A BDI-based Agent Architecture for Social Competent Agents

Inon Zuckerman\textsuperscript{a,*}, Meirav Hadad\textsuperscript{b}

\textsuperscript{a} Department of Industrial Engineering and Management, Ariel University, Ariel, 4070000, Israel
E-mail: inonzu@ariel.ac.il

\textsuperscript{b} Department of Computer Science, Bar Ilan University, Ramat-Gan, 5290002, Israel
E-mail: hadadm@biu.ac.il

Abstract.
In this work we suggest a Belief-Desires-Intentions mental model that provides agents with the social competence to capture and reason about their goals with respect to the goals of other agents/humans in the environment. The suggested architectural model would enable the implementation of generic social competent agents that would interact differently towards different groups.

We explore the agent’s behavior on the social spectrum by computationally describing the maximum attainable benefit when it belongs to different types of social groups. In addition, as the mental model requires the agent to have an 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 evaluate possible heuristics to speed-up the search.

Keywords: Social agents, Agent architecture, Mental model

1. 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 \[9\], 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 main goal of the model is to assist software engineers with 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 \[27\]), thus it must consider how its actions affects other group members. Classifying different types of desires and behaviors is an important step in allowing it to do so.

Our model builds upon Bratman’s philosophical foundations of the Belief-Desire-Intention (BDI) model \[4\] that was developed to explain human practical reasoning. The model was proved to be highly applicable as a software model for developing bounded rational intelligent agents (e.g. \[12\]). However, while previous adaptation of the model described only specific parts of the social behavior spectrum (in particular joint and individual desires), in the first part of this work we aim to provide a formal model that spans the entire social spectrum, including relations that have not yet received any attention.

To do so we first extend the common Desires model according to their subjective perception of the cooperative/competitive nature of the society.

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 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. The main goal of this analysis is to provide agents with the capabilities to reason about their different goals and to compute whether they should coordinate with the others, or form cooperative/competitive groups with them. Our analysis will define the different social groups and present a way to reason on them with the suggested mental architecture. In addition, as the mental model and social groups analysis required the agent to have any ability to reason about group membership, which we prove to be NP-complete, we also present a way to formulate the problem as a constraints satisfaction problem and present and evaluate possible heuristics to speed-up the search.

To summarize our main contributions can be seen as the following:

- The presentation of the Social Behavioral Activity (SBA) mental model for the construction of social competent agents. This mental model is an architectural framework that will allow engineers to create agents that have some reasoning capabilities about their social surroundings.
- An analytical exploration of different social groups, which will allow an agent to decide computationally which group to member (if any). As the above item is computationally expensive, we also present a reduction to the constraints satisfaction problem, present and evaluate several heuristics to the group membership problem.
- Following our presentation of the SBA model, we will explore its relationship to prior BDI models: SharedPlans (SP) [17], SharedActivity (SA) [18] and AdversarialActivity (AA) [38].

2. Related Works

The social oriented multi-agent literature includes domain-specific algorithms that provide some adaptability to agents with different personalities or social behaviors. For example, in [21] the authors used a joint social welfare function and allowed agents to change the weight of their social tendency level. In [29] the authors used the Colored Trails game to study a model in which agents’ helpfulness is characterized in terms of cooperation and reliability. Another interesting approach in the form of a dependency graph that is used to describe and capture the relationships inside a multi-agent system environment was presented in [26], while another line of work that explores agents’ social interactions comes from agent-based simulation approach [8]. While these work provide some reasoning capabilities about the agents or the emergent dynamics of the multi-agent system, they do not consider the architectural aspect of agent design within the BDI framework, that is the main focus of this work.

Another related line of work is on group beliefs. These are usually defined as collectively intentional attitudes that are based on what the group members accept as the group’s belief [19,32,13]. Group beliefs, as oppose to individual beliefs (or mutual beliefs in our context), are attributed only to the collective. The fact that all the group members believe that A is not sufficient (or necessary) for a group belief in A. It is required for group belief that the group members take A to be true when they are acting in the group context, that is, that the individuals accept A when they are acting as group members. In our case we are dealing solely with individual beliefs (and mutual beliefs which are based on those) and the formation of different social groups does not require the adaptation of group beliefs to become members in that group.

Other works explore the notion of norms as a regulatory mechanism to agents’ behavior. However, norm-regulated environments may experience problems when norms associated with their agents are in conflict. For instance, actions that are forbidden in one social context, may, at the same time, also be permitted in another context. There are works that try to cope with these inconsistencies by presenting various mechanisms to detect and solve these contradictions (e.g. [35]). Nevertheless, we are not aware of any work in which these inconsistencies affect the agent decision regarding group membership.

BDI stands for Beliefs-Desires-Intentions. It is a model of human planning and decision making proce-
The suggested framework in this work supports the full range of social interactions, from complete cooperation, through individualism, to zero-sum conflict. As such, while we have selected to use the SharedPlans formalism [17] as the underlying BDI based language, it is possible to use most of the above formal models [34,23,22] and augment the extended Desire model onto them.

3.1. Basic Definitions

Our BDI formalism is derived from the one proposed in [14], and [15]. We have a basic propositional language \( \mathcal{L} \) which is used by the agent to express beliefs about the world. We will simply assume that it contains the usual classical logic connectives \((\wedge, \lor, \neg, \rightarrow, \leftrightarrow)\), and their classical interpretation, together with the truth and falsity \((\top, \bot)\) constants, and the semantic entailment operator \((\vdash)\). \( \mathcal{F} \) is the set of sentences of this language. We assume that the belief change operations \(\vdash\) (revision) and \(\dashv\) (contraction) have been defined for \( \mathcal{L} \) [1].

