From Rational to Emotional Agents

Hong Jiang and **José M. Vidal**
Computer Science and Engineering
University of South Carolina
Columbia, SC 29208, USA
jiangh@engr.sc.edu, vidal@sc.edu

Abstract

This paper presents the Emotional-Belief-Desire-Intention architecture which reflects humans' practical reasoning by adding the influence of primary and secondary emotions into the decision making process of a traditional BDI architecture. Our architecture handles bounded resources by using primary emotions as the first filter for adjusting the priority of beliefs, thereby allowing the agents to speed up decision making. Secondary emotions are used to refine the decision when time permits. We present a sample EBDI agent for the Tileworld domain in order to show our architecture might be used.

Introduction

Most of the research into agents has focused on the development of rational utility-maximizing agents. This research assumes that decisions derive from an analysis of the future outcomes of various options and alternatives. The influence that emotions have on human decision-making is largely ignored. However, neurological studies of decision-making in patients with brain lesions that prevent them from processing emotions suggests that people make judgments not only by evaluating the consequences and their probability of occurring but also, and even sometimes primarily, at a gut or emotional level (Bechara 2004).

While some researchers have tried to expand traditional agents by adding emotion to them, an universally accepted generic model or architecture for emotional agents has not yet appeared. The closest candidate is (Pereira *et al.* 2005), which present a Emotional-BDI architecture including internal representations for agents' capabilities and resources. Unfortunately, this paper does not clearly represent the difference between emotional agents and normal rational agents. The capabilities and resources of these agents are independent of emotions, as such, they cannot reflect the relationship between emotions and belief, and how emotions influence agents' decision making.

Based on the idea of (Damasio 1994), our architecture takes into account both primary emotions and secondary emotions into decision making process and generates a conceptual emotional model based on the BDI agent architecture which we call the EBDI architecture. EBDI solves the problem of an agent's bounded computational resources by using primary emotions as a first filter for adjusting the priority of beliefs, such that the agents can speed up decision making. Secondary emotions are also used to refine the decision when time permits. Another addition to the standards BDI model is that instead of just considering beliefs from perception we add possible belief candidates directly from communication and contemplation. To improve the balance between being committed to and over-committed to one's intentions, which is a major problem in BDI agent, we left the control of reconsideration into the design of plan execution function, since the solution of this depends more on the specific agent type and agent's strategy. We describe the EBDI architecture and give a interpreter. We supply a example agent located in a Tileworld (Pollack & Ringuette 1990) so as to show how to use this model to implement a practical agent,

The paper starts by first giving some research background and describing our motivation for choosing BDI as a basic model, as well as some concerns on incorporating emotions into rational agents. We then provide a detailed description of the EBDI model, including the main components and functions as well as the architecture and interpreter. Finally, we provide an example EBDI agent that inhabits the Tileworld domain.

Research Background and Motivation

There are many studies of emotion in the psychology literature (Sousa 2003). The work of (Wright *et al.* 1996) and (Picard 2000) placed emotion into computational theory and has led to increasing interest in computational models of emotion. (Ekman & Davidson 1994) reveals the central issues in emotion research and theory in the words of many of the leading scientists working in the field today. (Davidson, Scherer, & Goldsmith 2002) provides a survey of current work in the affective sciences.

There is also some research into the problem of incorporating emotions into artificial agents. (Padgham & Taylor 1996) believes that emotions and personality interact with goal-oriented behavior and describes some simplifications to build an initial interactive environment for experimentation with animated agents that simulate personality alongside ra-

tional goal-oriented behavior. (Morgado & Gaspar 2005) presents an agent model where emotion and cognition are conceived as two integrated aspects of intelligent behavior. They show affective-emotional mechanisms that support the adaptation to changing environments and a controlled use of resources.(Meyer 2004) extends the KARO (Knowledge, Abilities, Results and Opportunities) framework – supplies a range of expressive modal logics for describing the behavior of intelligent agents (Hustadt *et al.* 2001), and use logic in reasoning about the emotional or affective states that an agent can reside in. While these studies describe some emotional models for some specific systems, a more general model is still desired.

