

Reasoning About Groups: A Cognitive Model for the Social Behavior Spectrum

Inon Zuckerman
Institute for Advanced Computer Studies
University of Maryland
College Park, Maryland 20742, USA
inon@cs.umd.edu

Meirav Hadad
Department of Computer Science
Bar-Ilan University
Ramat-Gan, 52900, Israel
hadad@cs.biu.ac.il

Abstract—An important aspect of *Social Intelligence* is the ability to correctly capture the social structure and use it to navigate and achieve ones goals. In this work we suggest a mental model that provides agents with similar social capabilities. The model captures the entire *social behavior spectrum*, and provides design principles that allow agents to reason and change their behavior according to their perception of the cooperative/competitive nature of the society. We also describe computationally the maximum attainable *benefits* when agents belong to different kinds of social groups. We conclude by exploring the group membership problem as a constraints satisfaction problem, and evaluate few heuristics.

Keywords-Autonomous agents, Social agents, BDI

I. INTRODUCTION

In human *society* individuals are pursuing their own interests while considering the impact of their actions on their surrounding. Accordingly, an important aspect of *Social Intelligence* is the ability to correctly capture the social structure and use it to navigate and achieve ones goals. In this work we suggest a mental model that provides agents with similar social capabilities. The model captures the entire *social behavior spectrum* [1], and provides design principles that will allow agents to reason and change their behavior according to their subjective perception of the cooperative/competitive nature of the society.

The suggested model would enable the implementation of *generic* social competent agents that would interact differently towards different groups. Such agents might, for instance, act cooperatively on some occasions, and on others they might reason (according to their beliefs) that competitive behavior should be applied to achieve their goals. Possible applications are quite obvious and span any autonomous agent (robot) that represents the interests of humans in virtual (physical) social environments. The model also intend to address a new class of important problems, where an agent can be part of different types of groups and acts accordingly (see [2]), thus it must consider how its actions affects other group members.

Our model builds upon the Belief-Desire-Intention (BDI) model [3] that was developed to explain human practical reasoning. The model is highly applicable as a software model for developing bounded rational intelligent agents (e.g. [4]). While previous models described specific parts

of the social behavior spectrum (in particular joint and individual desires), in this work we aim to provide a formal model that spans the full social spectrum, including relations that have not yet received any attention.

To do so we first extend the common *Desires* model to account for different relationship between desires. We then present the *Social Behavior Activity* (SBA) model, a BDI-based mental model that spans the social spectrum. The contribution of the SBA mental model is twofold: first, it provides a formal, valid theoretical foundation that could be used to explain and predict agent's social behavior in realistic environment. Second, it serves as a set of design principles to guide the creation of agents that would be able to engage in behaviors that span the entire social spectrum.

Following the presentation of the model, we explore the agent's behavior on the spectrum by computationally describing the maximum attainable *benefit* (defined later on) when it belongs to different types of social groups. The computational analysis provides a basis for building agents that are able to switch from competitive to cooperative interactions, or to individualistic behavior. Finally, as the mental model dictates the agent to have the ability to reason about group membership, which we prove to be NP-complete, we present a way to formulate the problem as a constraints satisfaction problem and present and evaluate possible heuristics to speed-up the search. The main contributions of our paper are:

- An extension of the BDI's *Desires* model to account for different relationship among desires.
- The presentation of the SBA model for the construction of social competent agents.
- An analytical exploration of different social groups.
- A constraints satisfaction based solution and heuristics to the group membership problem.

II. THE BDI LANGUAGE

A. Basic Definitions

Our BDI formalism is derived from the one proposed in [5]. We have a basic propositional language L which is used by the agent to express beliefs about the world. We will simply assume that it contains the usual classical logic

connectives ($\top, \perp, \wedge, \vee, \neg, \rightarrow, \leftrightarrow, \vdash$), and their classical interpretation. We assume that the belief change operations $\dot{+}$ (revision) and $\dot{-}$ (contraction) have been defined for L [6].

Next, we assume a set $\mathcal{A} = \{1, 2, \dots, n\}$ of agents. For each agent $i \in \mathcal{A}$ we assume a finite set of actions, $Ac_i = \{\alpha_1, \alpha_2, \dots\}$. The set Ac_i represents the abilities of agent i . Agent i 's knowledge about how to bring about states of affairs in the environment is captured by a set of recipes $R_i = \{\langle \alpha, \varphi \rangle \mid \alpha \in Ac_i \text{ and } \varphi \in L\}$, such that action α is executable if its preconditions are satisfied and the termination of α 's performance results in φ being true (see, e.g., [7] for discussion on the notion of ‘‘plan as recipe’’). For every recipe $r = \langle \alpha, \varphi \rangle$, we assume there is a proposition $r_{\alpha, \varphi} \in L$. Intuitively, $r_{\alpha, \varphi} \in L$ will be used to mean that: (1) the action α is executable, in that its preconditions are currently satisfied, and (2) the performance of α terminates and makes φ true. Only those recipes whose actions the agent believes can be executed are listed as beliefs using the propositions $r_{\alpha, \varphi}$. We will assume that in all recipes $\langle \alpha, \varphi \rangle$, the formula φ is an atomic proposition.

B. BDI Structures

Given the set of beliefs \mathcal{B}_i , desires \mathcal{D}_i , intentions \mathcal{I}_i , valuation function v^i and cost function c^i , A BDI structure for agent i is: $\mathcal{S}_i = \langle \mathcal{B}_i, \mathcal{D}_i, \mathcal{I}_i, v^i, c^i \rangle$.