Next, we assume a set \( \mathcal{A} = \{1, 2, \ldots, n\} \) of agents. For each agent \( i \in \mathcal{A} \), we assume a finite set of actions, \( \mathcal{Ac}_i = \{\alpha_1, \alpha_2, \ldots\} \). The set \( \mathcal{Ac}_i \) 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 \( R_i = \{\langle\alpha, \varphi\rangle | \alpha \in \mathcal{Ac}_i \text{ and } \varphi \in \mathcal{L}\} \), such that action \( \alpha \) is executable if its preconditions are satisfied and the termination of \( \alpha \)'s performance results in \( \varphi \) being true (see, e.g., [24] for discussion on the notion of “plan as recipe”). For every recipe \( r = \langle\alpha, \varphi\rangle \), we assume there is a proposition \( r_{\alpha, \varphi} \in \mathcal{L} \). Intuitively, \( r_{\alpha, \varphi} \in \mathcal{L} \) will be used to mean that: (1) the action \( \alpha \) is executable, in that its preconditions are currently satisfied, and (2) the performance of \( \alpha \) terminates and makes \( \varphi \) 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 \( \varphi \) is an atomic proposition.

3.2. 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: \( S_i = (\mathcal{B}_i, \mathcal{D}_i, \mathcal{I}_i, v^i, c^i) \).

- **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 description of a finite set of sentences. Formally, \( \mathcal{B}_i = \{b \in \mathcal{L} | D_i^b \vdash b\} \), where
$B_i^0 \subset L$ is a finite set that represents the basis set of beliefs for $i$.

- **Desires** — $D = D_1 \cup D_2 \ldots \cup D_n$, where $D_i \subset L$ represents the set of desires for agent $i$. We will use $d_i^1, \ldots, d_i^n$ as elements in $D_i$. We assert that desires will stay consistent by the belief revision operators.

- **Intentions** — $I_i$ stands for the intentions of the agent $i$, where $I_i \subseteq R_i$.

- **Valuation** — Defined as $v^i : 2^{(D_i)} \rightarrow R^+$ where $R^+$ is the set of positive real numbers.

- **Cost** — Defined for a finite set of actions $c^i : 2^{(A_i)} \rightarrow R^+$, which must satisfy the following condition: if $K \subseteq K' \subseteq A_i$ then $c^i(K') \geq c^i(K)$.

We require that $v^i$ satisfies the following “monotonicity” condition: for $T_i \subseteq D_i$, then $v^i(T_i) \leq v^i(D_i)$, 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 $v^i(\{d_i, \neg d_i\}) = \frac{1}{2}v^i(d_i) + \frac{1}{2}v^i(\neg d_i)$. We extend the value function $v^i$ to a function $\bar{v}^i$ on all subsets $X$ of $L$ as follows:

$$\bar{v}^i = \begin{cases} v^i(\{d \in D_i | X \vdash d\}) & \text{if } X \not\vdash \bot \\ 0 & \text{otherwise} \end{cases}$$

Additionally, where $I_i$ is a set of intentions, we assert $goals(I_i) = \{ \varphi < \alpha, \varphi > \in \varphi \}$ and $actions(I_i) = \{ \alpha < \alpha, \varphi > \in \alpha \}$. Where $B$ is a set of beliefs and $I$ is a set of intentions, we say that $I$ is feasible in the context of $B$ if $\forall (\alpha, \varphi) \in I, r_{\alpha, \varphi} \in B$, (i.e., the action part of every intention is believed to be executable). We write $feas(I, B)$ to denote the largest subset of $I$ that is feasible in the context of $B$. Note that since $B$ has a finite basis, $feas(I, B)$ must be finite.

Next, we defined a function $Ben_i(I, B) \rightarrow R$, the benefit that agent $i$ would obtain from the set of intentions $I_i$ if it were the case that beliefs $B$ were correct: $Ben_i(I, B) = \bar{v}^i(B + \{ goals feas(I, B) \}) - c^i(actions(I) \cap A_i)$. So, for example: $Ben_i(I_1, B_1)$ is the benefit that agent $i$ would obtain from its own intentions $I_i$ under the assumption that its own beliefs, $B_i$, were correct. Similarly, $Ben_i(I_1 \cup \cdots \cup I_n, B_n)$ is the benefit that agent $i$ would obtain from the intentions of all agents in the system, under the assumption that its own beliefs, $B_i$, were correct.

### 3.3 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 $I_i$ 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** ($D_i^{coop}$): $d \in D_i$ is $i$’s cooperative desire with respect to agent $j \in \mathcal{A}$, if $\exists j \neq i d \in D_j$.

2. **Individual Desire** ($D_i^{ind}$): $d \in D_i$ is an individual desire with respect to agent $j \in \mathcal{A}$, $i \neq j$, if $\neg (d \in D_j \lor \neg d \in D_j)$.

3. **Competitive Desire** ($D_i^{comp}$): $d \in D_i$ is $i$’s competitive desire with respect to agent $j \in \mathcal{A}$, if $\exists j \neq i \neg d \in D_j$.

**Running example:** Consider an academic department with three faculty members: Alice, Bob and Chen. Bob is already in a tenure track, while Alice and Chen have temporary positions. In addition, there is one newly available tenured position. Alice has three desires: (1) submitting a proposal. (2) winning an open tenured position in the department. (3) writing a paper. Bob’s desires are: (1) submit a proposal. (2) writing a paper. Chen’s desires are: (1) winning a tenured position in the department. (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.