One effort to incorporate emotion's into a BDI architecture is (Pereira *et al.* 2005) which presents a Emotional-BDI architecture including internal representations for agent's capabilities and resources. However, this paper does not represent the difference between emotional agents and normal rational agents. The capabilities and resources themselves are independent of emotions, as such, they cannot reflect the relationship between emotions and beliefs or how emotions influence agents' decision making. Another effort to incorporate emotions into a BDI architecture is given in (Parunak *et al.* 2006). They enhance the standard BDI model using the OCC(Ortony, Clore, Collins) model of emotion (Ortony, Clore, & Collins 1988) in a framework that can support large numbers of combatants. However, it is not a generic model.

We focus on how emotions influence an agent's decision making and propose a generic Emotional-Belief-Desire-Intention architecture. We also show an example of how to simulate a EBDI agent. The following subsections describe our motivation for choosing a BDI architecture and what issues should be considered about emotions in decision making.

Why based on BDI architecture

There are four types of traditional agent architectures (Weiss 1999):

- Logic based agents—in which decision making is realized through logical deduction;
- Reactive agents—in which decision making is implemented in some form of direct mapping from situation to action;
- Belief-desire-intention agents—in which decision making depends upon the manipulation of data structures representing the beliefs, desires, and intentions of the agent;
- Layered architectures—in which decision making is realized via various software layers, each of which is more or less explicitly reasoning about the environment based at different levels of abstraction.

For the logic based agents, decision making is predicated on the assumption of calculative rationality—the assumption that the world will not change in any significant way while the agent is deciding what to do, and that an action which is rational when decision making begins is still rational when it ends. However, most current multiagent systems can hardly guarantee a static and deterministic environment. The problems associated with representing and reasoning about complex, dynamic, possibly physical environments are unsolved.

Reactive agents make decisions based on local information, so they must have sufficient information available in their local environment for them to determine an acceptable action, and it is difficult to see how such decision making could take into account non-local information. On the other hand, for reactive agents, overall behavior emerges from the interaction of the component behaviors when the agent is placed in its environment, which suggests that the relationship between individual behaviors, environment, and overall behavior is not understandable, such that it is very hard to engineer agents to fulfill specific tasks.

The layered architectures are very general. The main problem is that while they are arguably a pragmatic solution, they lack the conceptual and semantic clarity of un-layered approaches. Another issue is that of interactions among layers, if each layer is an independent activity producing process, then it is necessary to consider all possible ways that the layers can interact with one another.

The BDI architectures reflect human's practical reasoning process, which is part of the reason why we choose it as a basic rational model for agent. It has shown to be a very successful architecture and it is attractive for several reasons: first, it has widely accepted philosophical roots; second, there are logical frameworks for modeling and reasoning about BDI agents; third, there are a considerable set of software systems which employ the architecture's concepts, such as PRS system (Georgeff & Ingrand 1989). The main problems about this architecture are: how to efficiently implement these functions; how to reach the balance between being committed to and over-committed to one's intentions. As they stand, BDI architectures ignore the influence of emotions in decision-making.

Concerns about Emotions for Agents

Many concerns are mentioned when researchers take emotions into agents. The main issues concerned are the following:

- Some researchers consider it necessary to incorporate human aspects such as personality and emotion in order to make agents more engaging and believable so that they can better play a role in various interactive systems involving simulation (Padgham & Taylor 1996). Entertainment is one obvious application area for such simulation systems, another is education and training. For example, a simulator that was able to realistically model emotional reactions of people could be used in training programs for staff who need to be trained to deal with the public.
- Some people believe that emotions play a functional role in the behavior of humans and animals, particularly behavior as part of complex social systems (Toda 1982). a successful modeling of emotion will enable us to come closer to the goal of building software agents which approach humans in their flexibility and ability to be adaptable and survive in complex, changing and unpredictable environments. For example, as in systems the Woggles

of Oz-world (Bates 1997), emotion modifies the physical behavior of agents: a happy agent moves faster, and more bouncily, while a sad agent is slower and flatter in its movements.

