rightarrow [0, 1]$  and  $U_R: \mathcal{F} \rightarrow [0, 1]$ .
- Both  $U_S$  and  $U_R$  should be monotonic; that is, if  $B_1 \subset B_2$ , then  $U_S(B_1) \leq U_S(B_2)$  and  $U_R(B_1) \leq U_R(B_2)$ . This condition means that the larger the set, the greater the degree of both failure avoidance and resource consumption.
- $U_S(\emptyset) = U_R(\emptyset) = 0$ . This condition establishes that the empty set can neither avoid failure nor consume resources.
- $U_S(\mathcal{A}) = U_R(\mathcal{A}) = 1$ . This condition ensures that the entire alternative set completely avoids failure and also consumes all resources.
- The incremental value, either in terms of failure avoidance or resource consumption, of any two disjoint sets should be additive. That is, if  $B_1 \cap B_2 = \emptyset$ , then the failure avoiding and resource consuming values of their union is equal to the sum of their respective individual failure avoiding and resource consuming values; that is,  $U_S(B_1 \cup B_2) = U_S(B_1) + U_S(B_2)$  and  $U_R(B_1 \cup B_2) = U_R(B_1) + U_R(B_2)$ .

These conditions imply that  $U_R$  and  $U_S$  possess the syntactical structure of probability measures over  $\mathcal{F}$ , although they do not possess the traditional semantic interpretations of probability.<sup>6</sup> Instead, they are used to quantify resource consuming

<sup>5</sup> In the epistemological domain, the probabilistic analogue to this concept is that  $\text{Prob}(B)$  quantifies the degree to which focusing on  $B$  avoids error.

<sup>6</sup> The use of the probability syntax for utilities is not new. Berhold (1973) and Castagnoli and LiCalzi (1996) have interpreted normalized utility functions as probability mass functions, and Abbas and Howard (2005) and Abbas (2009) have applied the probability syntax to utilities for the study of multi-attribute decision problems.

and failure avoidance attributes. The larger the set  $B$ , the more failure is avoided but, also, the more resources are consumed. Thus, our approach is to choose a set  $B \in \mathcal{F}$  that achieves a balance between avoiding failure and consuming resources. For any set  $B$ ,  $U_R(B)$  is the amount of resource consumed by focusing on  $B$ .  $U_R(B^c) = 1 - U_R(B)$  is the amount of resource that is conserved by focusing on  $B$ . We define the utility of simultaneously avoiding failure and conserving resources by retaining  $B$  as the convex combination of  $U_R(B)$  and  $U_R(B^c)$ , namely

$$\Psi(B) = \alpha U_S(B) + (1 - \alpha) U_R(B^c), \quad (1)$$

where  $0 \leq \alpha \leq 1$ . The parameter  $\alpha$  represents the relative importance of avoiding failure versus conserving resources. Setting  $\alpha = 1$  places a premium on avoiding failure; setting  $\alpha = \frac{1}{2}$  places equal weight on conserving resources and avoiding failure; and setting  $\alpha < \frac{1}{2}$  emphasizes conservation of resources over avoiding failure.

Since utilities are invariant to scale and zero level, we may transform  $\Psi$  by a positive affine transformation of the form

$$\begin{aligned} \Phi(B) &= \frac{1}{\alpha} \Psi(B) - \frac{1 - \alpha}{\alpha} \\ &= U_S(B) - q U_R(B), \end{aligned} \quad (2)$$

where  $q = \frac{1 - \alpha}{\alpha}$ . The function  $\Phi$  is the satisficing utility. If we assume that  $\mathcal{F}$  contains all singleton sets, then for any  $\{a\} \in \mathcal{F}$ ,  $\Phi(\{a\}) = u_S(a) - q u_R(a)$ , where  $u_S(a) = U_S(\{a\})$  and  $u_R(a) = U_R(\{a\})$  are mass functions, that is,  $u_S(a) \geq 0$  and  $u_R(a) \geq 0$  for all  $a \in \mathcal{A}$ , and  $\sum_{a \in \mathcal{A}} u_S(a) = \sum_{a \in \mathcal{A}} u_R(a) = 1$ . Thus,

$$\Phi(B) = \sum_{a \in B} [u_S(a) - q u_R(a)]. \quad (3)$$

The set that maximizes this utility comprises all of the alternatives for which the utility of avoiding failure dominates the inutility of consuming resources (as scaled by  $q$ ). This is the satisficing set, defined as

$$\mathcal{S} = \{a \in \mathcal{A} : u_S(a) \geq q u_R(a)\}. \quad (4)$$

We interpret  $q$  as a boldness index. As  $q \rightarrow 0$  (i.e.,  $\alpha \rightarrow 1$ ), the penalty for consuming resources decreases, and  $X$  considers more alternatives as being acceptable. The larger  $q$  becomes, however, the more alternatives are rejected on the basis of inefficiency, and  $X$  becomes more willing to accept failure in the interest of conserving resources. Setting  $q = 1$  establishes an equal balance between avoiding failure and conserving resources. The following theorem establishes sufficient conditions for the existence of a satisficing solution.

**Theorem 1**  $q \leq 1$  implies  $\mathcal{S} \neq \emptyset$ .

*Proof* If  $\mathcal{S} = \emptyset$ , then  $u_S(a) < q u_R(a)$  for all  $a \in \mathcal{A}$ , in which case,

**Table 2** Epistemology-practical analogical concepts

Epistemological context	Practical context
Belief	Effectiveness
Informative	Efficient
Credal probability	Selectability
Information determining probability	Rejectability

$$1 = \sum_a u_S(a) < q \sum_a u_R(a) = q, \quad (5)$$

resulting in a contradiction. Thus,  $\mathcal{S}$  must contain at least one element.  $\square$

The satisficing set is analogous to what [Levi \(1980\)](#) terms the set of serious possibilities. Essentially, we have migrated his concept of epistemic utility into a concept of practical utility by introducing practical analogies to the epistemological concepts. Just as a credal probability mass function is a measure of the degree of belief regarding a proposition without regard for its informational value, we shall term  $u_S$  a selectability mass function as a measure of the degree to which an alternative achieves the objective without regard for the consumption of resources. And, just as an information determining probability of a proposition is a measure of the degree to which information is conserved by rejecting it,  $u_R$  is termed a rejectability mass function as a measure of the degree to which rejecting an alternative conserves resources without regard for achieving the objective. [Table 2](#) summarizes the relationships between epistemological and practical concepts.

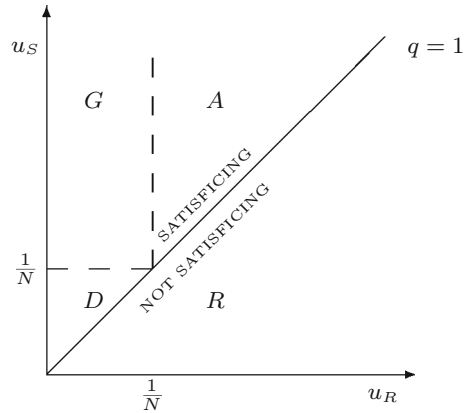
### 3.2 Quantitative evaluations

Expressing evaluations according to selectability and rejectability permits qualitative analysis in addition to quantitative evaluations. Let  $N$  denote the cardinality of  $\mathcal{A}$  and assume  $q = 1$ . If  $u_R(a) = \frac{1}{N}$  ( $u_S(a) = \frac{1}{N}$ ) for all  $a \in \mathcal{A}$ , then the rejecting (selecting) self is in a state of complete preference neutrality regarding efficiency (effectiveness). Qualitatively, if  $u_R(a) < \frac{1}{N}$  then  $a$  is deemed to be efficient, and if  $u_S(a) > \frac{1}{N}$ , then  $a$  is deemed to be effective. This suggests that we can partition outcomes into four general qualitative categories or dispositional modes:

- Gratification:  $u_S(a) > \frac{1}{N}$  and  $u_R(a) < \frac{1}{N}$  (effective and efficient)
- Ambivalence:  $u_S(a) > u_R(a) > \frac{1}{N}$  (effective and inefficient)
- Dubiety:  $u_R(a) < u_S(a) < \frac{1}{N}$  (ineffective and efficient)
- Repulsion:  $u_S(a) < u_R(a)$  (not satisficing)

These dispositional mode regions are illustrated in [Fig. 1](#). Gratification and repulsion are qualitative modes of contentment, while dubiety and ambivalence are modes of conflict. These qualitative categories provide additional insight that is difficult to ascertain with optimization-based approaches. Under the optimization paradigm, although the decision maker may achieve the best possible outcome (whatever comfort that

**Fig. 1** Dispositional regions:  $G$  = gratification,  $A$  = ambivalence,  $D$  = dubiety,  $R$  = repulsion



knowledge affords), it may only be making the best of a bad situation. The qualitative assessments available with neo-satisficing, on the other hand, can alert the decision maker of limitations regarding its ability to cope adequately with the decision problem and may trigger a re-evaluation of the scenario. For example, the agent may be motivated to seek assistance or advice if the dispositional mode is conflictive.

