The Social Landscape: Reasoning on the Social Behavior Spectrum

Inon Zuckerman, Ariel University Center of Samaria
Meirav Hadad, Bar Ilan University

A society can be loosely defined as a group of individuals connected to each other in some form of persistent relationship. Humans’ social competence is the product of a long evolutionary process that provides individuals with the required skills to survive and flourish in a tight social structure.

Proliferating social integration between humans and computational entities increases the need for agent architectures that span the entire social behavior spectrum. Accordingly, an important problem in artificial social intelligence is the ability to correctly capture the social structure and use the resulting model to navigate and achieve one’s goals.

Over the years, researchers have suggested many different approaches to augmenting agents with social know-how. Our goal in this article is to present one such approach, a mental model that provides computational entities (agents and robots) with social capabilities similar to those of humans. The model captures the entire social behavior spectrum and provides design principles that allow agents to reason and change their behavior according to their subjective perception of the cooperative and competitive natures of the society.

This social behavior activity (SBA) model builds upon the belief-desire-intention (BDI) model of human practical reasoning. BDI models are highly applicable as software models for developing bounded rational intelligent agents. Although previous models described specific parts of the social behavior spectrum (particularly joint goals), our model spans the full social spectrum, including relations that haven’t yet received attention. The contribution of such a mental model is twofold: first, it provides a formal, valid theoretical foundation for explaining and predicting agents’ social behavior. Second, it serves as a set of design principles to guide the creation of agents capable of engaging in behaviors that span the entire spectrum.

The Social Behavior Spectrum
Sociological models posit a spectrum of social interpersonal behavior, starting with altruistic behavior at one end (defined as increasing the other person’s outcome) and extending through cooperation (increasing both parties’ outcomes), individualism (increasing your own outcome), and competition (increasing your own outcome while...
decreasing your opponent’s outcome) to aggression (decreasing your opponent’s outcome without increasing your own, for the sole purpose of inflicting injury) at the other end.

Figure 1 shows a graphical description of the major interpersonal orientations that can occur between two players. In this model, one player’s outcome lies on the horizontal axis, and the “other” player’s is on the vertical axis. Each outcome increases along each axis, and the values reflect a linear combination of payoffs to both players. The model treats multiple agents as pairwise aggregations of the two-person model.

Our social behavior activity (SBA) model is a BDI-based formalism written in propositional logic with modality operators that seeks to describe an individual’s mental states during a social interaction. The model lets us build agents that can reason about intertwined beliefs, desires, and intentions with a valuation and with cost functions. We can define a benefit function as valuation minus cost.

Previous BDI models focused only on cooperation while working solely with individual or joint desires, but in a complex society, the desires of the different agents can be cooperative, individual, or competitive. Our model relies on the following two definitions:

- Given the set of beliefs $B_i$, desires $D_i$, intentions $I_i$, valuation function $v^i$, and cost function $c^i$, a BDI structure for agent $i$ is $S_i = (B_i, D_i, I_i, v^i, c^i)$.
- A multiagent system will comprise multiple computational entities, and will be denoted as $S$, where $\forall i, S_i \in S$.

To avoid describing a new BDI framework from scratch, we’ve built our model on the well-known Shared-Plans framework, which provides all the operators and predicates for facilitating a joint activity. Nevertheless, the extension is general and can be used to extend any other model.

The SBA has three characteristics. First, it implies the agents’ ability to identify themselves as members of some social group. Second, each individual in the group has its own life history, development patterns, and behavior patterns, which are represented in that individual’s profile to the extent they are known. Thus, the members must have beliefs concerning a partial profile of the others. And third, mutual dependence means that the benefits of all the parties are affected to some extent by the behaviors of the others. We differentiate (using exclusive or, symbolized by ⊕) between two cases of mutual dependence: competitive, in which agents believe in a negative dependency between their benefits’ functions; and cooperative, with a positive dependency between the benefit functions.