- Emotions can effect an agent's goals, hence affecting their actions. Emotional effects on goals can be via reordering or re-prioritizing, existing goals, or by introducing completely new goals. The goals' success or failure can affect emotional states. An agent which experiences a goal failure may feel unhappy while one experiencing goal success may feel glad. (Dyer 1987) develops a comprehensive lexicon of emotional states based on goal success and failure.
- (Frijda & Swagerman 1987) postulates emotions as processes which safeguard long-term persistent goals or concerns of the agents, such as survival, a desire for stimulation or a wish to avoid cold and damp.
- (Toda 1982) postulates emotions as processes which affect the rational system of the agent, and which are based on basic urges: emergency urges, biological urges, cognitive urges and social urges. Emotions are seen as varying in intensity where the intensity level is an important factor in determining the effect on the rational processing of the agent.
- Rational agents often are thought as self-interest, that is, they always want to maximize their own wealth or other material goals. However, practically, people may sometimes choose to spend their wealth to punish others who have harmed them, reward those who have helped, or to make outcomes fairer (Camerer, Loewenstein, & Rabin 2003).
- (Damasio 1994) finds that people with relatively minor emotional impairments have trouble making decisions and, when they do, they often make disastrous ones. Other research shows that what appears to be deliberative decision making may actually be driven by gut-level emotions or drives, then rationalized as a thoughtful decision (Wegner & Wheatley 1999). (Bechara 2004) also mentions that most theories assume that decisions derive from an assessment of the future outcomes of various options and alternatives through some type of cost-benefit analysis. The influence of emotions on decision-making is largely ignored. The studies of decision-making in neurological patients who can no longer process emotional information normally suggest that people make judgments not only by evaluating the consequences and their probability of occurring, but also and even sometimes primarily at a gut or emotional level.

The above research confirms that emotions do have an important impact on human decision-making as such, if we hope to build agents that behave like humans then we must incorporate emotions into our design. Also, even if we don't want to build human-like agents, emotions can still be helpful as they serve as an efficient way to prioritize an agent's multiple goals. In this way they can reduce the computational load of an otherwise rational agent.

Emotional Belief-Desire-Intention Model

The original Belief-Desire-Intention model was developed by the philosopher Michael Bratman (Bratman 1987). It has been a very successful model and reflects human's practical reasoning in some sense. However, it does not take into account the effect of emotions on desires or beliefs (and emotions do have influence on beliefs (Frijda, Manstead, & Bem 2000)). By adding the idea of primary and secondary emotions (Damasio 1994) we can filter the decision making process of an agent.

Main Components and Functions

According to (Wooldridge 2001), practical reasoning involves two important processes: deciding what state of affairs we want to achieve— deliberation, and deciding how we want to achieve this state of affairs—means-ends reasoning. In the EBDI model, we still follow this processes. We divide the process into four components: Emotion, Belief, Desire and Intention, and we connect these four components through some main functions.

Since no real agent has unlimited memory and can deliberate indefinitely, we assume that *resource bounds* are applicable to all above items, which means there are fixed amount of memory and a fixed processor available to carry out its computations. It also means that means-ends reasoning and deliberation must be carried out in a fixed, finite number of processor cycles, with a fixed, finite amount of memory space. For our model we assume that all emotions, beliefs, desires and intentions are stored according to some priority. For all processes, the emotion, belief, desire or intention with the highest priority is considered first, and then next. During information updating, if the resource bound is reached then the one with the lowest priority will be removed or replaced.

Emotion: There is no standard definition for emotions. (Kleinginna & Kleinginna 1981) mentions that there are as many as 92 different definitions in the literature. Here we use the one that defines emotions as conscious states (LeDoux 1994). For this component, we did not limit the representation method of emotions, which can be stored as first-order logic, multidimensional logic (Gershenson 1999), some numerical measurement method as PAD emotion scales (Mehrabian 1998), or something else.

In this model, we follow the idea of (Damasio 1994) and take into account both primary emotions and secondary emotions: Primary emotions are those that we feel first, as a first response to a situation; Secondary emotions appear after primary emotions, which may be caused directly by primary emotions, or come from more complex chains of thinking. These processes are described in emotional update functions.

Belief: Belief is usually defined as a conviction to the truth of a proposition. Beliefs can be acquired through perception, contemplation or communication. In the psychological sense, belief is a representational mental state that takes the form of a propositional attitude. Knowledge is often defined as justified true belief, in which the belief must be consid-

ered to correspond to reality and must be derived from valid evidence and arguments. However, this definition has been challenged by the Gettier problem (Gettier 1963) which suggests that justified true belief does not provide a complete picture of knowledge. Still, we believe the component of belief in the original BDI model is enough to cover the idea of resources added by David Pereira (Pereira *et al.* 2005), that is, the resources mentioned can actually be looked as kind of beliefs.