#### 4 Neo-satisficing in a social context

Thus far we have provided a non-heuristic concept of satisficing for a single agent only, and we have not yet addressed another central issue that motivates our approach, namely the need to accommodate social influence relationships that may exist among the individuals. Furthermore, since the satisficing set will generally not be a singleton set, we must address the issue of how to choose a unique element from this set as the one to be implemented.

Let  $\mathcal{X} = \{X_1, \dots, X_n\}$  be a group of autonomous decision-making agents, and let  $\mathcal{A}_i$  denote  $X_i$ 's alternative set,  $i = 1, \dots, n$ . The joint alternative set is the  $n$ -fold Cartesian product set  $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ . The elements of this set, which comprise vectors of the form  $\mathbf{a} = (a_1, \dots, a_n)$  with  $a_i \in \mathcal{A}_i$ , are called profiles. In terms of our satisficing development, we may express each  $X_i$  as the composite of a selecting self, denoted  $S_i$ , and a rejecting self, denoted  $R_i$ . We denote the selecting group as  $\mathcal{S} = \{S_1, \dots, S_n\}$  and the rejecting group as  $\mathcal{R} = \{R_1, \dots, R_n\}$ . We may also associate profiles with each of these groups. A selecting profile  $\mathbf{s} = (s_1, \dots, s_n) \in \mathcal{A}$  is a profile considered in terms of effectiveness, and a rejecting profile  $\mathbf{r} = (r_1, \dots, r_n) \in \mathcal{A}$  is a profile considered in terms of inefficiency.

##### 4.1 Categorical satisficing games

Classical game theory assumes that each individual comes to a multiagent decision-making encounter with ex ante preferences that are fixed and immutable; they are

categorical. They represent the individual's evaluation after taking into consideration all factors that could serve to benefit the individual.

Expressing a multiagent decision problem in terms of selecting and rejecting selves offers a mechanism for implementing our goal of facilitating compromise and negotiation. In keeping with the classical approach of ascribing a utility to each agent, we may ascribe a utility to each self. Suppose that each  $S_i$  possesses a selectability utility mass function,  $u_{S_i}: \mathcal{A} \rightarrow [0, 1]$ , that evaluates profiles in terms of achieving  $S_i$ 's objective of effective behavior and that each  $R_i$  possesses a rejectability utility mass function,  $u_{R_i}: \mathcal{A} \rightarrow [0, 1]$ , that evaluates profiles in terms of achieving  $R_i$ 's objective of conserving resources. These ex ante utilities are also categorical, since they represent the fixed opinions of the selecting and rejecting selves.

**Definition 1** A categorical satisficing game for a collective  $\mathcal{X}$  comprises the sets  $\mathcal{S}$  and  $\mathcal{R}$  of selecting and rejecting selves, respectively; the product alternative set  $\mathcal{A}$ ; and the set of categorical utilities  $\{u_{S_i}, u_{R_i}, i = 1, \dots, n\}$ .

An obvious way to solve a categorical satisficing game is to apply the satisficing concept by computing the set of individually satisficing profiles, yielding

$$\mathcal{S}_i = \{\mathbf{a} \in \mathcal{A}: u_{S_i}(\mathbf{a}) \geq q_i u_{R_i}(\mathbf{a})\}, \quad (6)$$

$i = 1, \dots, n$ , where  $q_i$  is  $X_i$ 's individual  $q$ -value.

**Definition 2** Given a collection  $\{\mathcal{S}_1, \dots, \mathcal{S}_n\}$  of individually satisficing sets, the categorical compromise is the intersection

$$\mathcal{C}_{\mathcal{X}} = \mathcal{S}_1 \cap \dots \cap \mathcal{S}_n. \quad (7)$$

The categorical compromise the set of profiles that are simultaneously satisficing for all individuals. If  $\mathcal{C}_{\mathcal{X}}$  contains multiple profiles, then ties can be broken by standard techniques. If  $\mathcal{C}_{\mathcal{X}} = \emptyset$ , then, in an attempt to reach a compromise, the agents may decrement their  $q$ -values, thus allowing more alternatives to be satisficing. If enough agents are willing to incrementally reduce their satisficing standards in this way, a compromise will eventually be reached. If  $\mathcal{C}_{\mathcal{X}}$  remains empty after all individuals have reached their decrementing limits, then a compromise cannot be reached.

**Example 1** (Tom and Martha) Suppose Tom and Martha are planning a party. Tom is in charge of selecting the venue, either the beach ( $b$ ) or the country club ( $c$ ). Thus,  $\mathcal{A}_T = \{b, c\}$ . Suppose that Martha is in charge of selecting the menu, either hamburger ( $h$ ) or lasagna ( $l$ ), yielding  $\mathcal{A}_M = \{h, l\}$ . Let Tom's and Martha's selecting and rejecting selves be denoted  $S_T, S_M, R_T$ , and  $R_M$ , respectively. The joint alternative set is  $\mathcal{A} = \mathcal{A}_T \times \mathcal{A}_M = \{(b, h), (b, l), (c, h), (c, l)\}$ . Natural operational definitions for the selecting and rejecting utilities would be the enjoyment of the planners and the cost of the party, respectively. Numerical values for these utilities are displayed in Table 3. Setting boldness values at  $q_T = q_M = 1$  and applying (6), the satisficing set for Tom is  $\mathcal{S}_T = \{(b, h), (c, h)\}$  and the satisficing set for Martha is  $\mathcal{S}_M = \{(b, l), (c, l)\}$ . At these boldness values,  $\mathcal{C}_{TM} = \emptyset$ , but a compromise could be achieved if, say, Tom were to reduce  $q_T$  to  $\frac{2}{3}$ , yielding  $\mathcal{C}_{TM} = \{(b, l)\}$ .

**Table 3** Categorical selecting and rejecting utilities for Tom and Martha

	$(b, h)$	$(b, l)$	$(c, h)$	$(c, l)$
$u_{S_T}$	0.4	0.2	0.2	0.2
$u_{R_T}$	0.1	0.3	0.2	0.4
$u_{S_M}$	0.0	0.5	0.0	0.5
$u_{R_M}$	0.1	0.3	0.2	0.4

**4.2 Conditional satisficing games**

As is the case with traditional game theory, the satisficing model developed with (6) and (7) is based on the concept of narrowly construed self-interest, where each individual is concerned for its own welfare, regardless of the possible effect on others. This model of rational behavior may be appropriate in an environment of pure competition where the only concern is for one's material benefit. But as (Arrow 1986, p. 203) has noted, "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."

In social environments, expanded notions of rational behavior may include social objectives such as cooperation, compromise, and altruism, or even such antisocial objectives as conflict, avarice, meanness, and manipulation. A social objective, however, cannot be easily expressed directly as a material payoff or other tangible or easily quantifiable reward. Rather, it is more naturally expressed by individuals evaluating their material benefit in terms of the social influence that others exert on them. An individual whose concept of rational behavior includes achieving social objectives as well as material objectives may do so by dynamically modulating their material objectives in response to social influence. Our approach is to expand the utility structure to accommodate an expanded notion of self-interest, where individuals incorporate the interests of others into their own rationality.

A natural way for one to respond to the interests of others is to structure one's preferences as hypothetical propositions of the form if  $p$  then  $q$ , where the antecedent  $p$  is a hypothetical conjecture and the consequent  $q$  is the response. We demonstrate this structure with the following example: Suppose the preferences of  $S_1$  are influenced by the interests of  $R_2$ , that is,  $X_2$ 's concern for conserving resources exerts influence on  $X_1$ 's interest in effectiveness. Then  $S_1$  could respond to this influence by defining a conditional utility of the form  $u_{S_1|R_2}(s_1|r_2)$ , where the profile  $r_2$  on the right side of the conditioning symbol " $|$ " is the hypothesis that  $R_2$  views that profile as the one that should or will be chosen by the group. This is the antecedent. The profile  $s_1$  on the left side of the conditioning symbol is the profile under consideration by  $S_1$ , and the consequent  $u_{S_1|R_2}(s_1|r_2)$  is  $S_1$ 's resultant utility of  $s_1$ , given that  $R_2$  hypothesizes  $r_2$ .  $S_1$  can account for  $R_2$ 's influence by defining its response for each  $s_1 \in \mathcal{A}$  for all  $r_2 \in \mathcal{A}$ . We now formalize this concept.

Graph theory provides a powerful and expressive way to represent the relationships that exist among a collective of individuals. A graph comprises a pair  $(\mathcal{Z}, E)$ , where  $\mathcal{Z}$  is a finite set of vertices (individuals) and  $E$  is a set of edges (influence relationships) linking pairs of vertices. If  $Z_j$  is directly influenced by  $Z_i$  but  $Z_j$  does not directly influence  $Z_i$ , then there is a directed edge, denoted “ $\rightarrow$ ” from  $Z_i$  to  $Z_j$ . A path from  $Z_i$  to  $Z_j$  is a sequence of vertices  $\{Z_i, Z_{k_1}, Z_{k_2}, \dots, Z_j\}$  such that  $Z_i \rightarrow Z_{k_1} \rightarrow Z_{k_2} \rightarrow \dots \rightarrow Z_j$ . We write  $Z_i \mapsto Z_j$  if there is a path from  $Z_i$  to  $Z_j$ . If there are no paths such that  $Z_i \mapsto Z_i$  for any  $i$ , the graph is said to be acyclic; if, in addition, all edges are directed,  $(\mathcal{Z}, E)$  is a directed acyclic graph, or DAG. The set of all vertices that influence  $Z_i$  is called the parent set of  $Z_i$ , denoted  $\text{pa}(Z_i)$ . If a vertex has no parents, it is called a root vertex. A DAG of course must necessarily have at least one root vertex.