\footnote{Another option is to assume $v^i(\{d_i, \neg d_i\}) = -\infty$. The appropriate assumption is dependent on the use of the model.}
Adding the information that the proposal desires are cooperative as all three are working jointly on the same proposal, i.e., $d^A_1 = d^B_2 = d^C_3$. 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.

The definitions of the different relations between goals yield the following:

**Proposition 1** These properties satisfy the desire model:

- **Mutual Exclusion (with respect to a single agent)**
  
  $$(D^\text{cop}_i \cap D^\text{ind}_i) = \emptyset, \quad (D^\text{cop}_i \cap D^\text{com}_i) = \emptyset, \quad (D^\text{ind}_i \cap D^\text{com}_i) = \emptyset.$$  

- **Additivity** — Let $D^\text{cop}_i$ and $D^\text{v}_j$ be the set of desires from different types, then $v^i(D^\text{cop}_i \cup D^\text{v}_j) = v^i(D^\text{cop}_i) + v^i(D^\text{v}_j)$.

- **Containment** — $D^\text{cop}_i \subseteq D_i$, $D^\text{ind}_i \subseteq D_i$, $D^\text{com}_i \subseteq D_i$.

- **Self Desires** — $D_i = D^\text{cop}_i \cup D^\text{ind}_i \cup D^\text{com}_i$.

**Proofs:** 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. Mutual exclusion is with respect to a specific observable agent $j$. That is, agent’s $i$ desire that is classified as a cooperative desire with respect to agent $j$, cannot be classified as an individualistic or a competitive desire towards $j$. 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 $d \in D^\text{cop}_i \lor d \in D^\text{com}_i$ (depending on $v_i$), or it is not in some other agent’s set, hence $d \in D^\text{ind}_i$.

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: $A^b_{Cl} = \{\alpha | \alpha < \alpha, \varphi \in I_i, \varphi \in D^b_i\}$, where $b \in \{\text{cop, ind, com}\}$.

### 4. The Social Behavior Spectrum

Figure 1 describes a two-person preference model of the major interpersonal orientations that can occur between players. In this model [16], 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.3

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 [17], that provide all the operators and predicates for facilitating a joint activity.4

Next we present the main components of the SharedPlans formalism, but refer the reader to [17] for a complete description. The operator Int.To($i, \alpha, T_n, T_\alpha, S$) represents $i$’s intentions at time $T_n$ to do action $\alpha$ at time $T_\alpha$, in the context of the BDI structure $S$. The operator Int.Th($i, \text{prop}, T_n, T_\text{prop}, S$) represents $i$’s intention at time $T_n$ that a certain proposition $\text{prop}$ will hold at time $T_\text{prop}$ in the context of $S$. The potential in-

3 Note that the behaviors on the left side of the graph are deliberately excluded from our work, since besides behaviors such as sacrifice or gifts, in which ones outcome is reduced, most of them are considered mental disorders (e.g. Masochism, Sado-Masochism).

4 Nevertheless, our extension is general and can be used to extend any other model.
tention operators, \(\text{Pot.Int.To}(\ldots)\) and \(\text{Pot.Int.Th}(\ldots)\), 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 by the formula \(f\) not really constituting the argument, but its name ‘\(f\)’). \(\text{MB}(A, f, T_f)\) represents Mutual Belief of the agents in \(A\). In addition, the operator \(\text{Do}(i, \alpha, T_\alpha)\) holds when \(i\) performs action \(\alpha\) at time \(T_\alpha\). The following are the five major social orientations:

A1. Altruistic act axiom — In its purest form, an altruistic act is an action that improves the condition of the other, while not improving (and possibly worsening) your own condition. It appears to contradict the rational behavior assumption at the heart of classical economic theories. While rational behavior can lead to various problems when there is conflict between an individualist’s own utility value. This kind of treatment does not consider these implicit benefits. The formulation of the cooperative axiom is based on the theory presented in [18] 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 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 \(\alpha\) which achieves the desire (the \(\alpha\)’s may be different), and there are three options. First, \(i\) may consider doing \(\alpha\) by itself. Second, \(i\) may adopt a potential intention that \(\alpha\) will be done by \(j\). Third, \(\alpha\) will be performed by \(i\) and \(j\) jointly.

A3. Individualistic act axiom — Individualism is the case of showing no interest in the actions and utilities of others, while conducting actions that will maximize the individualist’s own utility value. This kind of social behavior is considered rational behavior in economic.

A2. Cooperative act axiom — In a cooperative act, both parties share some effort to gain some benefit. In a pure cooperative scenario (such as the \(\text{SharedPlan}\) model [17]), the parties are members of the same group and attain part of their benefit from the group’s mutual utility function. In these situations, the parties have the incentive to use all of their resources in order to complete the joint goal. No member has any incentive to “save” some of its resources because the group’s utility is the main utility function.

\[
\begin{align*}
\forall i, j \in A, S_i, S_j, \langle \alpha, d \rangle \notin I_i, d \notin D_j, T_n) \\
[\text{Bel}(i, M_L_{alt} < \text{ben}_i(I_i, B_i) < 0, T_n) \land \text{Bel}(i, d \in D_j, T_n) \rightarrow \text{Pot.Int.To}(i, \alpha, T_\alpha, S_i)]
\end{align*}
\]

It is important to note that as researchers still debate on the nature of altruistic behavior and whether they do provide some implicit benefit to its actuator, our definition does not consider these implicit benefits.

\[
\begin{align*}
\forall i, j \in A, S_i, S_j, \langle \alpha, d \rangle \notin I_i, d \notin D_j, T_n) \\
[\text{Bel}(i, \text{ben}_i(I_i, B_i) \geq 0, T_n) \\
\rightarrow \text{Pot.Int.To}(i, \alpha, T_\alpha, S_i)]
\end{align*}
\]

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 — There are situations in which increasing one’s utility will decrease the other’s, thus they are competitive interactions. The middle point of these interactions is pure competition, which occurs in zero-sum interactions. In these competitions, each party wants to complete a goal that directly conflicts with the goal of the other, and one’s gain is exactly the other’s loss. The competitive range
also includes non-zero-sum encounters, in which the parties’ aggregate gains and losses can be more or less than zero (reflected in the ranges above and below the competition line).