- **Beliefs** — modeled in the conventional way and describe knowledge about the world or about other agents. \mathcal{B}_i stands for the beliefs of the agent, and it is a logical closure of a finite set of sentences. Formally, $\mathcal{B}_i = \{b \in L \mid \mathcal{B}_i^0 \vdash b\}$, where $\mathcal{B}_i^0 \subset L$ is a finite set that represents the basis set of beliefs for i .
- **Desires** — $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \dots \cup \mathcal{D}_n$, where $\mathcal{D}_i \subset L$ represents the set of desires for agent i . We will use d_i^1, \dots, d_i^m as elements in \mathcal{D}_i . We assert that desires will stay consistent by the belief revision operators.
- **Intentions** — \mathcal{I}_i stands for the intentions of the agent i , where $\mathcal{I}_i \subseteq \mathcal{R}_i$.
- **Valuation** — Defined as $v^i : 2^{(\mathcal{D}_i)} \rightarrow R^+$ where R^+ is the set of positive real numbers.
- **Cost** — Defined for a finite set of actions $c^i : 2^{(Ac_i)} \rightarrow R^+$, which must satisfy the following condition: if $K \subseteq K' \subseteq Ac_i$ then $c^i(K') \geq c^i(K)$.

We require that v^i satisfies the following ‘‘monotonicity’’ condition: for $T_i \subseteq \mathcal{D}_i$, then $v^i(T_i) \leq v^i(\mathcal{D}_i)$, and ‘‘additivity’’ condition: $v^i(\{d_i^1, d_i^2\}) = v^i(d_i^1) + v^i(d_i^2)$. In addition, we assert that the results of conflicting actions will be determined by nature: i.e. when two agents trying to achieve d_i and $\neg d_i$ at the same time, only one of them will be instantiated, and their values will be the expected values, that is $v^i(\{d_i, \neg d_i\}) = \frac{1}{2}v^i(d_i) + \frac{1}{2}v^i(\neg d_i)$.¹

¹Another option is to assume the following: $v^i(\{d_i, \neg d_i\}) = -\infty$. The appropriate assumption is dependent on the use of the model.

Additionally, where \mathcal{I}_i is a set of intentions, we assert $goals(\mathcal{I}_i) = \{\varphi \mid \langle \alpha, \varphi \rangle \in \mathcal{I}_i\}$ and $actions(\mathcal{I}_i) = \{\alpha \mid \langle \alpha, \varphi \rangle \in \mathcal{I}_i\}$. Where \mathcal{B} is a set of beliefs and \mathcal{I} is a set of intentions, we say that \mathcal{I} is feasible in the context of \mathcal{B} if $\forall \langle \alpha, \varphi \rangle \in \mathcal{I}, r_{\alpha, \varphi} \in \mathcal{B}$, (i.e., the action part of every intention is believed to be executable). We write $feas(\mathcal{I}, \mathcal{B})$ to denote the largest subset of \mathcal{I} that is feasible in the context of \mathcal{B} . Note that since \mathcal{B} has a finite basis, $feas(\mathcal{I}, \mathcal{B})$ must be finite. We extend the value function v^i to a function \bar{v}^i on all subsets X of L as follows:

$$\bar{v}^i = \begin{cases} v^i(\{d \in \mathcal{D}_i \mid X \vdash d\}) & \text{if } X \not\vdash \perp \\ 0 & \text{otherwise} \end{cases}$$

Next, we defined a function $ben_i(\mathcal{I}, \mathcal{B}) \rightarrow R$, which defines the *benefit* that agent i would obtain from the set of intentions \mathcal{I} if it were the case that beliefs \mathcal{B} were correct: $ben_i(\mathcal{I}, \mathcal{B}) = \bar{v}^i(\mathcal{B} \dot{+} \wedge \{goals(feas(\mathcal{I}, \mathcal{B}))\}) - c^i(actions(\mathcal{I}) \cap Ac_i)$. So, for example: $ben_i(\mathcal{I}_i, \mathcal{B}_i)$ is the benefit that agent i would obtain from its own intentions \mathcal{I}_i under the assumption that its own beliefs, \mathcal{B}_i were correct. Similarly, $ben_i(\mathcal{I}_1 \cup \dots \cup \mathcal{I}_n, \mathcal{B}_i)$ is the benefit that agent i would obtain from the intentions of all agents in the system, under the assumption that its own beliefs, \mathcal{B}_i were correct.

C. Desires Model

Our formalization extends former BDI frameworks by providing the ability to reason whether one should be part of a social group \mathcal{A} , such that its intentions \mathcal{I}_i are adopted according to its beliefs about the states of the other members as well as their desires. Specifically, while previous BDI models only focused on cooperation while working solely with individual or joint desires, we take a new approach by differentiating between three types of desires that agent i may consider according to its relationship with some j :

1. **Cooperative Desire** (\mathcal{D}_i^{cop}): $d \in \mathcal{D}_i$ is a cooperative desire with respect to agent $j \in \mathcal{A}$, if $d \in \mathcal{D}_j$.
2. **Individual Desire** (\mathcal{D}_i^{ind}): $d \in \mathcal{D}_i$ is an individual desire with respect to agent $j \in \mathcal{A}$, $i \neq j$, if $\neg(d \in \mathcal{D}_j \vee \neg d \in \mathcal{D}_j)$.
3. **Competitive Desire** (\mathcal{D}_i^{com}): $d \in \mathcal{D}_i$ is i 's competitive desire with respect to agent $j \in \mathcal{A}$, if i believes that $(\exists_{j \neq i}) \neg d \in \mathcal{D}_j$.

Running example: Consider an academic department with three faculty members: Alice, Bob and Chen. Alice has three desires: (1) submitting a proposal. (2) winning a position. (3) writing a paper. Bob's desires are: (1) submit a proposal. (2) writing a paper. Chen's desires are: (1) winning a position. (2) submitting a proposal. Formally, the set of agents is $\{A, B, C\}$. The Desires are $\{d_A^1, d_A^2, d_A^3\}$ for Alice, $\{d_B^1, d_B^2\}$ for Bob and $\{d_C^1, d_C^2\}$ for Chen. The valuation function for Alice is $v^A(d_A^1) = 5, v^A(d_A^2) = 12, v^A(d_A^3) =$

10. The costs of the actions that brings about these desires (e.g. $\langle \alpha, d_A^1 \rangle$) are 5, 8, 3 respectively. The valuation function for Bob is $v^B(d_B^1) = 8, v^B(d_B^2) = 10$, with action costs of 6, 6 respectively. The valuation function for Chen is $v^C(d_C^1) = 8, v^C(d_C^2) = 6$ with actions costs of 5, 4.