**Definition 3** Let  $m = 2n$ , define

$$\mathcal{Z} = \{Z_1, \dots, Z_m\} = \{S_1, \dots, S_n, R_1, \dots, R_n\}, \quad (8)$$

and let  $\text{pa}(Z_i) = \{Z_{i_1}, \dots, Z_{i_{p_i}}\} \subset \mathcal{Z}$  denote the subset of  $\mathcal{Z}$  consisting of the  $q_i$  selves that exert social influence on  $Z_i$ . A profile  $\mathbf{z}_i \in \mathcal{A}$  is a conjecture for  $Z_i$  if  $\mathbf{z}_i$  is the profile that is under consideration for  $Z_i$  as the one that should or will be implemented. A joint conjecture is a set of profiles  $(\mathbf{z}_1, \dots, \mathbf{z}_m)$ , where  $\mathbf{z}_i$  is a conjecture for  $Z_i$ .

A conditioning conjecture for  $\text{pa}(Z_i)$  is a set of profiles  $\boldsymbol{\zeta}_i = (\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_{q_i}})$ , where  $\mathbf{z}_{i_k}$  is a conjecture for  $Z_{i_k}, k = 1, \dots, q_i$ . A conditional utility, denoted  $u_{Z_i|\text{pa}(Z)}(\mathbf{z}_i|\boldsymbol{\zeta}_i)$ , is the consequent of a hypothetical proposition whose antecedent is the conditioning conjecture  $\boldsymbol{\zeta}_i$  for  $\text{pa}(Z_i)$ . If  $q_i = 0$ , then  $Z_i$  is not socially influenced and possesses a categorical utility, denoted  $u_{Z_i}$ .

Conditional utilities are mass functions, that is,  $u_{Z_i|\text{pa}(Z)}(\mathbf{z}_i|\boldsymbol{\zeta}_i) \geq 0$  for all  $\mathbf{z}_i \in \mathcal{A}$  and  $\sum_{\mathbf{z}_i \in \mathcal{A}} u_{Z_i|\text{pa}(Z)}(\mathbf{z}_i|\boldsymbol{\zeta}_i) = 1$  for all  $\boldsymbol{\zeta}_i \in \mathcal{A}^{p_i} = \mathcal{A} \times \dots \times \mathcal{A}$  ( $p_i$  times). A conditional utility quantifies the amount selectability utility mass (if  $Z_i \in \mathcal{S}$ ) or rejectability utility mass (if  $Z_i \in \mathcal{R}$ ) that  $Z_i$  ascribes to  $\mathbf{z}_i$ , given the hypothesis (i.e., conditioning conjecture) that each  $Z_{i_k} \in \text{pa}(Z_i)$  ascribes its entire unit of utility mass to  $\mathbf{z}_{i_k}$ . Specifically, let

$$\boldsymbol{\sigma}_i = (\mathbf{s}_{i_1}, \dots, \mathbf{s}_{i_{p_i}}, \mathbf{r}_{k_1}, \dots, \mathbf{r}_{k_{r_i}}) \quad (9)$$

denote the conditioning conjecture for

$$\text{pa}(S_i) = \{S_{i_1}, \dots, S_{i_{p_i}}, R_{k_1}, \dots, R_{k_{r_i}}\},$$

and let

$$\boldsymbol{\rho}_j = (\mathbf{s}_{j_1}, \dots, \mathbf{s}_{j_{l_j}}, \mathbf{r}_{l_1}, \dots, \mathbf{r}_{l_{v_j}}) \quad (10)$$

denote the conditioning conjecture for

$$\text{pa}(R_j) = \{S_{j_1}, \dots, S_{j_{l_j}}, R_{l_1}, \dots, R_{l_{v_j}}\}.$$

The utility mass functions  $u_{S_i|\text{pa}(S_i)}$  and  $u_{R_j|\text{pa}(R_j)}$  are conditional selectability utility and conditional rejectability utility mass functions for  $S_i$  and  $R_j$ , respectively. If

pa  $(S_i) = \emptyset$ , then  $u_{S_i|pa(S_i)} = u_{S_i}$ , a categorical utility. Similarly, if pa  $(R_i) = \emptyset$ , then  $u_{R_i|pa(R_i)} = u_{R_i}$ .

**Definition 4** A conditional satisficing game for a collective  $\mathcal{X}$  comprises the sets  $\mathcal{S}$  and  $\mathcal{R}$  of selecting and rejecting selves, respectively; the product alternative set  $\mathcal{A}$ ; and the collections  $\{u_{S_i|pa(S_i)}, u_{R_i|pa(R_i)}, i = 1, \dots, n\}$  of conditional utilities. If the influence relationships among the elements of  $\mathcal{S}$  and  $\mathcal{R}$  contain no cycles, then the collective may be represented by a DAG where the vertices correspond to the selecting and rejecting selves and the edges correspond to the conditional utilities. Such a collective is a hierarchical satisficing game.

It will sometimes be convenient to express a satisficing game using the alternate notation  $\mathcal{Z}$  and the collections  $\{u_{Z_i|pa(Z_i)}, i = 1, \dots, m\}$  as introduced in Definition 3.

### 4.3 Satisficing solution concepts

The introduction of conditional utilities offers the possibility of developing new solution concepts for satisficing games. Certainly one could revert to traditional game-theoretic concepts such as equilibrium-based solutions, but doing so would not comport with our objective of establishing an approach that favors compromise and negotiation over constrained optimization. A concept that is compatible with this objective is the creation of a social model for the group as a function of the social relationships as expressed with the conditional utilities.

In the course of a social engagement, the individual ex ante conditional preferences will propagate through the collective. As the individuals interact, social relationships will be formed, resulting in a complex social structure. The result is a comprehensive social model that accounts for all of the relationships that exist among the individuals. For a group whose conditional preferences dispose them to cooperate, the interests of the individuals will tend to align in a positive, or reinforcing, way. For a group whose conditional preferences dispose them to conflict, such as would be the case for athletic contests or military engagements, the alignment of interests would be negative, that is, in opposition. In either case, such a social model would provide a comprehensive description of the propensity of the collective to coordinate.

Although sometimes used synonymously in common vernacular, coordination and cooperation are distinct concepts. Individuals may cooperate (or conflict) to pursue their own narrow self-interest without the existence of a group-level objective (e.g., the Prisoner's Dilemma), but coordination involves more than narrowly construed self-interest. According to the Oxford English Dictionary, "[to coordinate] is to place or arrange (things) in proper position relative to each other and to the system of which they form parts; to bring into proper combined order as parts of a whole" (Murray et al. 1991). A collective coordinates to the extent that the interests of the individuals correspond, or conform, to some degree of systematic group-level behavior, either positively or negatively. At one extreme, all of the participants would have the same most preferred profile (which may be conflictive), and the coordination of that set of profiles would be high. At another extreme, each individual would have a different most preferred profile, and the coordination of that set of profiles would be low. Thus,

it would be desirable to combine, or aggregate, the conditional utilities to produce a coordination ordering over all possible joint conjectures for each individual. Such an aggregation would provide a quantitative measure of the degree of alignment of individual interests and would enable the assessment of such social characteristics as the seriousness of disputes and the possibilities of compromise.

Aggregation is a central concept of social choice theory, where the individual preferences over a single set of alternatives are combined to form a choice function for the group. The concept of aggregation is less relevant to conventional game theory, however, since in that context each individual is endowed with its own set of alternatives from which it may autonomously choose, and the notion of group choice or group preference is generally eschewed: "It may be meaningful, in a given setting, to say that a group 'chooses' or 'decides' something. It is rather less likely to be meaningful to say that the group 'wants' or 'prefers' something" (Shubik 1982, p. 124). This sentiment may be appropriate when narrow self-interest dominates individual behavior and pure competition is the operative social climate. But in more general settings, where individuals have social as well as material interests, systematic group-level behavior may emerge as socially motivated individuals interact, thereby creating a social climate that enables cooperation, compromise, altruism, and negotiation to occur. Furthermore, the concept of aggregation that we advocate is not merely to define a choice function. Rather, our concept is to generate an emergent notion of group preference over the set of all joint conjectures as a result of social influence expressed in terms of the degree of alignment that exists among their individual conjectures.

**Definition 5** The preferences of a collective  $\mathcal{X}$ , expressed in terms of a collective of selecting and rejecting selves  $\{S_1, \dots, S_n, R_1, \dots, R_n\}$ , is endogenously aggregated if there exists a function  $\mathcal{U}_{\mathcal{X}}: \mathcal{A}^m \rightarrow \mathbb{R}$  (recall that  $m = 2n$ ), termed a coordination utility, of the form

$$\begin{aligned} \mathcal{U}_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n) \\ = f[u_{S_i | \text{pa}(S_i)}(s_i | \sigma_i), u_{R_j | \text{pa}(R_j)}(r_j | \rho_j), i, j \in \{1, \dots, n\}], \end{aligned} \quad (11)$$

where  $f$  is yet to be defined.