\( \forall i, j \in A, S_i, S_j, (\alpha, d) \notin I_i, d \notin D_i, T_n \) \[ Bel(i, \text{ben}_i(I_i, B_i) \geq 0, T_n) \land Bel(i, (\exists j \neq i) \neg d_i \in D_j, T_n) \rightarrow \text{Pot.Int.To}(i, \alpha, T_n, T_\alpha, S_i) \]

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** — In aggressive behavior, the agent’s only goal is to hurt the other party without trying to increase its own benefit. Such behavior can occur due to personal social and psychological circumstances. E.g. a player with no chance to win in an encounter might decide to focus his efforts on preventing one of the other competitors from winning, due to personal intrigues against him. This player might gain some benefit from its “vengeful” behavior, but no benefit will be gained with respect to the original goal.

\( \forall i, j \in A, S_i, S_j, (\alpha, d) \notin I_i, d \notin D_i, T_n \) \[ Bel(i, \text{M}L_{\alpha, agg} < \text{ben}_i(I_i, B_i) \leq 0, T_n) \land Bel(i, \neg d \in D_j, T_n) \rightarrow \text{Pot.Int.To}(i, \alpha, T_n, T_\alpha, S_i) \]

Formally, agent \( i \) may consider adopting an intention to achieve an aggressive desire with respect to \( j \), if it loses from achieving \( d \), as long as member \( j \) also loses. The axiom states that an agent \( i \) will consider adopting the intention of performing \( \alpha \) which may decrease its benefit up to some lower bound \( M_{L_{\alpha, agg}} \), if it believes that by achieving \( d \) \( j \)'s benefit will decrease as well, as \( \neg d \) is in his set of desires.

### 5. 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 \( A \) in the context of the BDI structure \( S \) and profiles \( P \) at time \( T_n \). We use the notation \( P \) to represent the profiles of the members, and we use \( P_i \) 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. The operator \( \text{member}(i, A) \) in the definition holds if \( i \) is a member of \( A \), and \( \max() \) is the maximum function. The Social Behavior Activity model is as follows:

\[ \text{SBA}(S, A, P, T_n) \]

1. \( A \) mutually believes that all members are part of \( A \):
   \[ MB(A, (\forall i \in A) \text{member}(i, A), T_n) \]
2. Members of \( A \) have (partial) beliefs about the profiles of the other agents:
   \[ (\forall i \in A) Bel(i, (j \in A)(\exists P_i \subseteq P), T_n) \]
3. \( A \) mutually believes that either:
   
   (a) [all members of \( A \) have the intention that they will attain the maximum positive difference between their and their opponent’s benefits:]
   \[ MB(A, f, T_f), f = (\forall i \in A) \text{Int.Th}(i, \max(\text{ben}_i(I_i \cup \cdots \cup I_n, B_i) - (\forall j \neq i) \text{ben}_j(I_j \cup \cdots \cup I_n, B_i), T_n, T_{\max}(\_ \_ \_ \_ \_)), T_n) \]
   
   (b) [being a member obtains a better value:
   \[ MB(A, f, T_f), f = (\forall i \in A) \text{Int.Th}(i, \text{member}(i, A), T_n, T_{\text{member}(\_ \_ \_ \_ \_)}, T_n) \]

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 profiles may be given explicitly or implicitly (e.g., learning the profile by observation, overhearing, etc.), and their exact structure is domain dependent. The third item represents two forms of mutual inter-group dependence. Dependence refers to the relation in which the benefit that one member obtains from its own behaviors is affected at least partly by the activities of another party. Mutual dependence (Item 3) means that the benefits of all the parties are affected to some extent by the behaviors of the other members [9]. We differentiate (using exclusive or, \( \oplus \)) between two cases of mutual dependence: competitive \( (SBA_c) \) and cooperative \( (SBA_s) \) dependence.

---

\(^5\)As the profile definition is domain dependent, we refer the reader to [18] for an example of profile usage in an application.
Proposition 2 (Consistency of Altruism)
(a) Agent $i$ which his $S_i$ satisfies the mental states for SBA$_a$ (competitive social group) will not behave according to axiom A.1. (b) Agent $i$ which his $S_i$ satisfies the mental states for SBA$_a$ (competitive social group) might behave according to axiom A.5.

Proof: (a) When $i$ satisfies SBA$_a$ he believes to be in a competitive social group. Item 3.a in the mental state model dictates a mutual belief on a need to maximize the difference between his own benefit and the other members, which is inconsistency rationality with adopting an $ben_i(I_i, B_i) < 0$ intention. (b) When $i$ satisfies SBA$_a$ he believes to be in a cooperative social group. Item 3.b in the mental state model dictates a mutual belief that $ben_i(I_i \cup \ldots \cup I_n, B_i) \geq ben_i(I_i, B_i)$. In order to satisfy that, $i$ might need to adopt a $ben_i(I_i, B_i) < 0$ intention, to be consistent with the former condition.