Practically, beliefs are subjective for humans. The original BDI model gets its beliefs based on see function, which perceive from the environment objectively. In our model, beliefs are influenced by the agent's emotional status and, instead of acquiring beliefs through perception only, we also add the alternative methods to acquire beliefs through contemplation and communication.

Desire: Desires point to the options that are available to the agent, or the possible courses of actions available to the agent. Desires are obtained through an *option generation function*, on the basis of its current beliefs and current intentions.

Intention: Intentions play a crucial role in the practical reasoning process, because they tend to lead to action. (Wooldridge 2001) summarizes four important roles in practical reasoning, here we modify them to five after taking into account of emotions:

- *Intentions drive means-ends reasoning.* Like human being, once an agent has formed an intention, it will attempt to achieve the intention and decide how to achieve it; if one particular course of action fails to achieve an intention then the agent will typically try others.
- *Intentions persist*. An agent will not give up its intentions until the agent believes that it has successfully achieved them; it is impossible to achieve them; or the reason for the intention is no longer present.
- *Current emotions influence the determination of intentions.* BDI agents determine their intentions based on their desires and beliefs. If the available options are equally reasonable then a human making the decision might rely on emotions. Some researches (Camerer, Loewenstein, & Rabin 2003) point out that deliberative decision making may actually be driven by emotions, since research shows that people with relatively minor emotional impairments have trouble making decisions. In our model, emotions set the priority of desires and help decide intentions.
- *Intentions constrain future deliberation.* In other words, an agent will not entertain options that are inconsistent with the current intentions.
- *Intentions influence emotions upon which together with beliefs future practical reasoning is based.* Once an agent adopts an intention it can plan for the future on the assumption that it will achieve the intention. Based on that, if there is some belief that the agent cannot benefit from the intention, the agent will feel unhappy, and this emotion will also influence future beliefs and reasoning. BDI agents avoid *intention-belief inconsistency* (Bratman 1987)—the status of having an intention to bring about φ while believing that the agent will not bring about φ, because it is irrational. In contrast with the original BDI model, our model does not avoid intention-belief inconsistency completely, since intention did not influence beliefs directly but indirectly through emotions. For example, when agent i has an intention to bring about φ, it is possible to have a belief that φ will not be true in our model: assume that this belief is obtained through a message from j, and i has some negative emotion on j, such that i does not care about the belief very much. Later on, if such emotion becomes very strong, then i may remove this belief with the limitation of the resource bound; or if i's emotion on i changes to some positive one, and the belief becomes subjectively important to i, then the intention might be canceled. Thus, using emotions as a tool to balance intention and belief, such inconsistencies can be solved naturally.

More specifically, intentions in our model represent the agent's current focus—those states of affairs that it has committed to trying to bring about, and are affected by current emotional status together with current desires and working intentions.

We now formally define an EBDI architecture. Let E be the set of all possible emotions; B be the set of all possible beliefs, D be the set of all possible desires, and I be the set of all possible intentions. Thus, the state of a EBDI agent at any given moment is given by its current set of emotions, beliefs, desires, and intentions. These components are connected via the following functions:

Belief Revision Functions: Beliefs can be acquired through perception, contemplation or communication, unlike the original BDI model which uses only perception. We define three belief revision functions which map input from these three areas into beliefs. The input from perception is treated the same as in the BDI architecture. Input from communication is treated similarly but we take into consideration the identity of the sender. The input from contemplation comes from the agent's beliefs themselves and from deliberation. As with human beings, the beliefs related to the current intentions will be given higher priority and the rest will be given lower priority or ignored. For convenience, we combine effects of emotions and intentions together since they involve some common issues about the priority.

The three belief revision functions are defined as follows:

Belief Revision Function through perception ($brf\text{-}see$) generates belief candidates from the environment:

$$brf\text{-}see : Env \rightarrow B_p$$

where Env denotes the environment, $B_p \subseteq B$ is the set of possible belief candidates from perception.