Endogeneity precludes the possibility that the aggregation structure is imposed from outside the group. Rather, it must emerge as a consequence of the influence relationships that exist among its members.

#### 4.3.1 Social coherence

The purpose of the coordination utility is to provide an ordering over all combinations of joint selecting-rejecting conjectures. For such an ordering to be meaningful, it must comply with what is perhaps the only universally accepted notion of group preference, the Pareto principle: If every member of a group views an alternative as best for it, then that alternative is best for the group.

Suppose each member of  $\mathcal{Z} = \{Z_1, \dots, Z_m\}$  (see Definition 3) were to possess a categorical utility  $u_{Z_i}$ , and suppose further that  $u_{Z_i}(\mathbf{z}^*) > u_{Z_i}(\mathbf{z})$  for all  $\mathbf{z} \in \mathcal{A}(\{\mathbf{z}^*\})$

and for all  $i \in \{1, \dots, m\}$ . Then the Pareto principle requires that

$$\mathcal{U}_{\mathcal{X}}(\mathbf{z}^*, \dots, \mathbf{z}^*) > \mathcal{U}_{\mathcal{X}}(\mathbf{z}_1, \dots, \mathbf{z}_m) \quad \forall (\mathbf{z}_1, \dots, \mathbf{z}_m) \in \mathcal{A}^{m \setminus \{(z^*, \dots, z^*)\}}. \quad (12)$$

Now suppose that  $Z_i$  views a particular profile  $\mathbf{z}^*$  as its most preferred solution. That profile, although perhaps not the one that maximizes the coordination utility, should not be barred from consideration just because it is best for  $Z_i$ . In other words, there should be no categorical discrimination against any member of the community. This form of discrimination is related to the notion of suppression, as discussed by [Hansson \(1972\)](#) and [Fishburn \(1973\)](#). An individual is suppressed if, whenever it prefers alternative  $a$  to  $a'$ , the group chooses  $a'$  over  $a$ . We extend this notion as follows:

**Definition 6** Suppose there exists a conjecture  $\mathbf{z}^* \in \mathcal{A}$  such that

$$u_{Z_i}(\mathbf{z}^*) > u_{Z_i}(\mathbf{z}) \quad \forall \mathbf{z} \in \mathcal{A} \setminus \{\mathbf{z}^*\}, \quad (13)$$

but

$$\mathcal{U}_{\mathcal{X}}(\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}^*, \mathbf{z}_{i+1}, \dots, \mathbf{z}_m) < \mathcal{U}_{\mathcal{X}}(\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}_i, \mathbf{z}_{i+1}, \dots, \mathbf{z}_m) \quad (14)$$

for all  $\mathbf{z}_i \in \mathcal{A} \setminus \{\mathbf{z}^*\}$  and all  $(\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \dots, \mathbf{z}_m) \in \mathcal{A}^{m-1}$ .  $Z_i$  is said to be subjugated. The aggregation defined by (11) is said to be socially coherent if no  $Z_i$  can be subjugated.

If  $Z_i$  is subjugated, then, whenever  $\mathbf{z}^*$  is  $Z_i$ 's most preferred profile, the coordination utility of any joint conjecture for which  $Z_i$ 's individual conjecture is  $\mathbf{z}^*$  is less than the coordination utility of the corresponding joint conjecture with  $Z_i$ 's conjecture replaced by any  $\mathbf{z}_i \neq \mathbf{z}^*$ . In particular, suppose that each individual (including  $Z_i$ ) views  $\mathbf{z}^*$  as best for it. But if  $Z_i$  is subjugated, then

$$\mathcal{U}_{\mathcal{X}}(\mathbf{z}^*, \dots, \mathbf{z}^*) < \mathcal{U}_{\mathcal{X}}(\mathbf{z}^*, \dots, \mathbf{z}^*, \mathbf{z}_i, \mathbf{z}^*, \dots, \mathbf{z}^*) \quad \forall \mathbf{z}_i \neq \mathbf{z}^*, \quad (15)$$

which contradicts (12). Thus, social coherence is required for the coordination utility to comply with the Pareto principle.

### 4.3.2 Aggregation

Our objective is to establish necessary and sufficient conditions to ensure socially coherent aggregation. We address this issue by recognizing that an isomorphic relationship exists between the concept of subjugation in the practical domain and the concept of sure loss in the epistemological domain. A sure loss is a so-called Dutch book gamble such that the gambler loses regardless of the outcome of the gamble.

**Theorem 2** (Dutch Book Theorem) *Suppose a gambler places a bet to win a payout of  $S$ . A fair entry fee for this gamble is  $pS$ , where  $p$  is the gambler's degree of belief of winning.*

(Necessity) If  $p$  violates the probability axioms, then it is possible to construct a bet such that the payout is less than the entry fee—a sure loss.

(Sufficiency) If  $p$  conforms to the probability axioms, then it is not possible to construct a bet such that the gambler sustains a sure loss.<sup>7</sup>

The relevance of the Dutch Book Theorem regarding coherent aggregation established by the following lemma, which is proved in Appendix 1:

**Lemma 1** *Subjugation is isomorphic to sure loss.*

One of the main features of probability theory is that the statistical dependencies for a collective of discrete random variables can be expressed via a DAG whose vertices are the random variables and whose edges are conditional probabilities. Such a graph is a Bayesian network (Pearl 1988; Cowell et al. 1999; Jensen 2001). Consequently, if there is to be more than a metaphorical relationship between probability theory and utility theory, we must restrict social influence to be acyclic; that is to hierarchical games. This restriction limits the generality of our approach. Nevertheless this model still represents an important generalization from mutual social independence, where all agents possess categorical utilities and thus all social relationships are trivially acyclical. The following theorem establishes conditions for socially coherent aggregation:

**Theorem 3** (The aggregation theorem) *Let  $\mathcal{X}$  be a hierarchical satisficing game as defined by Definition 4. The coordination utility is a socially coherent aggregation if, and only if,*

$$\mathcal{U}_{\mathcal{X}}(\mathbf{s}_1, \dots, \mathbf{s}_n, \mathbf{r}_1, \dots, \mathbf{r}_n) = \prod_{i=1}^n \prod_{j=1}^n u_{S_i | \text{pa}(S_i)}(\mathbf{s}_i | \boldsymbol{\sigma}_i) u_{R_j | \text{pa}(R_j)}(\mathbf{r}_j | \boldsymbol{\rho}_j). \quad (16)$$

*Proof* By the Dutch Book Theorem, sure loss cannot occur if, and only if, all belief functions are probability mass functions and if beliefs are combined according to the axioms of probability theory. Thus, by the isomorphism established by Lemma 1, subjugation cannot occur if all utility functions possess the mathematical structure of probability mass functions and are combined and manipulated according to the mathematical syntax of probability theory. Thus, a hierarchical satisficing game is isomorphic to a Bayesian network.

A fundamental property of Bayesian networks is the Markov condition: nondescendent nonparents of a vertex have no influence on the vertex, given the state of its parent vertices (Cozman 2000). Consequently, if a collective is isomorphic to a Bayesian network, the conditional utility of a self is dependent only upon the preferences of its parents. Thus we may apply the fundamental theorem of Bayesian networks, namely that the joint mass function of the vertices is the product of the conditional mass functions of all non-root vertices and the marginal mass functions of all root vertices (Pearl

<sup>7</sup> Necessity (the original theorem) was independently established by de Finetti (1937) and Ramsey (1978), and sufficiency (the converse theorem) was independently established by Kemeny (1955) and Lehman (1955).

1988; Cowell et al. 1999; Jensen 2001). Stated another way, (16) is an application of the chain rule of probability. Consequently, (16) is satisfied and social coherence is established.  $\square$

This theorem does not mean that if the utilities are not structured according to the theorem, then subjugation of individual preferences is sure to occur. Rather, it means that if the utilities are so structured and aggregated, then subjugation is impossible. This weak notion of equity is a minimal requirement for meaningful negotiations to be possible. Essentially it ensures that no individual is a victim of categorical discrimination. Each individual has a “seat at the table” in the sense that it is at least theoretically possible for its most preferred profile to be the one most preferred by the collective as well.

This aggregation structure is syntactically identical to the way conditional probability mass functions are aggregated to form a joint probability mass function for a collective of random vectors. Although the semantics are different, this theorem establishes that the preferential influence among the selves propagates through the system to create a social model in exactly the same way that probabilistic dependence among a set of random vectors propagates through the collective to create a joint distribution.

#### 4.4 Sociation

Thus far our development has assumed the full generality of conditioning, namely that a conditional utility depends on the entire conjecture profiles of all of the parents and that the conditional utility is a function of all elements of the profile. This fully general model can be extremely complex since each self is under obligation to define its preferences for every possible joint conjecture of its parents. A possible criticism of this approach is that it is unduly complex; however, it need not be more complex than is necessary to characterize the relevant social relationships.<sup>8</sup> Fortunately, there are practical ways to mitigate complexity. Although it is necessary for the theory to have the ability to accommodate maximal complexity, it can be the case that a self does not condition its preferences on the entire conjecture profiles of its parents. It can also be the case that a self’s utility does not depend upon the entire profile. To account for such situations, we introduce the notion of sociation. It is convenient to develop this concept in terms of the alternate notation introduced in Definition 3, where  $Z_i$  denotes either  $S_i$  or  $R_i$ , and  $\mathbf{z}_i$  denotes the corresponding  $\mathbf{s}_i$  or  $\mathbf{r}_i$ .