Proposition 3 (Consistency of Aggression)
(a) Agent $i$ which his $S_i$ satisfies the mental states for SBA$_a$ (cooperative social group) will not behave according to axiom A.5. (b) Agent $i$ which his $S_i$ satisfies the mental states for SBA$_a$ (cooperative social group) might behave according to axiom A.5.

Proof: (a) As the cooperative social group does not incentivize aggressive actions, a null-action will be always preferred to an aggressive action which lowers both his benefits. (b) It is intuitive to construct an example in which a “sacrificing” type of action will increase the gap in benefits, which is consistent with 3.1.

In a cooperative scenario (SBA$_a$), 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 (proposition 3). On the other hand, in a competitive scenario (SBA$_a$), the agents believe in a negative dependency between their benefit 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 acts 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 (proposition 2).

6. 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. This spectrum will allow him to reason about which group is expected to be most beneficial for himself, and adopt its intentions accordingly. We exemplify the various groups using the running example.

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

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

(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 from page 4, 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 $A$, agents $i, j \in A$, $i \neq j$, 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:

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

$$\forall \alpha \in actions(I_i), \alpha \in \{C_i^{cop} \cup C_i^{ind} \cup C_i^{com}\}$$

(2)

As seen before, when a rational agent is alone in its environment it selects the set of desires that max-
may adopt a different set of coordinated cooperative desires where the group selects the best way to achieve the positive desires with other members to optimize its benefit. However, 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 (\(C_{i}^{\text{cop}} \subseteq C_{i}^{\text{cop}}\) is the subset that \(i\) performs by itself), thus decreasing \(i\)’s total cost to \(c(C_{i}^{\text{cop}} \setminus C_{i}^{\text{cop}})\). (2) negative impact - by achieving part of its competitive desires (\(D_{i}^{\text{conf}} \subseteq D_{i}^{\text{com}}\)), thus decreasing its total value from not receiving their benefits.

**Proposition 4 (Benefits of uncoordinated social group)**

If \(D_{i}^{\text{com}} = \emptyset\) \(\forall i \in \alpha \in \{C_{i}^{\text{cop}}, C_{i}^{\text{conf}}\} \leq v^{\prime}(D_{i}^{\text{com}} \setminus D_{i}^{\text{com}}) = c^{\prime}(C_{i}^{\text{com}})\), then \(\text{Ben}_{i}^{\text{usg}} \geq \text{Ben}_{i}^{\text{ind}}\).

**Proof:** 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 \(\text{Ben}_{i}^{\text{usg}} = \text{Ben}_{i}^{\text{ind}}\). The second case is when the positive impact (when other agents achieve \(i\)’s desires) is greater or equal to the negative impact, and \(\text{Ben}_{i}^{\text{usg}} \geq \text{Ben}_{i}^{\text{ind}}\).

The above proposition states that \(i\) is always better off being a member 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.

\[
\text{Ben}_{i}^{\text{usg}}(A, B) = 0 + (0.5 \times 4 - 0.5 \times 8) + 7 = 5,
\]

Bob with \(\text{Ben}_{i}^{\text{usg}}(B, B) = 2 + 4 = 6\) and Chen with \(\text{Ben}_{i}^{\text{usg}}(C, B) = (0.5 \times 3 - 0.5 \times 5) + 2 = 1\).

**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. A case of cooperative desires with coordination is equivalent to previous models (e.g., SharedPlans), where the group selects the best way to achieve the shared cooperative desire. In such cases, the agents may adopt a different set of coordinated cooperative actions, \(C_{i}^{\text{cop}}\), than in cases where coordination is not possible. Similarly, the agent can negotiate its competitive desires with other members to optimize its benefits, and thus agree on a partial set of competitive desires \(D_{i}^{\text{conf}}(\alpha)\) (with the respective \(C_{i}^{\text{conf}}\) 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 \((D_{i}^{\text{ind}'} \subseteq D_{i}^{\text{ind}}\) with respective \(C_{i}^{\text{ind}'} \subseteq C_{i}^{\text{ind}}\) set). Given a social group \(A\), agents \(i, j \in A, i \neq j\), the maximum total benefit of \(i\), as a member of a Coordinated Social Group (CSG), denoted as \(\text{Ben}_{i}^{\text{csg}}(I_{i} \cup I_{j}, B_{i})\) is composed of:

\[
\forall \varphi \in \text{goals}(I_{i}), \varphi \in \{D_{i}^{\text{cop}}, D_{i}^{\text{ind}'} \cup D_{i}^{\text{conf}}(\alpha)\} \land
\forall \alpha \in \text{actions}(I_{i}), \alpha \in \{C_{i}^{\text{cop}}, C_{i}^{\text{conf}}, C_{i}^{\text{ind}'} \}
\]

**Proposition 5 (Benefits of coordination)**

\(\text{Ben}_{i}^{\text{csg}} \geq \text{Ben}_{i}^{\text{usg}}\), when \(v^{\prime}\{D_{i}^{\text{conf}}(\alpha), D_{i}^{\text{ind}'}(\alpha)\} - c^{\prime}\{C_{i}^{\text{conf}}(\alpha), C_{i}^{\text{conf}N}\} > v^{\prime}\{D_{i}^{\text{com}}(\alpha), D_{i}^{\text{ind}}\} - c^{\prime}\{C_{i}^{\text{conf}}(\alpha), C_{i}^{\text{ind}'}(\alpha), C_{i}^{\text{conf}N}\}\)