Adding the information that the proposal desires are cooperative as all three are working jointly on the same proposal, i.e., $d_A^1 = d_B^1 = d_C^1$. In a well-coordinated interaction only a single agent will spend the cost associated with this desire. In contrast, the position appointment is a competitive desire for Alice and Chen. As such, without coordination both will spend the cost of perusing the appointment, but only one will win. The others are individual desires.

Proposition 1: These properties satisfy the desire model:

- Mutual Exclusion — $(D_i^{cop} \cap D_i^{ind}) = \emptyset$. $(D_i^{cop} \cap D_i^{com}) = \emptyset$. $(D_i^{ind} \cap D_i^{com}) = \emptyset$.
- Additivity — Let D_i^x and D_i^y be the set of desires from different types, then $v^i(D_i^x \cup D_i^y) = v^i(D_i^x) + v^i(D_i^y)$.
- Containment — $D_i^{cop} \subseteq D_i$, $D_i^{ind} \subseteq D_i$, $D_i^{com} \subseteq D_i$.
- Self Desires — $D_i = D_i^{cop} \cup D_i^{ind} \cup D_i^{com}$.

Proofs sketch: The Mutual Exclusion, that its existence can be seen from direct inspection of the definitions, provides the ability to ascribe the desire of agent i to one of the three types of desires. The Additivity property, stating that the value of the union of desires is their individual sum, is a corollary of Mutual Exclusion. The Containment property is obvious from its definition. The Self Desires property is derived using propositions Mutual Exclusion and Containment; For every $d \in D_i$, d can either be in some other agent's set, in that case $d \in D_i^{cop} \vee d \in D_i^{com}$ (depending on v_i), or it is not in some other agent's set, hence $d \in D_i^{ind}$.

From the above basic properties we can see that agent i 's set of desires (D_i) contains three types of desire relations: cooperative, individualistic and competitive desires. We define three types of action sets, where the actions in each set are part of the intentions that an agent adopts when trying to achieve the respective type of desire, as follows: $A_{C_i}^b = \{\alpha \mid \langle \alpha, \varphi \rangle \in I_i, \varphi \in D_i^b\}$, where $b \in \{cop, ind, com\}$.

III. THE SOCIAL BEHAVIOR SPECTRUM

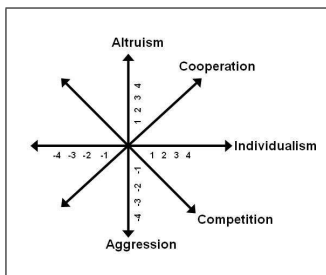


Figure 1. social behaviors spectrum

Figure 1 describes a two-person preference model of the major interpersonal orientations that can occur between

players. In this model [8], the player's utility is defined on the horizontal axis, and the outcome of the "other" player is on the vertical axis. Each outcome increases monotonically along each axis, and the values reflect a linear combination of payoffs to both players. Multiple agents will be regarded as pairwise aggregation of the two-person model.²

To allow the description of different relationships between the agents' desires, we now describe a set of behavioral axioms that correspond to the interpersonal orientations exhibited in the social behavior spectrum. The presented axioms, together with the mental state model presented later on, will constitute the design guidelines for the construction of social agents. In order to avoid describing a BDI framework from scratch, we build our presentation on a well-known framework, the *SharedPlans* [9], that provide all the operators and predicates for facilitating a joint activity.³

Next we present the main components of the *SharedPlans* formalism, but refer the reader to [9] for a complete description. The operator $\text{Int.To}(i, \alpha, T_n, T_\alpha, \mathcal{S})$ represents i 's intentions at time T_n to do action α at time T_α in the context of the BDI structure \mathcal{S} . The operator $\text{Int.Th}(i, prop, T_n, T_{prop}, \mathcal{S})$ represents i 's intention at time T_n that a certain proposition $prop$ will hold at time T_{prop} in the context of \mathcal{S} . The potential intention operators, $\text{Pot.Int.To}(\dots)$ and $\text{Pot.Int.Th}(\dots)$, are used to represent the mental state when an agent considers adopting an intention, but has not yet deliberated about its interaction with the other intentions it holds. The operator $\text{Bel}(i, f, T_f)$ indicates i believes the statement expressed by formula f at time T_f (we abuse notation with the formula f not really constituting the argument, but its name 'f'). $\text{MB}(\mathcal{A}, f, T_f)$ represents Mutual Belief of the agents in \mathcal{A} . In addition, the operator $\text{Do}(i, \alpha, T_\alpha)$ holds when i performs action α at time T_α . The following are the five major social orientations:

A1. Altruistic act axiom

$$\begin{aligned} &(\forall i, j \in \mathcal{A}, S_i, S_j, \langle \alpha, d \rangle \in I_i, d \notin D_i, T_n) \\ &[\text{Bel}(i, ML_{alt} < \text{ben}_i(I_i, B_i) < 0, T_n) \wedge \\ &\text{Bel}(i, d \in D_j, T_n) \rightarrow \text{Pot.Int.To}(i, \alpha, T_n, T_\alpha, S_i)] \end{aligned}$$

Formally, agent i may consider adopting an intention to achieve an altruistic desire with respect to j , even if i does not obtain any benefit from attaining it (hence, altruism). The axiom states that an agent i will adopt an intention which achieves $d \in D_j$ if it believes that by performing it agent j will obtain a value, and that the decrease in its own benefit, due to the additional costs, will not be larger than some ML_{alt} constant.

A2. Cooperative act axiom

$$\begin{aligned} &(\forall i, j \in \mathcal{A}, S_i, S_j, \langle \alpha, d \rangle \in I_i, d \in D_i, T_n) \\ &[[\text{Bel}(i, d \in D_j, T_n)] \wedge \end{aligned}$$

²Note that the behaviors on the left side of the graph are deliberately excluded from our work, since they are considered mental disorders (e.g. *Masochism, Sado-Masochism*).