Suppose  $Z_i$  has  $p_i > 0$  parents  $\text{pa}(Z_i) = \{Z_{i_1}, \dots, Z_{i_{p_i}}\}$ , with conditioning conjecture  $\xi_i = (\mathbf{z}_{i_1}, \dots, \mathbf{z}_{i_{p_i}})$  and conditional utility  $u_{Z_i|\text{pa}(Z_i)}(\mathbf{z}_i|\xi_i)$ . A conditioning conjecture subprofile for  $Z_i$ , denoted  $\hat{\xi}_i$ , is the subset of  $\xi_i$  that influences  $Z_i$ ’s conditional utility. We then have

<sup>8</sup> “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” (Palmer 1971, p. 184).

$$u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \xi_i) = u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \hat{\xi}_i). \quad (17)$$

$Z_i$  is completely conditioning conjecture sociated if  $\hat{\xi}_i = \xi_i$ .  $Z_i$  is completely conditioning conjecture dissociated if  $\xi_i = (z_{i_1}, \dots, z_{i_{q_i}})$ , that is, the influence  $Z_{i_k}$  exerts on  $Z_i$  is with respect to only  $Z_{i_k}$ 's component of  $\mathbf{z}_{i_k}$ . If  $Z_i$  is completely conditioning conjecture dissociated, then

$$u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \xi_i) = u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | z_{i_1}, \dots, z_{i_{p_i}}). \quad (18)$$

Otherwise,  $Z_i$  is partially conditioning conjecture dissociated.

A utility subprofile, denoted  $\tilde{\mathbf{z}}_i$ , comprises the subset of the elements of  $\mathbf{z}_i$  that affect  $Z_i$ 's utility. We then have

$$u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \hat{\xi}_i) = \tilde{u}_{Z_i | \text{pa}(Z_i)}(\tilde{\mathbf{z}}_i | \hat{\xi}_i), \quad (19)$$

where  $\tilde{u}$  denotes  $u$  with the dissociated elements of its argument removed.  $Z_i$  is completely utility sociated if  $\tilde{\mathbf{z}}_i = \mathbf{z}_i$ .  $Z_i$  completely utility dissociated if  $\tilde{\mathbf{z}}_i = z_i$ , that is, if  $Z_i$ 's preferences are with respect to its own actions only, rather than for the entire profile. We then have

$$u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \xi_i) = \tilde{u}_{Z_i | \text{pa}(Z_i)}(z_i | \hat{\xi}_i). \quad (20)$$

Otherwise,  $Z_i$  is partially utility dissociated.

$Z_i$  is completely dissociated if it is both completely conditioning conjecture dissociated and completely utility dissociated, in which case,

$$u_{Z_i | \text{pa}(Z_i)}(\mathbf{z}_i | \xi_i) = \tilde{u}_{Z_i | \text{pa}(Z_i)}(z_i | z_{i_1}, \dots, z_{i_{p_i}}). \quad (21)$$

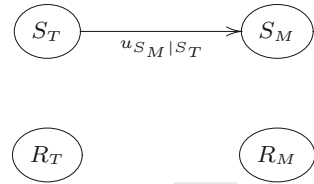
(The original treatment of satisficing games, as developed in [Stirling \(2003\)](#), was limited to the completely dissociated case).

Complete sociation is the standard assumption with classical game theory, where the utilities of all individuals are functions of the entire profile of possible actions of all individuals. For many applications, however, an individual's interest may be with respect to its own individual action, as illustrated in the following example:

**Example 2** (Tom and Martha (Continued)) We continue with the Tom and Martha party example introduced in [Example 1](#), where  $\mathcal{A}_T = \{b, c\}$ ,  $\mathcal{A}_M = \{h, l\}$ , and  $\mathcal{A} = \{(b, h), (b, l), (c, h), (c, l)\}$ . Let us suppose that Tom's selecting and rejecting selves have categorical preferences of the form  $u_{S_T}(\mathbf{s}_T)$  and  $u_{R_T}(\mathbf{r}_T)$ , respectively, but that Martha is amenable to adjusting her notion of benefit as a function of Tom's preferences, and thus possesses conditional selecting and rejecting utilities of the form  $u_{S_M|S_T R_T}(\mathbf{s}_M | \mathbf{s}_T, \mathbf{r}_T)$  and  $u_{R_M|S_T R_T}(\mathbf{r}_M | \mathbf{s}_T, \mathbf{r}_T)$ . The coordination utility is then

$$\begin{aligned} \mathcal{U}_{TM}(\mathbf{s}_T, \mathbf{s}_M, \mathbf{r}_T, \mathbf{r}_M) = \\ u_{S_T}(\mathbf{s}_T) u_{S_M|S_T R_T}(\mathbf{s}_M | \mathbf{s}_T, \mathbf{r}_T) u_{R_T}(\mathbf{r}_T) u_{R_M|S_T R_T}(\mathbf{r}_M | \mathbf{s}_T, \mathbf{r}_T). \end{aligned} \quad (22)$$

**Fig. 2** The DAG for the Tom and Martha party



Now suppose further that both of Tom's selves are only concerned with the location. Also, suppose that Martha's rejecting self is concerned only with the menu, and her selecting self preferences regarding both location and menu are influenced only by Tom's preferences for location. In this case, Tom's selecting and rejecting selves are both completely dissociated, Martha's rejecting self is completely dissociated, and her selecting self is partially utility sociated and completely conditioning conjecture dissociated. Under these assumptions, Tom's selecting and rejecting utilities are a functions of location only, and Martha's rejecting utility is a function of menu only. Martha's conditional selecting utility is a function of both location and menu as influenced only by Tom's preferences regarding location. The coordination utility for this partially dissociated scenario is

$$\mathcal{U}_{TM}(s_T, \mathbf{s}_M, r_T, r_M) = \tilde{u}_{s_T}(s_T) \hat{u}_{s_M|s_T}(\mathbf{s}_M|s_T) \tilde{u}_{r_T}(r_T) \tilde{u}_{r_M}(r_M). \quad (23)$$

It is important to appreciate the distinction between (22) and (23). In (22), the conjectures  $\mathbf{s}_T$ ,  $\mathbf{s}_M$ ,  $\mathbf{r}_T$ , and  $\mathbf{r}_M$  are all elements of  $\mathcal{A}_T \times \mathcal{A}_M$ , but in (23),  $s_T$  and  $r_T$  are elements of  $\mathcal{A}_T$ ,  $r_M$  is an element of  $\mathcal{A}_M$ , but  $\mathbf{s}_M$  is an element of  $\mathcal{A}_T \times \mathcal{A}_M$ . To emphasize this distinction, let us write Martha's selecting conjecture as  $\mathbf{s}_M = (\zeta_T, \varsigma_M)$ . Although  $s_T$  and  $\zeta_T$  are both selecting conjectures for Tom,  $s_T$  is Tom's selecting conjecture for himself, but  $\zeta_T$  is Martha's selecting conjecture for Tom. We may rewrite (23) as

$$\mathcal{U}_{TM}(s_T, \zeta_T, \varsigma_M, r_T, r_M) = \tilde{u}_{s_T}(s_T) \hat{u}_{s_M|s_T}(\zeta_T, \varsigma_M|s_T) \tilde{u}_{r_T}(r_T) \tilde{u}_{r_M}(r_M). \quad (24)$$

The DAG corresponding to this model is illustrated in Fig. 2. Notice that  $\text{pa}(S_M) = \{S_T\}$  and  $\text{pa}(S_T) = \text{pa}(R_T) = \text{pa}(R_M) = \emptyset$ .

This example illustrates the role of dissociation. The fully sociated coordination utility for Tom and Martha is a function of eight independent variables and assumes 256 values, but the dissociated coordination utility is a function of five independent variables and assumes only 32 values—an eightfold reduction in complexity from the fully sociated version.

## 5 Compromise and negotiation

By viewing an individual as the composite of two distinct selves, we open the possibility of developing an approach to decision making that takes advantage of this refinement. However, although the decision problem may be expressed in terms of separate selves, these selves are not empowered to act individually; they simply represent facets of  $X_i$ . Thus, to solve a satisficing game we must define a solution concept

for  $X_i$ . Our approach is to exploit the isomorphic relationship of the coordination utility to a joint probability mass function and compute the ex post marginal utilities for the individuals and for the group. These marginals can then be used to define satisficing sets for the individuals and for the group, which can serve as the basis for compromise and negotiation.