**Proof:** Trivial by simple arithmetic manipulations.

Proposition (5) 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. It will also depend on the agents individual negotiation skills and strategic reasoning capabilities. If might be direct negotiations between the agents, or it might be that a central mechanism will be the reasoning side who simply dictates the agents which intentions to adopt, and which one to avoid. This will change according to the type of BDI agents in question. Nevertheless, we can assume that an optimal solution will not suddenly emerge due to computation limitation (we discuss this problem in depth in section 7). 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, \(\text{Ben}_{i}^{\text{csg}}(A, B) = 0 + 0 + 7 = 5\), Bob with \(\text{Ben}_{i}^{\text{csg}}(B, B) = 8 + 4 = 12\) and Chen with \(\text{Ben}_{i}^{\text{csg}}(C, B) = 3 + 6 = 9\). 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 \( A \), agents \( i,j \in A, i \neq j \), 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:

\[
v^i(D_{i}^{cop} \cup D_{i}^{ind''} \cup D_{i}^{com''} \cup_{\forall_{j} D_{j}^{alt}} - e^i(C_{i}^{copN} \cup C_{i}^{ind''} \cup C_{i}^{com''} \cup C_{i}^{alt})
\]

As we noticed in the previous group type, when the agent is part of a coordinated group, proposition 5 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 6 (Benefits of cooperation)**

If \( Ben_{i}^{cop} \geq Ben_{i}^{sg} \), then \( Ben_{i}^{CopSG} \geq Ben_{i}^{sg} \), when \( v^i(D_{i}^{alt}) - e^i(C_{i}^{alt}) \geq 0 \).

**Proof:** By aligning 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**

If \( Ben_{i}^{CopSG} \geq Ben_{i}^{sg} \) and \( Ben_{j}^{CopSG} \leq Ben_{j}^{sg} \), then \( ML_{alt} = Ben_{i}^{CopSG} - Ben_{i}^{sg} \).

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 \( Ben_{j}^{CopSG} < Ben_{j}^{sg} \)), 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 \( D_{i}^{comN} \) will not be part of \( D_{i}^{comN'} \). Given a social group \( A \), agents \( i,j \in A, i \neq j \), 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:

\[
v^i(D_{i}^{comN'} \cup D_{i}^{ind} \cup D_{i}^{comN'} \cup_{\forall_{j}} D_{j}^{agg}) - e^i(C_{i}^{copN'} \cup C_{i}^{ind} \cup C_{i}^{comN'} \cup C_{i}^{agg})
\]

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 \( \exists MD \in D_{i}^{com} \text{ s.t.} d = \max{\{ben_{i}(I_{1} \cup \ldots \cup I_{n}, B_{i}) - (\forall j \in A, j \neq i)ben_{j}(I_{1} \cup \ldots \cup I_{n}, B_{i})\}} \).

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.

**7. 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 \( ben_{i}(I_{1} \cup \ldots \cup I_{n}, B_{i}) \geq ben_{i}(I_{1}, B_{i}) \), that is a main query behind the question: to member or not to member (a cooperative group)? Given the vast amount of information that is needed the answer to this ques-
tion relays on including the beliefs of the agent, his desires, his valuation and cost functions, and similar knowledge on the BDI structure of the other agents that member the group in question, it seems impractical to assume that an automated agent will be able to compute an answer. The following theorem shows that it is 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:** 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 \( v = \{x_1, x_2, ..., x_k\} \), 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 \( v' = \{x_1, x_2, ..., x_k\} \), 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. \( \sum_{i=1}^{n} v_p i \geq \sum_{i=1}^{n} v' p_i \) (higher social welfare).
2. \( (\forall i)v_p i \geq v' p_i - \epsilon \) (each individual is no worse than what he could have before minus some constant \( \epsilon \)).

However, as the optimal solution is not accessible, oftentimes the agents will settle with some approximated solution that they can attain rather promptly, as a function of the available computational resources, and the requested precision. As human being in real life are able to get an approximated decision to the membership question using fast computed heuristic function [33], we might as well not aspire for an optimal solution, but settle with some approximated solution that bounded rational automated agents can get rather promptly, as a function of the available computational resources, and the requested precision.

### 7.1. 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: \((\forall i)v_p i \geq v' p_i - \epsilon\).
3. Social Welfare constraint: \(\sum_{i=1}^{n} v_p i \geq \sum_{i=1}^{n} v' p_i\).

![Fig. 2. Example graph representation](image)

**Running example:** According to the previous example Alice has three desires \( \{d_{A1}, d_{A2}, d_{A3}\} \), Bob’s desires are \( \{d_{B1}, d_{B2}\} \), and Chen’s desires are \( \{d_{C1}, d_{C2}\} \). The submit proposal desires are cooperative \( \{d_{A1}, d_{B1}, d_{C1}\} \) desire as all three are working jointly on the same proposal. The position appointment \( \{d_{A1}, d_{C2}\} \) 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: \(\text{Ben}_A^{\text{UCG}} = 5\), \(\text{Ben}_B^{\text{UCG}} = 6\), and \(\text{Ben}_C^{\text{UCG}} = 1\), and the total social welfare, \(\sum_{i=1}^{n} v_p i = 12\). A

\(^6\)Recall that we assume all parties have an equal chance of attaining competitive desires, and compute the expected benefit accordingly.
graph representation of the problem is depicted in figure 2. One solution for the above problem is \( v' = [d_A^1, d_A^3, d_A^2, d_B^1, d_B^2, -d_C^2], \) which results in respective individual benefits of: 7 for Alice, 12 for Bob, and 9 for Chen. The social welfare also increases to \( \sum_{i=1}^{n} v'p_i = 28. \)

7.2. 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.\(^7\) 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 pursued by the agents. This situation resembles a real life scenario in which the agents sit by the table and openly communicate and try to form a cooperative group, or situations in which a central principal, who have all the information, tries to coordinate their actions.