Expressed in propositional logic, the SBA model defines a social behavior activity, SBA $(S, A, P, T)$, for a group of agents $A$ in the context of the BDI structure $S$ and profiles $P$ at time $T_n$, with the following characteristics:

1. A mutually believes that all members are part of $A$: $MB(A, \forall i \in A)$ member$(i, A)$, $T_n$.
2. Members of $A$ have (partial) beliefs about the profiles of the other agents: $\forall i \in A)Bel(i, j, (j \in A)(\exists p_i \subseteq P), T_n)$.
3. A mutually believes that either a. all members of $A$ have the intention to attain the maximum positive difference between their own and their opponent’s benefits: $MB(A, f_i, f_j, f = (\forall i \in A)Int. Th_{i}, \max (\text{ben}(I_1 \cup \cdots \cup I_{n}, B_i) - (\forall n)\text{ben}(I_1 \cup \cdots \cup I_{n}, B_j)), T_n, T_{\text{max}(\ldots)}, S) \oplus$ b. being a member obtains a better benefit value: $MB(A, f_i, T_j), f = (\forall i \in A)\text{ben}(I_1 \cup \cdots \cup I_{n}, B_i) \geq \text{ben}(I_1, B_j), T_n)$, and $A$ has a mutual belief that each member intends to maintain the group: $MB(A, f_i, T_j), f = (\forall i \in A)Int. Th_{i}, \text{member}(i, A), T_n, T_{\text{member}(\ldots)}, S)$.

**Social Reasoning**

Because the BDI structure defines the valuation and cost functions, it lets us explore the spectrum by way of utility computation. An agent can compute and determine which social group will be most beneficial according to its beliefs about the relationships among desires. The following example will provide the basis for discussing how this plays out in three types of social groups.
Consider an academic department with three faculty members: Alice, Bob, and Chen. Alice has three desires:

- submitting a proposal,
- winning a position, and
- writing a paper.

Bob’s desires are

- submitting a proposal and
- writing a paper.

Chen’s desires are

- winning a position and
- submitting a proposal.

Formally, the set of agents is \{A, B, C\}. Alice’s desires are \(d_{1}, d_{2}, d_{3}\), with valuation of \(v = <5, 12, 10>\) and costs of \(c = <5, 8, 3>\). Bob’s desires are \(d_{4}, d_{5}\), with valuation of \(v = <8, 10>\) and costs of \(c = <6, 6>\), and Chen’s desires are \(d_{6}, d_{7}\), with valuation of \(v = <8, 6>\) and costs of \(c = <5, 4>\).

The proposal desires are cooperative, because all three are working on the same proposal \(d_{A} = d_{B} = d_{C}\). In a well-coordinated interaction, only a single agent will spend the cost associated with this desire. In contrast, winning a position is a competitive desire for Alice and Chen, and without coordination both will spend the cost of pursuing it, but only one will win. The others are individual desires.

Given that example, we can examine how the interactions work out in three types of social groups. Other types of groups exist in the social and behavioral literature, of course, which we could use in conjunction with the mental model.

**Maximum Individual Benefit**

One form of social group features the members acting alone in the environment. The maximum benefit that each may obtain is the sum of its own cooperative, individual, and competitive desires together with their relevant costs. In our example, Alice would receive a benefit of \(0 + 4 + 7 = 11\), Bob a benefit of \(2 + 4 = 6\), and Chen a benefit of \(3 + 2 = 5\).

**Uncoordinated Social Group**

In an uncoordinated social group, the members act alone according to their desires but share the environment with other agents. For that reason, their beliefs about the other agents’ intentions might influence their own actions. In this scenario, we assume that the agents do not engage in any sort of coordination activity.

Another agent in the environment can have two kinds of effects: a positive impact, by achieving part of its own desires from the set of cooperative desires; or a negative impact, by achieving part of its competitive desires at the expense of others, thus decreasing the group’s total benefit value.

Our example includes one cooperative desire that all three agents share, so they spend a lot of effort on that one without coordination. Alice and Chen also share a competitive desire; without coordinating, we can compute the benefit as follows: Alice’s is \(0 + (0.5 \times 4 - 0.5 \times 8) + 7 = 5\); Bob’s is still \(2 + 4 = 6\); and Chen’s is now \((0.5 \times 3 - 0.5 \times 5) + 2 = 1\).