³Nevertheless, our extension is general and can be used to extend any other model.

case 1: “ j loses from achieving d by itself”

$$[Bel(i, ben_j(I_j, B_i) < 0, T_n) \rightarrow Pot.Int.To(i, \alpha, T_n, T_\alpha, S_i)] \oplus$$

case 2: “ j gets a positive benefit from achieving d by itself”

$$[Bel(i, ben_j(I_j, B_i) \geq 0, T_n) \rightarrow Pot.Int.To(i, \alpha, T_n, T_\alpha, S_i) \vee \\ Pot.Int.Th(i, Do(j, \alpha, T_\alpha), T_n, T_{Do(\dots)}, S_i) \vee \\ Pot.Int.Th(i, Do(\{i, j\}, \alpha, T_\alpha), T_n, T_{Do(\dots)}, S_i)]$$

The formulation of the cooperative axiom is based on the theory presented in [10] that differentiate between a *task* and a *treatment* cooperative groups. Formally, an agent i may consider adopting an intention to achieve a cooperative desire with respect to j . This axiom states that if i believes that it shares the same desire with j , then there are two cases of cooperation. The first case states that i believes that j loses from completing the desire itself, then i may consider adopting an intention to achieve the desire. In the second case both agents, i and j , do not lose by performing α which achieves the desire (the α 's may be different), and there are three options. First, i may consider doing α by itself. Second, i may adopt a potential intention that α will be done by j . Third, α will be performed by i and j jointly.

A3. Individualistic act axiom

$$(\forall i, j \in \mathcal{A}, S_i, S_j, \langle \alpha, d \rangle \in I_i, d \in D_i, T_n) \\ [Bel(i, ben_i(I_i, B_i) \geq 0, T_n) \wedge \\ Bel(i, \forall_{(j \neq i)} (d \notin D_j) \wedge (\neg d \notin D_j), T_n) \\ \rightarrow Pot.Int.To(i, \alpha, T_n, T_\alpha, S_i)]$$

Formally, agent i may consider adopting an intention to achieve an individualistic desire when it believes that this desire does not belong to the other agent's set of desires.

A4. Competitive act axiom

$$(\forall i, j \in \mathcal{A}, S_i, S_j, \langle \alpha, d \rangle \in I_i, d \in D_i, T_n) \\ [Bel(i, ben_i(I_i, B_i) \geq 0, T_n) \wedge Bel(i, (\exists_{(j \neq i)}) \neg d_i \in D_j, T_n) \rightarrow \\ Pot.Int.To(i, \alpha, T_n, T_\alpha, S_i)]$$

Formally, agent i may consider adopting an intention to achieve a competitive desire with respect to j when it obtains a positive benefit from achieving the desire, even when i believes that agent j holds another desire, where when both desires will be achieved, agent i might lose.

A5. Aggressive act axiom

$$(\forall i, j \in \mathcal{A}, S_i, S_j, \langle \alpha, d \rangle \in I_i, d \notin D_i, T_n) \\ [Bel(i, ML_{agg} < ben_i(I_i, B_i) \leq 0, T_n) \wedge \\ Bel(i, \neg d \in D_j, T_n) \rightarrow Pot.Int.To(i, \alpha, T_n, T_\alpha, S_i)]$$

The axiom states that an agent i will consider adopting the intention of performing α which may decrease its benefit up to some lower bound ML_{agg} , if it believes that by achieving d j 's benefit will decrease as well.

IV. SOCIAL BEHAVIOR ACTIVITY

The Social Behavior Activity model (SBA) describes the mental states of a member in a social interaction; these possible mental states span the spectrum as presented above, whereby the agent's valuation of the desire relations affects its competitive/cooperative position on the spectrum. The SBA model defines a social behavior activity for a group of agents \mathcal{A} in the context of the BDI structure \mathcal{S} and profiles

\mathcal{P} at time T_n . We use the notation \mathcal{P} to represent the profiles of the members, and we use P_i^j to denote i 's beliefs about j 's profile. The profile of a member is a domain dependent object that captures information that is known about the agents.⁴ The operator $member(i, \mathcal{A})$ in the definition holds if i is a member of \mathcal{A} , and $max()$ is the maximum function. The Social Behavior Activity model is as follows:

$$SBA(\mathcal{S}, \mathcal{A}, \mathcal{P}, T_n)$$

- 1) \mathcal{A} mutually believes that all members are part of \mathcal{A} :
 $MB(\mathcal{A}, (\forall i \in \mathcal{A}) member(i, \mathcal{A}), T_n)$
- 2) Members of \mathcal{A} have (partial) beliefs about the profiles of the other agents:
 $(\forall i \in \mathcal{A}) Bel(i, (j \in \mathcal{A}) (\exists P_i^j \subseteq \mathcal{P}), T_n)$
- 3) \mathcal{A} mutually believes that either:
 - a) [all members of \mathcal{A} have the intention that they will attain the maximum positive difference between their and their opponent's benefits:
 $MB(\mathcal{A}, f, T_f), f = (\forall i \in \mathcal{A}) \\ Int.Th(i, max(ben_i(I_1 \cup \dots \cup I_n, B_i) - (\forall_{j \neq i}) \\ ben_j(I_1 \cup \dots \cup I_n, B_i)), T_n, T_{max(\dots)}, \mathcal{S})] \oplus$
 - b) [being a member obtains a better benefit value:
 $MB(\mathcal{A}, f, T_f), f = (\forall i \in \mathcal{A}) \\ (ben_i(I_1 \cup \dots \cup I_n, B_i) \geq ben_i(I_i, B_i), T_n) \wedge \\ \mathcal{A}$ has a mutual belief that each member has an intention that the group be maintained:
 $MB(\mathcal{A}, f, T_f), f = (\forall i \in \mathcal{A}) \\ Int.Th(i, member(i, \mathcal{A}), T_n, T_{member(\dots)}, \mathcal{S})]$

First, SBA implies the ability of the agents to identify themselves as members of some social group (Item 1) (an autonomous agent working in isolation will not be part of a social activity). Second, each individual in the group is characterized by life histories, development patterns, needs, goals, and behavior patterns. These characteristics might be known to some extent and are represented in each member's profile. Thus, the members must have beliefs concerning a partial profile of the others (Item 2). The third item represents two forms of mutual inter-group dependence. Mutual dependence (Item 3) means that the benefits of all the parties are affected to some extent by the behaviors of the other members [1]. We differentiate (using exclusive or, \oplus) between two cases of mutual dependence: competitive (SBA_a) and cooperative (SBA_b) dependence.