## 5.1 Individual satisficing

**Definition 7** Let  $s_i$  and  $r_i$  denote the  $i$ th elements of the profiles  $\mathbf{s}_i$  and  $\mathbf{r}_i$ , respectively. The individual selectability utility and individual rejectability utility functions, denoted  $w_{s_i}$  and  $w_{r_i}$ , are obtained by summing the coordination utility over all arguments except the  $s_i$  and  $r_i$ , respectively, yielding

$$w_{s_i}(s_i) = \sum_{\sim s_i} \mathcal{U}_{\mathcal{X}}(\mathbf{s}_1, \dots, \mathbf{s}_n, \mathbf{r}_1, \dots, \mathbf{r}_n) \quad (25)$$

$$w_{r_i}(r_i) = \sum_{\sim r_i} \mathcal{U}_{\mathcal{X}}(\mathbf{s}_1, \dots, \mathbf{s}_n, \mathbf{r}_1, \dots, \mathbf{r}_n), \quad (26)$$

where the notation  $\sum_{\sim a_i}$  means that for any function  $g(a_1, \dots, a_k)$ , the sum is taken over all arguments except  $a_i$ , that is,

$$\sum_{\sim a_i} g(a_1, \dots, a_k) = \sum_{a_j: j \neq i} g(a_1, \dots, a_k).$$

The individually satisficing set, denoted

$$\Sigma_i = \{a_i \in \mathcal{A}_i: w_{s_i}(a_i) \geq q_i w_{r_i}(a_i)\}, \quad (27)$$

is the set of all alternatives in  $\mathcal{A}_i$  that are satisficing for  $X_i$ . The satisficing rectangle is the Cartesian product of the individually satisficing sets, denoted

$$\mathcal{R}_{\mathcal{X}} = \Sigma_1 \times \dots \times \Sigma_n. \quad (28)$$

## 5.2 Group satisficing

The existence of a comprehensive social model in the form of a coordination utility generates an emergent notion of group preference. However, although the satisficing rectangle identifies the profiles that are simultaneously satisficing for all individuals, it does not necessarily correspond to the set of profiles that are satisficing from the group perspective. To define this set, we must compute the marginal utility for the group as a function of the conjectures of each individual over its own alternative set.

**Definition 8** Let  $s_i$  and  $r_i$  denote the  $i$ th elements of the profiles  $\mathbf{s}_i$  and  $\mathbf{r}_i$ , respectively, and form the profiles  $(s_1, \dots, s_n)$  and  $(r_1, \dots, r_n)$ . The satisficing group utility is the

marginal utility formed by summing over all elements of the profiles  $\mathbf{s}_i$  and  $\mathbf{r}_i$  except the  $i$ th elements of each, yielding

$$v_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n) = \sum_{\sim s_1} \dots \sum_{\sim s_n} \dots \sum_{\sim r_1} \dots \sum_{\sim r_n} \mathcal{U}_{\mathcal{X}}(\mathbf{s}_1, \dots, \mathbf{s}_n, \mathbf{r}_1, \dots, \mathbf{r}_n). \quad (29)$$

The satisficing group utility is a measure of the degree of alignment for the collective as a function of each individual's conjecture regarding its own interests only. The interpretation of alignment depends on the context. To illustrate, if the objective of a two-agent (four-self) collective is to cooperate (e.g., a team), then

$$v_{x_1 x_2}(s_1, s_2, r_1, r_2) > v_{x_1 x_2}(s'_1, s'_2, r'_1, r'_2) \quad (30)$$

means that the profile  $(s_1, s_2, r_1, r_2)$  induces greater positive alignment of interests than does  $(s'_1, s'_2, r'_1, r'_2)$ , that is, the more interests are positively aligned, the greater the cooperation. However, if the objective of the collective is to compete (e.g., an athletic contest or military engagement), then (30) means that  $(s_1, s_2, r_1, r_2)$  induces greater negative alignment of interests than does  $(s'_1, s'_2, r'_1, r'_2)$ , that is, the more the interests are negatively aligned, the greater the opposition.

**Definition 9** The group selectability utility and group rejectability utility functions are the marginal joint utilities obtained by summing the satisficing group utility over  $r_1, \dots, r_n$  and  $s_1, \dots, s_n$ , respectively, yielding

$$v_{\mathcal{S}}(s_1, \dots, s_n) = \sum_{r_1, \dots, r_n} v_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n) \quad (31)$$

$$v_{\mathcal{R}}(r_1, \dots, r_n) = \sum_{s_1, \dots, s_n} v_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n). \quad (32)$$

The group satisficing set is the set of all profiles that are satisficing from the perspective of the group, that is,

$$\Sigma_{\mathcal{X}} = \{\mathbf{a} \in \mathcal{A}: v_{\mathcal{S}}(\mathbf{a}) \geq q_{\mathcal{X}} v_{\mathcal{R}}(\mathbf{a})\}, \quad (33)$$

where  $q_{\mathcal{X}}$  is the  $q$ -value for the group.

The satisficing rectangle  $\mathcal{R}_{\mathcal{X}}$  (28), which defines the set of profiles such that the  $i$ th element is satisficing for  $X_i$ , is different from the group satisficing set  $\Sigma_{\mathcal{X}}$  (33), which defines the set of profiles that are satisficing for the group as a whole, and does not necessarily require each individual's satisficing criteria to be met. The following theorem provides insight regarding the relationship between the sets defined by (33) and (28).

**Theorem 4** For  $\mathcal{X} = \{X_1, \dots, X_n\}$ , consider a subprofile  $(a_1, \dots, a_k) \in \Sigma_1 \times \dots \times \Sigma_k$  for  $k < n$ . If  $q_{\mathcal{X}} \leq q_i$  for  $i \in \{1, \dots, k\}$ , then there exists a subvector  $(a_{k+1}^*, \dots, a_n^*) \in \mathcal{A}_{k+1} \times \dots \times \mathcal{A}_n$  such that  $(a_1, \dots, a_k, a_{k+1}^*, \dots, a_n^*) \in \Sigma_{\mathcal{X}}$ .

*Proof* We establish this result by a contrapositive argument. We first observe that since  $v_{s_i}(s_i)$  is the  $i$ th selectability marginal of  $\mathcal{U}_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n)$ , and since  $v_{\mathcal{S}}(s_1, \dots, s_n)$  is the joint selectability marginal of  $\mathcal{U}_{\mathcal{X}}(s_1, \dots, s_n, r_1, \dots, r_n)$ , it follows that  $v_{s_i}(s_i)$  is the  $i$ th selectability marginal  $v_{\mathcal{S}}(s_1, \dots, s_n)$ . By a similar argument,  $v_{r_i}(r_i)$  is the  $i$ th rejectability marginal of  $v_{\mathcal{R}}(r_1, \dots, r_n)$ .

Suppose

$$v_{\mathcal{S}}(a_1, \dots, a_k, a_{k+1}, \dots, a_n) < q_{\mathcal{X}} v_{\mathcal{R}}(a_1, \dots, a_k, a_{k+1}, \dots, a_n) \quad (34)$$

for all  $(a_{k+1}, \dots, a_n) \in \mathcal{A}_{k+1} \times \dots \times \mathcal{A}_n$ . Then, for  $i \in \{1, \dots, k\}$ ,

$$\begin{aligned} v_{s_i}(a_i) &= \sum_{\sim a_i} v_{\mathcal{S}}(a_1, \dots, a_k, a_{k+1}, \dots, a_n) \\ &< q_{\mathcal{X}} \sum_{\sim a_i} v_{\mathcal{R}}(a_1, \dots, a_k, a_{k+1}, \dots, a_n) \\ &= q_{\mathcal{X}} v_{r_i}(a_i) \leq q_i v_{r_i}(a_i), \end{aligned} \quad (35)$$

Hence,  $a_i \notin \Sigma_i$ , a contradiction.  $\square$

The condition  $q_{\mathcal{X}} \leq q_i$  is consistent with the principles of justice espoused by Rawls (1971), who argues, essentially, that a society as a whole cannot be better off than its least advantaged member. This theorem establishes an important result: no subset of agents can be categorically discriminated by the larger group. In other words, every individual has a seat at the table in the sense that if an alternative is individually satisficing for it, then that alternative is part of some satisficing profile for the group. Thus, each individual can be assured that it is can bargain in good faith. This condition is perhaps the weakest one possible for meaningful negotiations to occur.

The group satisficing set and the satisficing rectangle provide a basis for negotiation. Let  $\mathcal{C}_{\mathcal{X}} = \Sigma_{\mathcal{X}} \cap \mathcal{R}_{\mathcal{X}}$  denote the compromise set. If  $\mathcal{C}_{\mathcal{X}} \neq \emptyset$ , then all profiles in this set are both group and individually satisficing. If  $\mathcal{C}_{\mathcal{X}}$  is not a singleton set, then, depending on the context, a tie-breaking logic may be applied to define a unique solution. If  $\mathcal{C}_{\mathcal{X}} = \emptyset$ , then each individual has the ability to increase the size of its individually satisficing set by incrementally reducing  $q_i$ . In this way each individual has control over the amount of performance it is willing to sacrifice to accommodate the interests of the group. If the compromise set remains empty after each agent has reduced its  $q_i$  to its minimum acceptable level, then an impasse cannot be avoided. Such a society is dysfunctional.

*Example 3* [Tom and Martha (Concluded)] We can now compute the satisficing solution for the Tom and Martha party. Let us assume the following numerical values. The categorical selecting and rejecting utilities for Tom are  $u_{s_T}(b) = 0.6$ ,  $u_{s_T}(c) = 0.4$ ,  $u_{r_T}(b) = 0.4$ , and  $u_{r_T}(c) = 0.6$ . The rejecting utility for Martha is  $u_{r_M}(h) = 0.3$ , and  $u_{r_M}(l) = 0.7$ , and Martha's conditional selecting utility is displayed in Table 4.