Belief Revision Function through communication ($brf\text{-}msg$) generates belief candidates from the content of communication messages:

$$brf\text{-}msg : Cont \rightarrow B_m$$

where $Cont$ denotes the content of possible communication messages, B_m the set of possible belief candidates from message, and $B_m \subseteq B$.

Belief Revision Function through contemplation ($brf\text{-}in$) takes into consideration the current emotion status and intentions, and revises the current beliefs upon previous beliefs and the set of belief candidates from environment and communication messages:

$$brf\text{-}in : E \times I \times (B \cup B_p \cup B_m) \rightarrow B$$

Emotion Update Functions: We take into account both primary emotions and secondary emotions (Damasio 1994), correspondingly, we have two update functions.

Primary emotions are those that we feel first, as a first response to a situation. Thus, if we are threatened, we may feel fear. When we hear of a death, we may feel sadness. They appear before conscious thought and are thus instinctive or reactive functions of the human brain. When time is limited and we do not have enough time to think about something clearly, primary emotions become extremely useful in decision making. In agents, we can use primary emotions to speed up decision making similarly. Thus, the primary emotion update function ($euf1$) can be defined as:

$$euf1 : E \times I \times (B_p \cup B_m) \rightarrow E$$

Secondary emotions appear after the primary emotions. They may be caused directly by them, for example where the fear of a threat turns to anger that fuels the body for a fight reaction. They may also come from more complex chains of thinking. For agents, the secondary emotions come from the result of further deliberation and can replace the primary emotions. They are used to refine the decision making if time permits. The secondary emotion update function ($euf2$) is defined as:

$$euf2 : E \times I \times B \rightarrow E$$

Option Generate Function: This function is similar with the one in BDI model. The option generate function ($options$) is defined as:

$$options : B \times I \rightarrow D$$

Filter Function: This function is also similar with the one in BDI model, however we add emotions which are used to find the best option(s). The filter function ($filter$) is defined as:

$$filter : E \times B \times D \times I \rightarrow I$$

Plan Function: The functions above complete the process of deliberation and generate the best option(s)—intention(s), which can be some actions or states of mind. These intentions drive means-end reasoning, which can be represented by the plan function ($plan$):

$$plan : I \times Ac \rightarrow \pi$$

where Ac is the set of possible actions that the agent can do, and π denotes a plan which is a sequence of actions

$$\pi = (\alpha_1, \cdots, \alpha_n)$$

where α_i is an action, and $\alpha_i \in Ac$.

Plan Execution Function: This function is used to execute the sequence of actions produced by plan function. It is represented as

$$execute : \pi \rightarrow Env$$

During the execution, if π is empty, or the intention currently worked on is succeed, or the agent finds that the intention currently worked on is impossible to achieve, then this function will be terminated.

Architecture

Figure

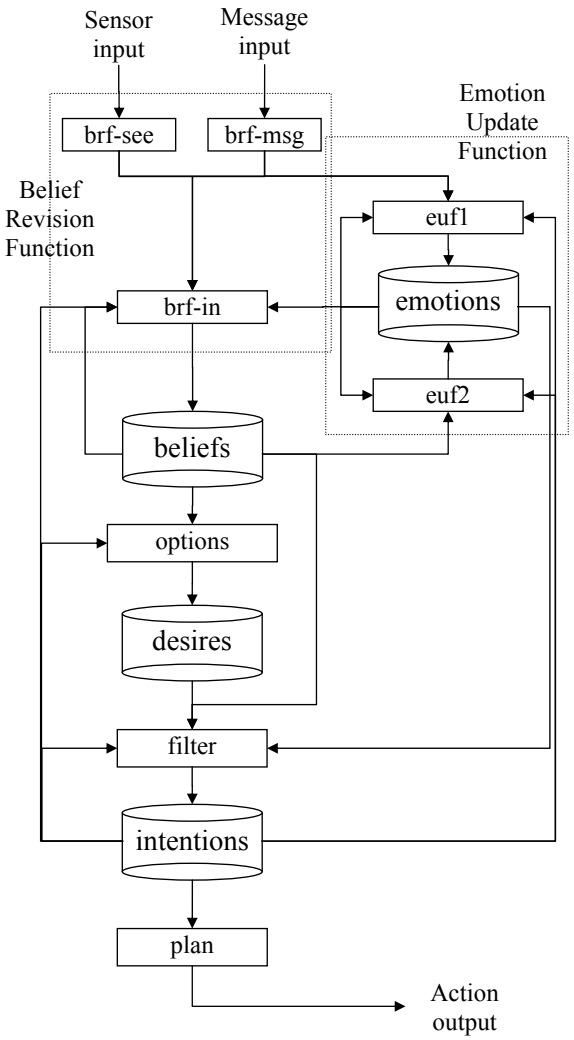

Figure 1: Schematic diagram of EBDI architecture.