In a cooperative scenario (SBA_b), agent i believes that there is some positive dependency between the benefit functions (Item 3.b), thus staying with the group will yield a higher benefit value than working alone. Therefore, when this holds the agent will need to adopt a belief that all members in the group (including itself) have the intention to remain group members (Item 3.b.i), and an aggressive behavior will not be possible.⁵ On the other hand, in a competitive scenario (SBA_a), the agents believe in a negative

⁴As the profile definition is domain dependent, we refer the reader to [10] for an example of profile usage in an application.

⁵The proof is straightforward and was left out due to space constraints.

dependency between their benefits functions, which happens when there are competitive desires. In these situations, i will have the intention that the difference in benefits with each of the other members j will be maximized. The agent will act according to axioms A.2–A.5, as aggressive act might exist to increase the benefits differences. It is easy to prove that altruistic actions are not valid in competitive groups.

V. EXPLORING THE SPECTRUM

In this section we will explore the spectrum by way of utility computation which will allow an agent to compute and find out which social group will be most beneficial for him, according to its beliefs about the desires' relationships. We exemplify the various groups using the running example.

Maximum Individual Benefit:

The maximum benefit that i may obtain while being alone in the environment, $Ben_i^{ind}(I_i, B_i)$, can be defined as the values of the following set of desires:

$$\forall \varphi \in goals(I_i), \varphi \in \{D_i^{cop} \cup D_i^{ind} \cup D_i^{com}\} \quad (1)$$

The agent's total benefit as an individual when there is no other agent in the surrounding is composed of its own desires D_i where $D_i = D_i^{cop} \cup D_i^{ind} \cup D_i^{com}$ and their relevant costs. Looking back at the example we can compute the benefit values for the agents, had they were alone in the environment. Alice would have got $Ben_A^{ind}(I_A, B_A) = 0 + 4 + 7 = 11$, Bob with $Ben_B^{ind}(I_B, B_B) = 2 + 4 = 6$ and Chen with $Ben_C^{ind}(I_C, B_C) = 3 + 2 = 5$.

Maximum Benefit - Uncoordinated Social Group:

An uncoordinated social group is one in which its members act alone according to their desires, but they are situated with other agents in the same environment. As such, their beliefs on the intentions of the other agents might influence their own intentions. We also assume that agents cannot be engaged in any sort of coordination activity. Given a Social Group \mathcal{A} , agents $i, j \in \mathcal{A}, i \neq j$, the maximum total benefit of i in an Uncoordinated Social Group (USG), denoted as $Ben_i^{usg}(I_i \cup I_j, B_i)$ is composed of:

$$\begin{aligned} \forall \varphi \in goals(I_i), \varphi \in \{D_i^{cop} \cup D_i^{ind} \cup D_i^{com'}\} \wedge \\ \forall \alpha \in actions(I_i), \alpha \in \{C_i^{cop'} \cup C_i^{ind} \cup C_i^{com}\} \end{aligned} \quad (2)$$

As there are other agents in the environment, they can affect the agent in two ways: (1) positive impact - by achieving part of its desires from its set of cooperative desires ($C_i^{cop'} \subseteq C_i^{cop}$ is the subset that i performs by itself), thus decreasing i 's total cost to $c^i(C_i^{cop} \setminus C_i^{cop'})$. (2) negative impact - by achieving part of its competitive desires ($D_i^{com'} \subseteq D_i^{com}$), thus decreasing its total value from not receiving their benefits .

Proposition 2: If $(D_i^{com} = \emptyset) \vee (c^i(C_i^{cop} \setminus C_i^{cop'}) \leq v^i(D_i^{com} \setminus D_i^{com'}) - c^i(C_i^{com'}))$, then $Ben_i^{usg} \geq Ben_i^{ind}$.

Proof sketch. In the first case, since the agent has no competitive desires, the total benefit from its desires can not increase due to other agents' intentions, thus $Ben_i^{usg} = Ben_i^{ind}$. The second case is when the positive impact (when other agents achieve i 's desires) is greater or equal to the negative impact, and $Ben_i^{usg} \geq Ben_i^{ind}$.

The above proposition states that i is always better off being part of an uncoordinated social group, when it does not have competitive goals, or when it believes that the contribution of others to its cooperative desires will be greater than its loss from the competitive desires it will give. In the example we have a cooperative desire that is shared among all three agents, as such much effort is being spent without coordination. We also have a competitive desire between Alice and Chen, therefore without coordinating, the benefit will be computed as follows. $Ben_A^{usg}(I_A, B_A) = 0 + (0.5 * 4 - 0.5 * 8) + 7 = 5$, Bob with $Ben_B^{usg}(I_B, B_B) = 2 + 4 = 6$ and Chen with $Ben_C^{usg}(I_C, B_C) = (0.5 * 3 - 0.5 * 5) + 2 = 1$.