By straightforward calculations, the group selecting and rejecting utilities are as displayed in Table 5, from which we observe that the group satisficing set with  $q_{TM} = 1$  is  $\Sigma_{TM} = \{(b, h), (b, l), (c, h)\}$ . Figure 3 provides a cross-plot of these values, from

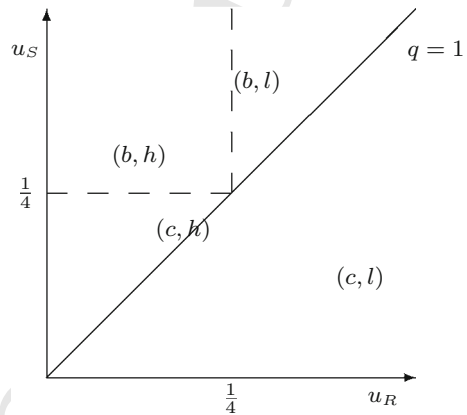
**Table 4** Martha's partially dissociated conditional selecting utilities

$s_T$	$u_{S_M S_T}(s_M s_T)$			
	$(b, h)$	$(b, l)$	$(c, h)$	$(c, l)$
$b$	0.5	0.2	0.0	0.3
$c$	0.25	0.25	0.25	0.25

**Table 5** Joint selecting and rejecting utilities for the Tom and Martha party

$(a_T, a_S)$	$v_{S_T S_M}(a_T, a_M)$	$v_{R_T R_M}(a_T, a_M)$
$(b, h)$	0.3	0.12
$(b, l)$	0.3	0.28
$(c, h)$	0.2	0.18
$(c, l)$	0.2	0.42

**Fig. 3** Cross-plot of group-level selectability versus rejectability for the Tom and Martha party



which it is evident, using the qualitative designations introduced in Sect. 3, that the collective is in a dispositional state of gratification with respect to  $(b, h)$ , ambivalence with respect to  $(b, l)$ , and dubiety with respect to  $(c, h)$ .

The individual selectability and rejectability utilities are

$$v_{S_T}(b) = 0.6 \quad v_{S_T}(c) = 0.4 \quad (36)$$

$$v_{R_T}(b) = 0.4 \quad v_{R_T}(c) = 0.6 \quad (37)$$

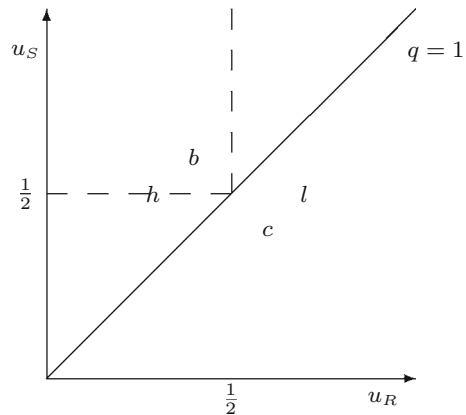
and

$$v_{S_M}(h) = 0.5 \quad v_{S_M}(l) = 0.5 \quad (38)$$

$$v_{R_M}(h) = 0.3 \quad v_{R_M}(l) = 0.7, \quad (39)$$

from which, setting  $q_T = q_S = 1$ ,  $\Sigma_T = \{b\}$  and  $\Sigma_M = \{h\}$ . Thus, the satisficing rectangle is the singleton set  $\mathcal{R}_{TM} = \{(b, h)\}$  and the compromise set is  $\mathcal{C}_{TM} = \Sigma_{TM} \cap \mathcal{R}_{TM} = \{(b, h)\}$ . Figure 4 displays the individual selectability versus rejectability

**Fig. 4** Cross-plot of individual selectability versus rejectability for Tom (options  $b$  and  $c$ ) and Martha (options  $h$  and  $l$ )



cross-plots. In terms of individual dispositions, Tom is gratified, but Martha is on the border between gratification and dubiety.

## 6 Conclusion

We have introduced a heuristic concept of “good enough” that serves as the basis for a formalized mathematical approach to decision-making that is founded on the premise that the perspective of seeking to avoid failure is more amenable to compromise and negotiation than is the perspective of seeking to maximize individual benefit. To implement this perspective, we view each individual as a composite of two selves, one concerned with effectiveness, the other with efficiency. Thus our approach does not take preferences as fixed but explicitly recognizes the tradeoffs between costs and decision failure.

Our contribution also naturally extends to multiagent settings and serves as a new framework within which to study the behavior of complex social systems. Specifically, we offer a more realistic conception of multiagent decision making where individual influence and social interaction play a central role in emergent choices. We introduce conditional preference relationships that enable individuals to incorporate the preferences of others into their own rationality as a function of social context. As these conditional preferences propagate through the group, they generate emergent social relationships that govern their collective behavior. Aggregation is the result of this propagation, which creates an emergent social model that characterizes the alignment of interests among the individuals that arises due to social influence they exert on each other. A key feature of this approach is the generation of both individually and group satisficing solutions that form the basis for negotiating protocols that lead to compromise solutions that are not only acceptable to the individuals, but further the (emergent) interests of the group as a whole.

One of the criticisms of using game theory for social modeling is that group behavior is derived exclusively from individual preferences (Gilbert 2008; Hollis 1998). Indeed, the classical way to construct a game-theoretic model is to make minimal assumptions

about the values of the individuals and then investigate what can be deduced about the behavior of the collective. The logical prescriptive consequence of this model is that group behavior should conform to a Nash equilibrium solution: a constrained optimal solution where each player seeks the best outcome for itself under the assumption that others will do likewise. But optimization is an intrinsically individual-based concept. The only way a group can optimize if it is viewed as a decision-making entity. However, expanding the concept of individual rationality to account for social as well as material benefits by means of conditional preferences opens the way to new group-level solution concepts that are not founded on optimization. In particular, the neo-satisficing provides a solution concept that is naturally amenable to truly coordinated decision making.

## Appendix: Proof of the isomorphism lemma

**Lemma 2** *Subjugation is isomorphic to sure loss.*

*Proof* To establish this result we first prove that the categorical and conditional utilities are order isomorphic to marginal and conditional probability mass functions. Without loss of generality, let us restrict attention to a two-agent collective  $\{Z_1, Z_2\}$ , with  $Z_1$  possessing a categorical utility  $u_{Z_1}: \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow [0, 1]$  and  $Z_2$  possessing a family of conditional utilities  $\{u_{Z_2|Z_1}(\cdot|z_1, z_2): \rightarrow [0, 1] \forall (z_1, z_2) \in \mathcal{A}_1 \times \mathcal{A}_2\}$ . To proceed, let  $\Omega_1$  and  $\Omega_2$  be an arbitrary probability sample space of distinct elements and cardinality equal to the cardinality of  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , respectively, and let  $\{Y_1, Y_2\}$  be discrete bijective function (random variables in statistical parlance) defined over  $\Omega_1 \times \Omega_2$  such that  $Y_1$  possesses a probability mass function  $p_{Y_1}$  over  $\Omega_1 \times \Omega_2$  and  $Y_2$  possesses a family of conditional probability mass functions  $\{p_{Y_2|Y_1}(\cdot|\omega_1, \omega_2): \Omega_1 \times \Omega_2 \rightarrow [0, 1] \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2\}$ . Now let  $g: \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow \Omega_1 \times \Omega_2$  be a bijective mapping of the form  $g(z_1, z_2) = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ . Define  $g$  such that

$$u_{Z_1}(z_1, z_2) = p_{Y_1}[g(z_1, z_2)] = p_{Y_1}[\omega_1, \omega_2]$$

for all  $(z_1, z_2) \in \mathcal{A}_1 \times \mathcal{A}_2$ , and

$$u_{Z_2|Z_1}(z_1, z_2|z'_1, z'_2) = p_{Y_2|Y_1}[g(z_1, z_2)|g(z'_1, z'_2)] = p_{Y_2|Y_1}[\omega_1, \omega_2|\omega'_1, \omega'_2]$$

for all  $(z'_1, z'_2) \in \mathcal{A}_1 \times \mathcal{A}_2$ . It is immediate that this mapping establishes the structural equivalence of the benefit criterion regarding  $\mathcal{A}_1 \times \mathcal{A}_2$  and the belief criterion regarding  $\Omega_1 \times \Omega_2$ . It remains to confirm that  $Z_2$  computing its utility of  $(z_1, z_2)$  given that  $Z_1$  conjectures  $(z'_1, z'_2)$  is equivalent to  $Y_2$  computing its belief that  $g(z_1, z_2)$  is realized given that  $Y_1$  asserts that  $g(z_1, z_2)$  is realized. To establish this equivalence, we simply observe that the conjecture by  $Z_1$  and the realization assertion by  $Y_1$  are both antecedents of hypothetical propositions whose consequents are  $u_{Z_2|Z_1}(z_1, z_2|z'_1, z'_2)$  and  $p_{Y_2|Y_1}[g(z_1, z_2)|g(z'_1, z'_2)]$ , respectively. This establishes the order isomorphism.