Figure

We can summarize the execution cycle as follows:

1. When there is some new information from the environment via sensor or communication messages, the EBDI agent generates belief candidates;

EBDI-MAIN-LOOP

1 $E \leftarrow E_0$; $\triangleright$ E_0 are initial emotions
2 $B \leftarrow B_0$; $\triangleright$ B_0 are initial beliefs
3 $I \leftarrow I_0$; $\triangleright$ I_0 are initial intentions
4 **while** true
5 **do** $B_p \leftarrow brf\text{-}see(Env)$;
6 $B_m \leftarrow brf\text{-}msg(Cont)$;
7 $E \leftarrow euf1(E, I, B_p \cup B_m)$;
8 $B \leftarrow brf\text{-}in(E, I, B \cup B_p \cup B_m)$;
9 $D \leftarrow options(B, I)$;
10 $I \leftarrow filter(E, B, D, I)$;
11 $E' \leftarrow E$
12 $E \leftarrow euf2(E, I, B)$;
13 **if** time permits **and** $E \neq E'$
14 **then** $B \leftarrow brf\text{-}in(E, I, B)$;
15 $D \leftarrow options(B, I)$;
16 $I \leftarrow filter(E, B, D, I)$;
17 $\pi \leftarrow plan(I, Ac)$;
18 $execute(\pi)$

Figure 2: Pseudo-code of an EBDI agent's main loop.

2. These belief candidates together with current intentions trigger emotion updating, that is, the agent obtains its first feeling about the information;
3. Based on the new emotion status and the new information, together with current intentions as a guide, the agent re-evaluates its beliefs;
4. From the beliefs and intentions, the agent generates desires;
5. Under influence of the emotions, the agent chooses the best options or intentions based on current beliefs, desires and intentions. Notice that, since intentions persist, current working intention always have the highest priority unless they are already achieved or are found impossible to achieve, or the reason for this intention is no longer present.
6. From this deliberation result, the secondary emotions are triggered, and this updating is based on current intentions, beliefs and previous emotions.
7. If there is no time for deeper consideration or emotion status is not changed, the agent will directly go to generate detail plan and execute it. Otherwise, the agent gets into deeper consideration and refines the decision making. It will reconsider if current beliefs are suitable, as in line 14, and reconsider the desires and intentions, as in line 15 and 16. After this reconsideration, the agent then generates a plan and executes it.

Our EBDI agent architecture thus manages to integrate emotions into the standard processing loop of a BDI agent.

Example: EBDI Agent in Tileworld

The Tileworld system, an experimental environment for evaluating agent architectures, is introduced (Pollack & Ringuette 1990). We chose Tileworld as a platform for experimentally investigating the behavior of various meta-level reasoning strategies since we can assess the success of alternative agent strategies in different environments by tuning the environmental parameters. The Tileworld can be seen in Figure

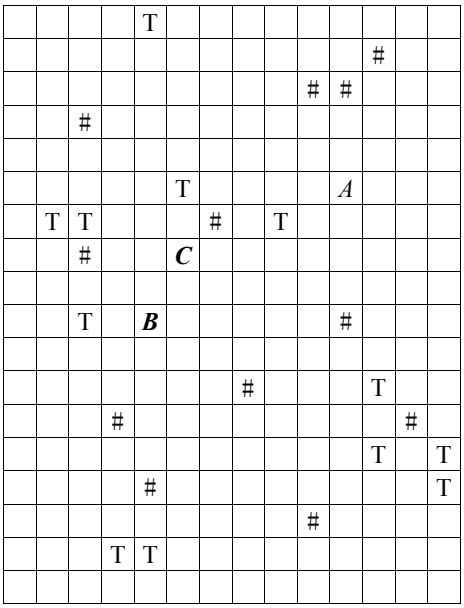

Figure 3: Simulation of Tileworld, *where* a, b, c *denote agents, T denotes tile, and # denotes hole.*

Instead of consisting of only one simulated robot agent and a simulated environment, as in (Pollack & Ringuette 1990), we design several agents and show how to apply our EBDI model to one agent in a specific Tileworld where both the agents and the environment are highly parameterized.