Maximum Benefit - Coordinated Social Group:

A coordinated social group is such that its members first coordinate their actions and intentions, in order to maximize their benefits. The agents may coordinate their cooperative desires as well as their competitive ones. In such cases, the agents may adopt a different set of coordinated cooperative actions, C_i^{copN} than in cases where coordination is not possible. Similarly, the agent can negotiate its competitive desires with other members to optimize its benefit, and thus agree on a partial set of competitive desires $D_i^{com''}$ (with the respective $C_i^{com''}$ set). In such cases, i can be requested to give up some weak competitive desires that decrease the benefit of other agents, and accordingly it may exchange its set of individual desires in order to achieve only part of them ($D_i^{ind''} \subseteq D_i^{ind}$ with respective $C_i^{ind''} \subseteq C_i^{ind}$ set). Given a social group \mathcal{A} , agents $i, j \in \mathcal{A}, i \neq j$, the maximum total benefit of i , as a member of a Coordinated Social Group (CSG), denoted as $Ben_i^{csg}(I_i \cup I_j, B_i)$ is composed of:

$$\begin{aligned} \forall \varphi \in goals(I_i), \varphi \in \{D_i^{cop} \cup D_i^{ind''} \cup D_i^{com''}\} \wedge \\ \forall \alpha \in actions(I_i), \alpha \in \{C_i^{copN} \cup C_i^{ind''} \cup C_i^{com''}\} \end{aligned} \quad (3)$$

Proposition 3: $Ben_i^{csg} \geq Ben_i^{usg}$, when $v^i(\{D_i^{com''}, D_i^{ind''}\}) - c^i(\{C_i^{com''}, C_i^{ind''}, C_i^{copN}\}) > v^i(\{D_i^{com'}, D_i^{ind}\}) - c^i(\{C_i^{com'}, C_i^{ind}, C_i^{cop}\})$

Proof sketch. Trivial by simple arithmetic manipulations.

Proposition (3) justifies the need for strong negotiation skills in agents, as they may attain higher benefits from coordinating their actions. Going back to the example in this case (and the two that follow) we cannot know exactly which of the possible solutions will be agreed upon by the agents. Nevertheless, we can assume that an optimal solution will not suddenly emerge due to computation limitation (we discuss this problem in depth in section VI). One solution that might emerge is that Alice will drop its

intention to be elected and will be the one who submits the proposal. As such, $Ben_A^{csg}(I_A, B_A) = 0 + 0 + 7 = 5$, Bob with $Ben_B^{csg}(I_B, B_B) = 8 + 4 = 12$ and Chen with $Ben_C^{csg}(I_C, B_C) = 3 + 6 = 9$. Two things to note about the coordinated solution: (1) all players are in not worse situation than before the coordination: $7 > 5$, $9 > 6$ and $9 > 1$. (2) The social welfare has doubled from 12 to 28.

Maximum Benefit - Cooperative Social Group:

When i becomes a member in a Cooperative Social Group (CopSG), its total benefit will be composed of the coordinated social group components, together with the costs of the altruistic actions it contributed, C_i^{alt} (which decreases its own benefit), and the altruistic actions that other agents performed to complete some of his desires, that will be denoted as D_j^{alt} (increasing its benefit). Given a social group \mathcal{A} , agents $i, j \in \mathcal{A}, i \neq j$, the maximum total benefit of i as a Cooperative Social Group (CopSG) member, denoted as $Ben_i^{CopSG}(I_i \cup I_j, B_i)$ is composed of:

$$v^i(D_i^{cop} \cup D_i^{ind''} \cup D_i^{com''} \cup_{\forall j} D_j^{alt}) - c^i(C_i^{copN} \cup C_i^{ind''} \cup C_i^{com''} \cup C_i^{alt}) \quad (4)$$

As we noticed in the previous group type, when the agent is part of a coordinated group, proposition 3 holds (otherwise a rational agent would leave the group). The next proposition states the basic condition in which the agent would prefer to “upgrade” the coordinated group to a cooperative one.

Proposition 4: Given agent i , if $Ben_i^{csg} \geq Ben_i^{usg}$, then $Ben_i^{CopSG} \geq Ben_i^{csg}$, when $v^i(D_i^{alt}) - c^i(C_i^{alt}) \geq 0$
Proof sketch. By aligning both equations and canceling out $D_i^{ind''}, D_i^{cop}, D_i^{com''}$ terms with their respective actions.

The main condition in the above proposition is that the coordinated social group achieves a higher benefit for the agent, than the uncoordinated group. Otherwise, the question of whether to join a cooperative social group is not solely related to altruistic actions.

Corollary 1: Given agents $i, j \in \mathcal{A}$, if $(Ben_i^{CopSG} \geq Ben_i^{csg}) \wedge (Ben_j^{CopSG} < Ben_j^{csg})$, then $ML_{alt} = Ben_i^{CopSG} - Ben_i^{csg}$

In order for the social group to be cooperative, its group members must have a mutual belief that being part of the group will yield higher benefits for them (item 3.b). Therefore, corollary 1 states that when agent i wants to be in a cooperative social group with another agent j , but it believes that it is not beneficial for j to be in the group (as $Ben_j^{CopSG} < Ben_j^{csg}$), it will be willing to contribute altruistic desires with respect to j to “convince” it to be part of the group. However, this contribution will be bounded by ML_{alt} to prevent i from eventually losing.

Maximum Benefit - Competitive Social Group:

In a similar way, when i is a member of a competitive social group (ComSG), its total benefit will comprise the sum of its

own aggressive actions, together with the aggressive actions taken by j , and the benefit gained through cooperative desires that are performed by itself and other members. In addition, it will consider achieving only cooperative desires that will contribute more value to himself than to other agents ($D_i^{copN'}$). i 's total benefit also includes the set of its individual desires and the subset of competitive desires $D_i^{comN'}$. In this case, the main goal of i will be to maximize its benefit with respect to each j in its competitive group. Thus, $D_i^{comN'}$ includes only the competitive desires that will cause a greater decrease in the benefit to others, thus some of the competitive desires in its original D_i^{com} will not be part of $D_i^{comN'}$. Given a social group \mathcal{A} , agents $i, j \in \mathcal{A}, i \neq j$, the maximum total benefit of i as a Competitive Social Group (ComSG) member, denoted as $Ben_i^{ComSG}(I_i \cup I_j, B_i)$ is composed of:

$$v^i(\{D_i^{copN'} \cup D_i^{ind} \cup D_i^{comN'}\} \setminus \cup_{\forall j} D_j^{agg}) - c^i(C_i^{copN'} \cup C_i^{ind} \cup C_i^{comN'} \cup C_i^{agg}) \quad (5)$$

When examining the competitive social group, we can see that the situation worsens in terms of the agent's total benefit as the benefit from the individual's set of desires decreases, and new aggressive desires are added. Thus, we can conclude the following:

Corollary 2: If $(Ben_i^{ComSG} > Ben_i^{CopSG})$ then $\exists MD \in D_i^{com}$ s.t. $d = \max(ben_i(I_1 \cup \dots \cup I_n, B_i) - (\forall j \in \mathcal{A}, j \neq i) ben_j(I_1 \cup \dots \cup I_n, B_i))$.