**Coordinated Social Group**

A coordinated social group is one in which the members first coordinate their actions and intentions—cooperative as well as competitive—to maximize their benefits. This can lead to a different set of coordinated cooperative actions than in cases where coordination is not possible. Similarly, an agent can negotiate its competitive desires with other members to optimize its benefit, thus agreeing on a partial set of competitive desires. In such cases, the agent might be asked to give up some weak competitive desires that decrease the benefits of other agents, and accordingly it might exchange its set of individual desires to achieve only part of them.

In the example, we can’t know exactly which of the possible solutions the agents will agree upon. It depends on their individual negotiation skills and strategic-reasoning capabilities. One possible solution would be for Alice to drop the intention to be elected but agree to be the one that submits the proposal. In that case, Alice’s benefit will be \(0 + 0 + 7 = 7\), Bob’s will be \(8 + 4 = 12\), and Chen’s will be \(3 + 6 = 9\). All three players end up better off than before the coordination, and the social welfare has increased from 12 to 28.

**To Become a Member or Not**

One key ingredient is still missing to make the model practical for architectural purposes. It lies behind the question implicit in element 3b of the SBA model: to join or not to join? For an individual to compute the optimal decision about whether to join a cooperative social group is an NP-hard problem. However, in real life, humans often arrive at an approximate solution through heuristics, so we might as well try a similar approach.

To empirically evaluate the practicality of the membership query in real applications, we formulated it as a Boolean constraints satisfaction problem. Each member of the set of agents has its own set of desires, for a total of \(N\) desires. We formulate each desire as an individual node, and the edges between nodes model the relationships between desires. We model the division of desires into cooperative,
individualistic, or competitive groups as follows:

- Cooperative desires have an or constraint edge connecting the nodes.
- Individual desires are simply the nodes themselves.
- Competitive desires have an exclusive or constraint edge.

A solution will result in the assignment of values 1 or 0 to the nodes—1 if the desire is fulfilled, 0 if not—such that the following constraints hold:

- The local desires constraints are not violated.
- Individual benefit constraint: \((\forall i) v_{p_i} \geq v'_{p_i} - \varepsilon\).
- Social welfare constraint: \(\sum_{i=1}^{n} v_{p_i} \geq \sum_{i=1}^{n} v'_{p_i}\).

In our example, the “submit proposal” desires are cooperative, because all three agents are working jointly on the same proposal. The “win a position” desires are competitive for Alice and Chen. The others are individual desires without any relationship among them. Given the valuation and cost values, we can compute the maximal uncoordinated valuation and cost values, we can jointly on the same proposal. The “win a position” desires are competitive, because each player’s goal is represented as either prioritized goals or conditional preference networks.

In Boolean games, each agent wants to accomplish its own individual goal, which is represented as a propositional-logic formula over some set of Boolean variables. Each agent is assumed to uniquely control a subset of the Boolean variables, while the set of valuations of these variables corresponds to the set of actions the agent can take. Setting a variable to some value, such as taking a particular action, might be associated with a nonuniform cost function.

The main difference between Boolean games and BDI-based frameworks is that the former are abstract mathematical frameworks that are easier to work with for analysis of strategic interactions. In contrast, BDI-based frameworks provide a more applicable design-oriented layer for constructing autonomous agents.

In cases where agents seek to apply complex strategic reasoning procedures (such as computing equilibrium points), it’s possible to reduce the proposed BDI framework to a Boolean game if you add two constraints.

We can define an \(n\)-player Boolean game as a four-tuple \((A, V, \pi, \Phi)\), where \(A = \{1, 2, \ldots, n\}\) is a set of players, \(V\) is a set of propositional variables, and \(\pi : A \rightarrow V\) is a control assignment function where \(\{\varphi_1, \ldots, \varphi_n\}\) forms a partition of \(V\). \(\Phi = \{\varphi_1, \ldots, \varphi_n\}\) are a collection of formulas of the language, where \(\varphi_i \rightarrow [0, 1]\).