A Tileworld with EBDI agent is described as follows:

Environment: It is both dynamic and unpredictable. There are holes and tiles, which appear with some probability at randomly selected location, live for some time and then disappear.

Agents: The task for agents is to push the tiles to cover as many holes as possible. For each agent i, it tries to obtain the highest utility of its own $u_i = num\text{-}holes\text{-}filled_i$, where $num\text{-}holes\text{-}filled_i$ denotes the number of holes filled by agent i in the environment.

We make following assumptions for agents in the system:

- At each time step, an agent can only change its facing direction once or move one step.
- They can move along four directions: north, south, east or west, and they can move to a direction only when they face it.
- An agent can only see holes and tiles along one direction, however it can change its facing direction. For example,

when an agent faces north, it can see all the holes and tiles in front of it, but can not see the holes and tiles at the east, west or south.

- Agents can communicate with each other about what they see, but they may not tell the truth.
- An agent can save information about the location of holes and tiles that it sees or gets from communication messages as its beliefs.
- Considering the resource bound, we limit the storage space of beliefs for each agent, such that each agent can't store all information it see or information from communication message.

To make a comparison between EBDI agent and rational agent, we design three agents in the system: one EBDI agent and two rational BDI agents. These three agents have the same basic strategy in how to choose tiles to cover holes, which is described as follows:

1. First ask from each other by telling them its location;
2. Look around to get information about tiles and holes;
3. Deal with request from other agents;
4. Figure out the closest tile using beliefs;
5. Move toward the closest tile;
6. Once the agent reaches the tile, asking from each other by telling them its location;
7. Look around to get information about tiles and holes;
8. Deal with request from other agents;
9. Find out the closest hole, and move the tile to the hole.
10. Repeat step 1-9.

The differences between the three agents are described as follows:

- EBDI agent a: has a specific state to store emotion status, which is initially set to neutral. According to the basic strategy, if it decides to choose a belief told to it by some other agent i, then it will take actions just like the strategy mentioned. If it finds out later the hole or tile that the agent i mentioned is not there, it will think that i lies to him, and hate i; on the other hands, if it finds out the information is correct, it will feel thankful to i. Later on, when agent A receives another message from agent i, it will correspondingly decrease or increase the priority the information from the agent i. Also, if agent a hates i to some degree it will lie to i about what it knows.
- Truth telling BDI agent b: Always tells the truth when asked about a tile or hole.
- Selfish lying BDI agent c: Always lies to other about tiles and holes so as to increase its chances of getting all the tiles into their holes.

Let's focus on the EBDI agent a, and see how to apply EBDI model to this agent in detail. We first consider the four main components.

- Assume six level emotions are used for this agent a, and these emotions are given the set

$$Em = \{hate, dislike, unhappy, happy, like, love\}$$

Then the emotion status can be represented as the set

$$E = \{(Agent_i, e) | Agent_i \in Ag, e \in Em\}$$

where Ag denotes the set of agents in the system. In this specific example, $Ag = \{a, b, c\}$. $(Agent_i, e)$ means that current agent has a emotion e on $Agent_i$. For example, in agent a's emotion status set, there is an element $\{b, hate\}$, which means a hates agent b. The initial emotion status set $E_0 = \emptyset$, which means that in the beginning the emotion status for agent a is neutral.

The above emotion set is just and example, one can use different methods to present the emotional status.