In a competitive group, both $D_i^{comN'}$ and $D_i^{copN'}$ are subsets, thus receive lower or equal values. In addition, $\cup_{\forall j} D_j^{agg}$ are desires that further lower A_i 's benefit (by definition). As the total benefit is lower, only a “meta” set of desires (MD) on the difference of the benefits values themselves can attain a higher total benefit.

VI. TO MEMBER OR NOT TO MEMBER?

Following the presentation of the *SBA* model, it appears that a key ingredient is still missing in order to be able to practically use the model for architectural purposes. The missing ingredient appear in the *SBA*'s model in item 3.b under the proposition $ben_i(I_i \cup \dots \cup I_n, B_i) \geq ben_i(I_i, B_i)$, that is a main query behind the question: to member or not to member (a cooperative group)? The following theorem shows that it impractical to compute an optimal answer.

Theorem 1: The individual agent's task of computing whether to join a cooperative social group is NP-Complete. **Proof sketch:** The reduction is to the Knapsack problem, where the set of n agents with individual payoff functions p_i try to fill the knapsack with a subset of the k intentions, without any weight constraints.

Specifically, we can use the above problem as an instance of the 0-1 knapsack problem by trying to optimize the vector $v = \{x_1, x_2, \dots, x_k\}$, where $x_i = 1$ if intention i is in the selected set, 0 otherwise. Instead of having a single p_i vector, we have n such vectors and each agent's

original and individual set of intentions will be denoted as $v' = \{x_1, x_2, \dots, x_k\}$, where $x_i = 1$ if intention i is in the agent's own individual set of intentions, 0 otherwise. The constraints are as follows: (1) $\sum_{i=1}^n vp_i \geq \sum_{i=1}^n v'p_i$ (higher social welfare). (2) $(\forall i)vp_i \geq v'p_i - \epsilon$ (each individual is no worse than what he could have before minus some constant ϵ).

However, as humans often are able to use heuristics to approximate solutions, we might as well not aspire for an optimal solution, but settle on some approximated solution that bounded rational agents can get rather promptly, as a function of the available computational resources, and the requested precision.

A. Membership as a CSP

In order to empirically evaluate the practicality of the membership query in real applications that the SBA model is directed to, we formulated the membership problem as a boolean constraints satisfaction problem: The set of agents, each with its own set of desires, for a total of N desires. Each of the N desires will be formulated as an individual node, that can get the value of 0 or 1 which signifies whether this desire will be fulfilled. Edges between nodes will model relationships between desires. The division to the cooperative, individualistic or competitive groups of desires, will be modeled as follows: (1) Cooperative desires - will be modeled with "or" constraint edges connecting the nodes. (2) Individual desires - are simply the nodes themselves. (3) Competitive desires - will be modeled with "exclusive or" constraint edge. A solution will be an assignment of values 1 or 0 to nodes s.t. the following constraints hold:

- 1) The local desires constraints are not violated.
- 2) Individual Benefit constraint: $(\forall i)vp_i \geq v'p_i - \epsilon$.
- 3) Social Welfare constraint: $\sum_{i=1}^n vp_i \geq \sum_{i=1}^n v'p_i$.

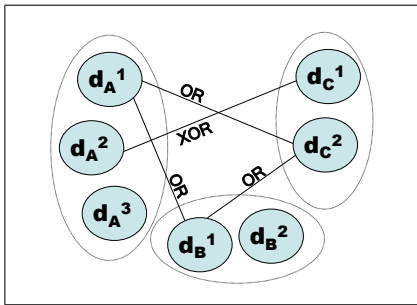


Figure 2. Example graph representation

Running example: According to the previous example Alice has three desires $\{d_A^1, d_A^2, d_A^3\}$, Bob's desires are $\{d_B^1, d_B^2\}$, and Chen's desires are $\{d_C^1, d_C^2\}$. The submit proposal desires are cooperative $\{d_A^1, d_B^1, d_C^2\}$ desire as all three are working jointly on the same proposal. The position appointment $\{d_A^2, d_C^1\}$ are competitive desires for Alice and Chen. The others are individual desires without any

relationship among them. Let's further assume the following costs/value functions to each of them: Alice costs $[5, 8, 3]$, values $[5, 12, 10]$, Bob costs $[6, 6]$ and values $[8, 10]$ and Chen costs $[5, 4]$ and values $[8, 6]$. We can compute the Maximal Uncoordinated Benefit and get the following:⁶ $Ben_A^{ucg} = 5$, $Ben_B^{ucg} = 6$, and $Ben_C^{ucg} = 1$, and the total social welfare, $\sum_{i=1}^n vp_i = 12$. A graph representation of the problem is depicted in figure 2. One solution for the above problem is $v' = [d_A^1, \neg d_A^2, d_A^3, \neg d_B^1, d_B^2, d_C^1, \neg d_C^2]$, which results in respective individual benefits of: 7 for Alice, 12 for Bob, and 9 for Chen. The social welfare also increases to $\sum_{i=1}^n v'p_i = 28$.

B. Heuristic Functions

The above solution is one of many existing solutions, and the small size of the problem (7 desires) constitutes a relatively easy problem to solve using brute-force search. However, for larger problems, bounded resources automated agents will need some heuristics to guide their search for a solution. As such, we suggest the following heuristic functions that are applicable to situations in which the agents commit their information to a third party (e.g. mediator or oracle) and get a solution that dictates which desires should be perused by the agents:

- H1 — smallest number of desires.
- H2 — largest number of competitive desires.
- H3 — smallest number of cooperative desires.
- H4 — maximize $H1 - H2 + H3$.
- H5 — largest Ben_i^{ucg} value.