To establish the isomorphism between subjugation and sure loss, let  $\{Y_1, Y_2\}$  be a collective of discrete random variables, each defined over the product sample space  $\Omega = \Omega_1 \times \Omega_2$ . Let  $p_{Y_i}: \Omega \rightarrow \mathbb{R}$  be a belief function (not necessarily a probability)

such that  $p_{Y_1}(\omega) \geq p_{Y_1}(\omega')$  means that the belief that  $\omega$  will be realized is at least as great as the belief that  $\omega'$  will be realized. Also, let  $p_{Y_1 Y_2}: \Omega^2 \rightarrow \mathbb{R}$  be a belief function such that  $p_{Y_1 Y_2}(\omega_1, \omega_2) \geq p_{Y_1 Y_2}(\omega'_1, \omega'_2)$  means that the belief that  $(\omega_1, \omega_2)$  is realized is at least as great as the belief that  $(\omega'_1, \omega'_2)$  is realized.

Suppose there exists  $\omega_1^* \in \Omega$  such that

$$p_{Y_1}(\omega_1^*) > p_{Y_1}(\omega_1) \quad \forall \omega_1 \in \Omega \setminus \{\omega_1^*\} \quad (40)$$

and that

$$p_{Y_1 Y_2}(\omega_1^*, \omega_2) < p_{Y_1 Y_2}(\omega_1, \omega_2) \quad (41)$$

for all  $\omega_1 \in \Omega \setminus \{\omega_1^*\}$  and for all  $\omega_2 \in \Omega$ . Thus, even though  $\omega_1^*$  is  $Y_1$ 's most strongly believed event, the belief regarding the realization of any joint event for which  $\omega_1^*$  is  $Y_1$ 's realization is weaker than the belief regarding the realization of the corresponding joint event with any other  $\omega_1$  as  $Y_1$ 's realization.

If, on the basis of (40) one were to enter a lottery to earn \$1 if  $\omega_1^*$  is realized, a fair entry fee would be  $q_1 > \frac{1}{2}$ . On the other hand, if, on the basis of (41), one were to earn \$1 if  $\omega_1^*$  is not realized, then a fair entry fee would be  $q_2 > \frac{1}{2}$ . By combining bets, one would be certain of winning \$1 with a fair entry fee of  $q_1 + q_2 > 1$ , therefore resulting in a sure loss.

It is immediate that the relationships given by (13) and (40) and by (14) and (41) are identical. Thus, sure loss is isomorphic to subjugation.  $\square$

## References

- Abbas, A.E. (2009). From bayes' nets to utility nets. In: *Proceedings of the 29th international workshop on bayesian inference and maximum entropy methods in science and engineering* (pp 3–12)
- Abbas, A. E., & Howard, R. A. (2005). Attribute dominance utility. *Decision Analysis*, 2(4), 185–206.
- Arrow, K. J. (1986). Rationality of self and others in an economic system. In R. M. Hogarth & M. W. Reder (Eds.), *Rational choice*. Chicago: University of Chicago Press.
- Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of cooperation. *Research in Economics*, 23, 117–147.
- Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton: Princeton University Press.
- Battigalli, P., & Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144, 1–35.
- Bendor, J. B., Kumar, S., & Siegel, D. A. (2009). Satisficing: A 'pretty good' heuristic. *The B E Journal of Theoretical Economics*, 9(1), 1–36.
- Berhold, M. H. (1973). The use of distribution functions to represent utility functions. *Management Science*, 23, 825–829.
- Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological review*, 120(3), 512.
- Bolton, G. E., & Ockenfels, A. (2005). A stress test of fairness measures in models of social utility. *Economic Theory*, 24(4), 957–982.
- Busemyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432.
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton Univ. Press.
- Camerer, C., Lowenstein, G., & Rabin, M. (Eds.). (2004a). *Advances in behavioral economics*. Princeton: Princeton Univ. Press.
- Camerer, C., et al. (2004b). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.

- Castagnoli, E., & LiCalzi, M. (1996). Expected utility without utility. *Theory and Decision*, 41, 281–301.
- Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26, 139–198.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. New York: Springer.
- Cozman, F. G. (2000). Credal networks. *Artificial Intelligence*, 120, 199–233.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7:1–68 (translated as 'Forsight. Its Logical Laws, Its Subjective Sources', in *Studies in Subjective Probability*, H. E. Kyburg Jr. and H. E. Smokler (eds.), Wiley, New York, NY, 1964, pages 93–158).
- Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268–298.
- Dyer, J. S., & Sarin, R. K. (1979). Measurable multiattribute value functions. *Operations Research*, 27, 810–822.
- Elster, J. (Ed.). (1985). *The multiple self*. Cambridge: Cambridge University Press.
- Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114, 817–868.
- Fishburn, P. C. (1973). *The theory of social choice*. Princeton: Princeton University Press.
- Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1, 60–79.
- Gilbert, M. (2008). Social convention revisited. *Topoi*, 27, 5–16.
- Gilboa, I., & Schmeidler, D. (1988). Information dependent games. *Economics Letters*, 27, 215–221.
- Hansson, B. (1972). The independence condition on the theory of social choice, working paper no. 2, The Mattias Frenling Society, Department of Philosophy, Lund.
- Harsanyi, J. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63, 315.
- Henrich, J., et al. (2005). "Economic Man" in cross-culturative perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795–855.
- Hicks, J. R. (1939). Foundations of welfare economics. *Economic Journal*, 49(196), 696–712.
- Hollis, M. (1998). *Trust within reason*. Cambridge: Cambridge University Press.
- James, W. (1956). *The will to believe and other essays*. New York: Dover.
- Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer.
- Kaldor, N. (1939). Welfare propositions of economic and interpersonal comparisons of utility. *Economic Journal*, 49(195), 549–552.
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives*. Cambridge: Cambridge University Press.
- Kemeny, J. (1955). Fair bets and inductive probabilities. *Journal of Symbolic Logic*, 20(1), 251–262.
- Lam, S. W., Ng, T. S., Sim, M., & Song, J. H. (2013). Multiple objective satisficing under uncertainty. *Operations Research*, 61, 214–217.
- Landau, M., & Chisholm, D. (1995). The arrogance of optimism: Notes on failure-avoidance management. *Journal of Contingencies and Crisis Management*, 3, 67–80.
- Lehman, R. S. (1955). On conformation and rational betting. *Journal of Symbolic Logic*, 20(1), 263–273.
- Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge: MIT Press.
- Margolis, H. (1990). Dual utilities and rational choice. In J. Mansbridge (Ed.), *Beyond self-interest*, chap 15 (pp. 239–253). Chicago II: University of Chicago Press.
- Murray, J.A.H., Bradley, H., Craigie, W.A., & Onions, C.T. (Eds.). (1991). *The compact oxford english dictionary*. Oxford: The Oxford Univ. Press.
- Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18, 155–162.
- Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, 54, 289–295.
- von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior*. Princeton: Princeton University Press. (2nd ed., 1947).
- Palmer, F. R. (1971). *Grammar*. Harmondsworth: Penguin.
- Pazgal, A. (1997). Satisficing leads to cooperation in mutual interests games. *International Journal of Game Theory*, 26(4), 439–453.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.
- Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper & Row.

- Ramsey, F. P. (1978). Truth and probability. In D. H. Mellor (Ed.), *Foundations: Essays in philosophy, logic, mathematics, and economics* (pp. 58–100). Atlantic Highlands: The Humanities Press.
- Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.
- Ross, D. (2014). Theory of conditional games. *Journal of Economic Methodology*, 21, 193–198.
- Scitovsky, T. (1941). A note on welfare propositions in economics. *Review of Economic Studies*, 9(1), 77–88.
- Sen, A. K. (1990). Rational fools: A critique of the behavioral foundations of economic theory. In J. J. Mansbridge (Ed.), *Beyond self-interest, chap 2* (pp. 25–43). Chicago, IL: University of Chicago Press.
- Shubik, M. (1982). *Game theory in the social sciences*. Cambridge: MIT Press.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 59, 99–118.
- Simon, H.A. (1982). *The role of expectations in an adaptive or behavioristic model*. In: Models of Bounded Rationality, vol 2 (pp 380–399). Cambridge: MIT Press.
- Steedman, I., & Krause, U. (1985). Goethe's *Faust*, arrow's possibility theorem and the individual decision maker. In J. Elster (Ed.), *The multiple self, chap 8* (pp. 197–231). Cambridge: Cambridge University Press.
- Stirling, W. (2012). *Theory of conditional games*. Cambridge: Cambridge University Press.
- Stirling, W. C. (2003). *Satisficing games and decision making: With applications to engineering and computer science*. Cambridge: Cambridge University Press.
- Stirling, W. C., & Felin, T. (2013). Game theory, conditional preferences, and social influence. *PLoS One*, 8(2), e56751. doi:[10.1371/journal.pone.0056751](https://doi.org/10.1371/journal.pone.0056751).
- Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, 10, 69–89.
- Sugden, R. (2000). Team preferences. *Economics and Philosophy*, 16, 175–204.
- Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations*, 6, 165–181.
- Unger, P. (1975). *Ignorance: a case for skepticism*. Oxford: Oxford University Press.
- Whitehead, A. N. (1937). *Adventures in ideas*. London: Macmillan.