This definition poses two challenges when trying to reduce a multi-agent BDI system \(S\) to a Boolean game:

- because each player’s goal \(\varphi_i \in [0, 1]\), different goals cannot carry different weights; and
- because each variable is controlled by only one agent, it can’t represent competitive and cooperative desires.

With respect to the first challenge, we can follow Elise Bonzon and her colleagues’ extension and define a preference relation on the goals. A preference relation \(\succeq\) is a reflexive and transitive binary relation on the set of truth assignments for all variables that \(i\) controls.

The second challenge, the requirement that only one agent control each variable, obviously restricts the Boolean game’s expressivity when compared to a multiagent BDI system where cooperative and competitive goals do exist. This constraint can be relaxed either by providing coherent semantics to describe the result of two agents simultaneously changing the same control variable (for example, having an equal chance of success), or by restricting the BDI framework from permitting...
such occurrences. After taking one of those steps, we can easily reduce a multi-agent BDI system $S$ to a (prioritized) $n$-player Boolean game $BG$ as follows: $A \subseteq S$ maps to $A \subseteq BG$, $D \subseteq S$ to $\Phi \subseteq BG$, and $I \subseteq S$ to $V \subseteq BG$ (where each $I_i$ is mapped according to the respective $\phi_i$). Finally, we map $v$, $c$, together as the value for $\phi$.

Following the reduction, we can exploit the analytic power of Boolean games and compute dominant strategies, Nash equilibrium, and other common aspects of game theory.

**Relationship to Other BDI-Based Models**

One question about the SBA model is how it relates to previous BDI formalisms. Furthermore, is it possible to migrate applications that were based on one of the previous models to the new SBA model? Theoretical exploration of the underlying model’s relationship to the new one can provide an answer to that question. To that end, we analyzed the SBA model’s relationship to the following BDI models: SharedPlans, SharedActivity, and Adversarial Activity.\(^6\),\(^10\),\(^11\)

The SharedPlans formalism is a good example of the 1990s “golden age” of BDI frameworks that built on Michael Bratman’s insightful philosophical work.\(^2\) It provides a sound and complete formalization of the pure-cooperation part of the social behavior spectrum—in other words, it describes cases in which all the agents have the exact same set of goals or share a single joint goal.

Although groups with joint goals (task groups, in psychology literature) are extremely important in multiagent interactions, they obviously don’t cover all possible cases. In a treatment group, individuals are working on different goals but can help or be helped by others; for example, in a graduate student lab, each student is working on his or her own work, but mutual help still exists in various forms. The SharedActivity formalism describes this form of interaction.

The SharedPlans and Shared-Activity formalisms cover some forms of cooperative interaction. The third BDI model we looked at, Adversarial Activity, presents a framework for the competitive part of the spectrum. In its current form, however, it applies only to zero-sum games.

To prove that one behavior model is contained in another, we can show that the definition of the first entails the definition of the second (that is, an agent that holds mental states of the second model must also hold these mental states in the first); and that all the axioms in the first model hold, and do not contradict, the second model. In our analysis, we were able to show the following series of containment relationships:

- $\text{SharedPlans} \subseteq \text{SharedActivity} \subseteq \text{SocialBehaviorActivity}$; and
- $\text{AdversarialActivity} \subseteq \text{SocialBehaviorActivity}$

**References**


---

**The Authors**

Inon Zuckerman is a faculty member at the Industrial Engineering and Management Department at Ariel University Center of Samaria. His research interests are in AI and its intersections with economy, psychology, biology, and the social sciences. He has a PhD in computer science from Bar-Ilan University. Contact him at inonz@ariel.ac.il.

Meirav Hadad is an adjunct lecturer at Bar Ilan University. Her research interests are AI and its intersections with economy, psychology, and the social sciences. She has a PhD in computer science from Bar-Ilan University. Contact her at hadad@biu.cs.ac.il.


Distinguish Yourself With the CSDP

In a world full of software developers, how do you stand out from the crowd? By adding the CSDP credential, you will be:

- Current with best software practices
- Connected with industry’s brightest minds
- Career-minded and ready for that next promotion
- Committed to advancing the software engineering profession

To read how our credentials have helped software development professionals as well as employers, go to:

www.computer.org/getcertified