In order to evaluate the usefulness of the suggested heuristics in guiding the search process, we used the ECLiPSe Constraint Logic Programming System. Our aim was to evaluate the different functions in 3 types of environments: cooperative, individualists, and competitive. The environments' names describe them in terms of the proportions of the respective desires. Let us define $pc + pi + pm = 1$ to be a the probabilities distributions that characterizes an environment in terms of the probability of a single desire to be cooperative (pc), individualistic (pi) or competitive (pm). An individualist environment is one in which, on average, the majority of the desires relations are individualistic. Specifically, we set the probabilities to be $pc = 0.15$, $pm = 0.15$ and $pi = 0.7$, in individualistic environment, and the cooperative/competitive environments were set with respective probabilities of $pc = 0.3$ or $pm = 0.3$. We created 25 random instances to each of the environments using the following set of parameters: the number of agents was in the $[4, 6]$ range, the number of desires per agent was in the $[2, 7]$ range (therefore the problem sizes ranged from 8 to 42 nodes). The costs/values numbers were in the $[1, 9]$ range. Each problem instance was solved using each of the suggested heuristics, and the solution time was recorded.

⁶Recall that we assume all parties have an equal chance of attaining competitive desires, and compute the expected benefit accordingly.

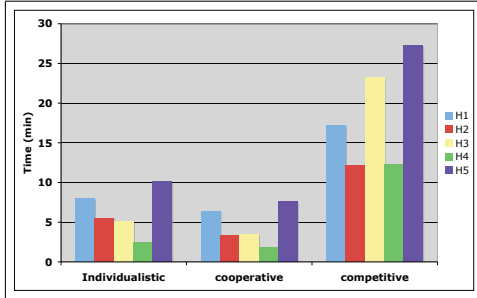


Figure 3. Simulations results

The results are depicted in figure 3, provide several insights on the behavior of the functions and the effect of the environment's characteristics on the solution time. First, it is easy to see competitive environments proved to be more challenging than the other environments. Second, H4 provide the best results in the first two environments and is equally effective to H2 in the competitive environment. Finally, not surprisingly, a larger number of cooperative desires results in easier problem instances in terms of solution time.

VII. RELATED WORK

The BDI literature is composed of work on formal logical languages [11] and practical solutions in the form of programming languages [12]. From the BDI literature perspective, there are works that looked at specific parts of the spectrum (e.g., [13], [10]), but we are not aware of any work that spans the entire social behavior spectrum. Other related work include Montagna et al. [11] that did consider classification to "friends" and "enemies" in the formal language, however these were regarded as predefined labels and were not connected to the agents' desires.

While the above-mentioned works and others bear some similarities to our work, we note two main differences. First, as the goal of our work is to be used as design guidelines for social competent agents, our choice of using the SharedPlans formalism [9] as the underlying language can be easily revised to most formal models found in the literature (e.g. [11], [13]), after augmenting the extended Desire model into them. Second, these works do not consider the case where agents might have conflicting *Desires*, which might lead to competitive or adversarial actions towards other agents. To the best of our knowledge, the idea of the complex relationship between the desires and their impact of the full range of social interpersonal relationship is new.

VIII. CONCLUSIONS

We presented the full social behavior spectrum as it has been modeled in social science research, and suggested the multiagent *Social Behavior Activity* model as a formal theory that describes agents' mental states and behavioral axioms across a range of behavioral possibilities. We explored the spectrum by providing a computational description of

various social groups and the maximum achievable benefit for an agent. The findings of our social spectrum exploration show that usually it is worthwhile for an agent to be in some sort of cooperative or coordinated social group (depending on its negotiation skills), while in some cases it is also worthwhile taking altruistic actions in order to keep the social group intact. At the same time, being in a competitive social group is the worst option in most cases, unless there is a strict competition. We then presented a way to tackle the "group membership" problem by remodeling the problem as a constraints satisfaction problem, and evaluated several heuristic functions to guide the search process on large problems. Regarding future research, we plan to implement the architecture in a simplified environment in which agents can be cooperative and competitive at different times.

REFERENCES

- [1] V. J. Derlega and J. Grzelak, *Cooperation and Helping Behavior Theoreis and Research*. Academic Press, 1982.
- [2] P. Stone, G. A. Kaminka, S. Kraus, and J. S. Rosenschein, "Ad hoc autonomous agent teams: Collaboration without pre-coordination," in *AAAI*, 2010.
- [3] M. E. Bratman, *Intention, Plans and Practical Reason*. Harvard University Press, Massachusetts, 1987.
- [4] M. P. Georgeff, "Decision-making in an embedded reasoning system," in *IJCAI*, 1989, pp. 972–978.
- [5] J. Grant, S. Kraus, and D. Perlis, "A logic based model of intention formation and action for multi-agent subcontracting," *AIJ*, vol. 163, no. 2, pp. 163–201, 2005.
- [6] C. E. Alchourròn, P. Gärdenfors, and D. Makinson, "On the logic of theory change: Partial meet contraction and revision functions," *Symbolic Logic*, vol. 50, pp. 510–530, 1985.
- [7] M. E. Pollack, "The uses of plans," *Artificial Intelligence*, vol. 57, no. 1, pp. 43–68, 1992.
- [8] D. Griesinger and J. Livingston, "Towards a model of interpersonal motivation in experimental games," *Behavioral Science*, vol. 18, pp. 173–188, 1973.
- [9] B. J. Grosz and S. Kraus, "Collaborative plans for complex group action," *Artificial Intelligence*, vol. 86, no. 2, pp. 269–357, 1996.
- [10] M. Hadad, G. Armon-kest, G. A. Kaminka, and S. Kraus, "Supporting collaborative activity," in *AAAI*, 2005, pp. 83–88.
- [11] R. Montagna, G. Delzanno, M. Martelli, and V. Mascardi, "Bdiatl : An alternating-time bdi logic for multiagent systems," in *EUMAS*, 2005, pp. 214–223.
- [12] A. S. Rao, "Agentspeak(l): Bdi agents speak out in a logical computable language," in *MAAMAW*, 1996, pp. 42–55.
- [13] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes, "On acting together," in *AAAI*, 1990, pp. 94–99.