**H1. Smallest number of desires** — This heuristic dictates searching first for a solution that satisfies the agent with the smallest number of desires. Obviously, less desires makes it easier to fully satisfy their goals.

**H2. Largest number of competitive desires** — This heuristic tries to assign the agents with the largest number of competitive desires. The intuition here is that competitive desires provide more constraint as they always affect at least two agents.

**H3. Smallest number of cooperative desires** — As cooperative desires are usually easier to assign (as at least two options are available), this function tries first to assign the agents that have the smallest amount of these.

**H4. Maximize H1 + H3 - H2** — This function aggregates the rational behind the first three functions and dictates starting with the agent that maximizes the sum of H1 and H3 minus H2.

**H5. Largest B\(_{\text{max}}\) value** — This function is very different from the previous four as it looks at the maximum benefit that an agent gets from being part of an uncoordinated social group. In other words, we will start by assigning the agent that would be in the best situation if such coordination effort would not take place at all.

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 \( pc = 0.15, pm = 0.15 \) 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. Recall that we are interested in finding a solution, and not necessarily the optimal one, and there were instances in which a solution was not available. We used a 2.16 Ghz Intel Core 2 Duo system with 2 GB of memory.

![Fig. 3. Simulations results](image)

The results are depicted in figure 3, provide several insights on the behavior of the functions and the ef-
fect of the environment’s characteristics on the solution time. The y-axis is the time in minutes that it took before getting a solution to the problem. First, it is easy to see competitive environments proved to be significantly more challenging than the other environments. One has to wait almost twice as long before getting the first solution. Second, it is easy to see that H4 provides 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. For sake of completeness, the average solution times when not using any heuristic function (and when solutions did exist) were: 104, 87.6 and 282.7 respectively. It is also interesting to observe that H5, which looked first at the agent who were better of without being coordinated, did not do as well as expected. It might be the case that ignoring the competition instead of trying to “manage” it is not the best course of action.

8. Relation with other BDI based models

The SharedPlans formalism, which we extend in this work, is a good exemplar of the golden age of BDI frameworks in the 90s. This formalism dictates all agents to have a joint utility function, thus describing a case of full cooperation. While joint group goal (or a “task group” as coined in the Psychology literature) is extremely important in multi-agent interactions, it obviously does not describe other forms of interactions. One example is a “treatment group” in which individuals are working on different goals, but can help or be helped by others. Such group describe well the inter-action in a graduate students lab: each student is working on his own work, but mutual help still exist in various forms. Such form of interaction was describe using the SharedActivity notation in [18] and was exemplified in a touring museum domain. Another, more re-cent model was presented in [38]. The AdversarialActivity model presents a BDI framework for the competitive part of the spectrum.

It is of crucial importance to connect any newly presented formalism to older versions in order to con-clude whether systems that implement any of the older frameworks, can be easily converted to include the new formalism. As such, we would like to show that the SBA model contains these three models. In other words, we show that the SBA can be regarded as a generalization of them without losing descriptive or behavorial power. We will show that through a series of containment relationships that show the following:

- SharedPlans $\subseteq$ SharedActivity $\subseteq$ SocialBehaviorActivity
- AdversarialActivity $\subseteq$ SocialBehaviorActivity

To prove that a model is contained in another, we show that: (1) The definition of the first entails the definition of the second (i.e., an agent that holds mental states of the second model must also hold these mental states in the first); (2) All the axioms in the first model hold, and do not contradict the second model.

Theorem 2 SharedActivity $\subseteq$ SocialBehaviorActivity.

Proof: This containment is quite obvious, as the definition of SA is equivalent to $SBA_b$ part of the SBA model, as clauses 1–4 in the SA model are identical to clauses 1, 3.b.i, 2, and 3.b in the SBA model, respectively. The helpful-behavior act and the cooperative act in the SA model are identical to A.1 and A.2 in the SBA model. Moreover, the union of A.3 and A.4 in the SBA model (individualism and competitive acts) are exactly the selfish act axiom from the SA model with their respective bounds in the cooperative case of the SBA model.

Theorem 3 SharedPlan $\subseteq$ SharedActivity.

Proof sketch. (complete proof in the Appendix) To show definition entailment we suggest explicitly adding two axioms to the SP model that hold during rational behavior, but were implicit in the definition. The first axiom (N1) states that having a mutual belief by a group A entails the ability of the agents to identify themselves as members of A. The second axiom (N2) states that a group A having some intention entails an intention by the group that the group will be maintained.

With those axioms in mind, it is easy to see that clause (0) of the partial SP model entails clause (1) of the $SBA_b$ model using axiom N1 above. In addition, axiom SP–A3 of the SP model together with N2 above entails clause (3.b) of the $SBA_b$. Clause (2) of SBA requires a belief in a (partial) profile of other members; Those parts are represented in the SP model by the meta-predicates CBA(...) (“can bring about”) and CBAG(...) (“can bring about group”). To prove that the $SBA_b$ model contains the axioms of the SP model, it is easy to see that an agent that acts according to the SP axiom is coherent with the SBA model.
Theorem 4
AdversarialActivity ⊆ SocialBehaviorActivity.

Proof: The AdversarialActivity (AA) model reflects the interactions from a single agent point of view (agent $A_0$ in the formulation) in order to clarify the model’s applicability. The following presentation gives the general version of the model, which quantifies over the point of view of the agent, and allows using any agent in $A$ as the reference agent via substitution:

$$
\text{AA}(A, G^*_A, T_n, w) = \\
(\forall A_i \in A)(\exists \alpha \in C_{A_i}, T_n) \cup \\
\text{Int.Th}(A_i, \text{Achieve}(G^*_A, \alpha), T_n, T_n, \text{AA}) = \\
(\forall A_i, A_o \in A)\text{Bel}(A_i, \text{FulConf}(G^*_A, G^*_A), T_n) = \\
(\forall A_i, A_o \in A)(\exists \alpha_o \in C_{A_o}, T_n) \cup \\
\text{Bel}(A_i, \text{Int.Th}(A_o, \text{Achieve}(G^*_A, \alpha_o), T_n, T_{\alpha_o}, \text{AA}), T_n) = \\
(\forall A_i, o \in A)(\exists P^*_A \in P_A)\text{Bel}(A_i, \text{Profile}(A_o, P^*_A), T_n)
$$

Intuitively, AA models a zero-sum interaction while the SBA model is general and spans the whole individual-aggression range. The heart of the containment comes from clause 2 of the AA model. All $A$’s agents are in a FulConf (full conflict) which (by definition) allows Achieve($G^*_A, \alpha$) to hold for only a single agent at a time. Using property 2 of the AA model (when Achieve($G^*_A, \alpha$) is true, agent $A_i$ has the maximum eval value) and AA clause 1, we see that the intention that Achieve($G^*_A, \alpha$) holds implies the intention to attain higher utility than the others, which increases their difference in a zero-sum interaction as stated in clause 3.b.1 of the SBA model. In addition, clause 4 in the AA model describes the agent’s belief in the profile of the adversaries, which is exactly what is stated in clause 2 of the SBA model.

9. Conclusions and Future Work

We presented the full social behavior spectrum as it has been modeled in social science research, and suggested the multiagent SocialBehaviorActivity 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. Our first step will be to implement an automated agent in the Du-Board game, and then continue to a more complex environment. This will allow us to empirically explore the benefits of using our architectural guidelines as a basis for the social agent’s design. Another limitation of our framework is the strong requirement for an explicit utility values and evaluation functions. Future iterations might tackle this by adding probabilistic values or fuzziness to the decision procedure.

References


[37] I. Zuckerman et al. / A BDI-based Agent Architecture for Social Competent Agents


Appendix

A. Proof of Containment Relation

Theorem 5 \( \text{SharedPlan} \subseteq \text{SharedActivity} \).

To show definition entailment we suggest explicitly adding two axioms to the SP model that hold during rational behavior, but were implicit in the definition. The first axiom states that having a mutual belief by a group \( A \) entails the ability of the agents to identify themselves as members of \( A \).

\[ N_1. MB(A, f, T_f) \Rightarrow MB(A, (\forall A_i \in A)\text{member}(A_i, A), T_f) \]

The second axiom states that a group \( A \) having some intention entails an intention by the group that the group will be maintained.

\[ N_2. \text{Int.Th}(A, \text{prop}, T_n, T_{prop}, S) \Rightarrow MB(A, (\forall A_i \in A)\text{Int.Th}(A_i, \text{member}(A_i, A), T_n, T_{prop}, S), T_n) \]

With those axioms in mind, it is easy to see that clause (0) of the partial SP model \( MB(A, (\forall A_j \in A)\text{Int.Th}(A_j, Do(\ldots), \ldots), T) \) entails clause (1) of the \( SB_{A_0} \) model using axiom \( N_1 \) above. In addition, axiom SP–A3 of the SP model:

\[ \text{Bel}(A, \text{Int.Th}(A, \text{prop}, T, T_{prop}, C_{prop}), T) \Rightarrow \text{Int.Th}(A, \text{prop}, T, T_{prop}, C_{prop}) \]

entails clause (3.b) of the \( SB_{A_0} \).

Clause (2) of SBA requires a belief in a (partial) profile of other members; the profile was defined as life histories, development patterns, needs, goals, and behavioral patterns. In the SP model, the agent’s intentions, capabilities and situations can be considered the profile. Those parts are represented in the SP model by the meta-predicates CBA(\ldots) (“can bring about”) and CBAG(\ldots) (“can bring about group”). Clause (1b1) of the partial SP definition entails FSP (“Full Shared Plan”) for selecting a recipe (using the Select_Rec_GR predicate), which in turn entails beliefs about the capabilities of the agent (i.e., about the CBA(\ldots) and CBAG(\ldots) meta-predicates).

Clause (3.b) of the \( SB_{A_0} \) states the necessity of mutual dependency. Thus, it remains to show that when the group’s members act according to the SP model, they have an MB that acting according to the SP generates better utility. However, the definition requires the agents to reconcile their intentions before adopting them, a process that causes the agents to perform the activities that generate the best possible utility.

To prove that the \( SB_{A_0} \) model contains the axioms of the SP model, we show that an agent that acts according to the SP axiom is coherent with the SBA model. It is important to note that since agents who act as a task group (modeled by the SP formalization) obtain part of their utility from the group, they have a stake in maximizing group utility. A greater group benefit means a greater share for each agent, which leads to greater individual utility. Therefore, only the Cooperation act axiom is relevant. There are 7 axioms in the SP model: axioms SP-A1 to SP-A5 state single-agent rational behavior and provide relations between intentions and belief in a single agent, and thus they must hold in the SA model. Moving on to SP-A6 and SP-A7, we encounter two axioms for the Int.Th(\ldots) (“intention that”) predicate, which are clearly valid and do not conflict with any of the SA axioms. The final axiom SP-A7 is the one that provides a basis for helpful behavior in an SP context. Being derived from the SA Cooperation act axiom (SA-A2), it is easy to see that this axiom is coherent with the SA model.