- The belief set for agent a stores the useful information about the Tileworld, which can be represented as

$$B = \{(Agent_i, obj, location) \quad | \quad Agent_i \in Ag, \\ obj \in \{tile, hole\}, \\ location = (x, y)\}$$

where $x \in [west\text{-}bound, east\text{-}bound], y \in [north\text{-}bound, south\text{-}bound]$. The constants $west\text{-}bound$, $east\text{-}bound$, $north\text{-}bound$, $south\text{-}bound$ are integers, and they satisfy $west\text{-}bound < east\text{-}bound$, $north\text{-}bound < south\text{-}bound$, which sets the boundary of the Tileworld environment where agents can move. $Agent_i$ is the agent who sends the message. If the belief comes from perception, then $Agent_i$ will be the agent a itself. If the belief is from the communication message sent by agent b, then $Agent_i = b$. For example, belief $(c, tile, (3, 5))$ means that the agent obtained a belief from agent C that there is a tile at location $(3, 5)$. The initial belief set $B_0 = \emptyset$.

- The desire set stores the agent's current desires (goals). For example, $find\text{-}tile$ is the desire to find a tile, $find\text{-}hole$ represents the desire to find a hole, and $carry\text{-}tile\text{-}to\text{-}hole(l)$ represents the desire to carry a tile, which we assume the agent is carrying, to a hole at location l. The EBDI agent also has plans associated with each one of these desires. For example, a $find\text{-}tile$ desire can be satisfied by either searching the space or asking other agents if they have seen any tiles. Since the agents start out with no tiles, initially they have $D_0 = find\text{-}tile$.
- Intention is the agent's currently executing plan. Initially $I_0 = \emptyset$.

We now consider the main functions for agent a:

- There are three belief revision functions:

  $brf\text{-}see$ gets the belief candidates from perception. For example, if the agent is located at $(3, 5)$ and it faces east, then the agent can see all the tiles and holes locate at $(x, 5)$, where $x \in (3, east\text{-}bound]$. Assume there is a hole at $(6, 5)$, then the agent a obtains a belief candidate as $(a, hole, (6, 5))$.

  $brf\text{-}msg$ obtains the belief candidates from communication messages. Assume agent a asked b about the closest hole to location $(3, 5)$, where a is located, and b returns a message to a that the tile at $(4, 4)$ is the closest one to $(3, 5)$ based on its beliefs. Then the agent a gets a belief candidate as $(b, hole, (4, 4))$.

$brf\text{-}in$ considers current emotion status and intention as a guide to revising the belief set. For example, assume $B = \{(a, hole, (8, 5))\}$, $B_p = \{(a, hole, (6, 5))\}$, and $B_m = \{(b, hole, (4, 4))\}$, if both the emotion status set and intention set are empty, the belief set will be

$$B = \{(b, hole, (4, 4)), (a, hole, (6, 5)), (a, hole, (8, 5))\}$$

which order the beliefs rationally according to the distance to the agent's current location $(3, 5)$, such that the front one is with the highest priority; If the current emotion status set has a member $(b, hate)$, which lowers the priority of belief $(b, hole, (4, 4))$, and gets result

$$B = \{(a, hole, (6, 5)), (a, hole, (8, 5)), (b, hole, (4, 4))\}$$

If the intention set is not empty and the intention is to reach a tile at $(8, 4)$, then the resulted belief set will be

$$B = \{(a, hole, (8, 5)), (a, hole, (6, 5)), (b, hole, (4, 4))\}$$

because the agent's future location will be around $(8, 4)$, and the hole at $(8, 5)$ will be the closest one by then.

- There are two emotion update functions:
  $euf1$ considers the primary emotions. For example, if $E = \emptyset$, and $I = \emptyset$, there are $B_p = \{(a, hole, (6, 5))\}$, and $B_m = \{(b, hole, (6, 5))\}$, though the agent a gets duplicate information about the hole at $(6, 5)$, it finds out b tells the truth, and think b is reliable, and then feels happy with b. Thus, an emotion status will be generated as $(b, happy)$. If there is already a emotion status $(b, happy)$ in set E, then the emotion status in E will be updated to $(b, like)$.
  $euf2$ considers secondary emotions. It works like $euf1$ but it uses the current beliefs and intentions.
- The $options$ function generates new desires based on the agent's current beliefs and intentions. In this example this function mostly serves to generate a new $find\text{-}tile$ desire after the agent drops its current tile.
- The filter function makes a decision on the intention. For example, if the current intention of a is to find a tile at $(6, 5)$, and this information is originally from agent c,
  $$I = \{find\text{-}tile(c, (6, 5))\}$$
  Assume a is currently at $(6, 1)$, and there is emotion status set
  $$E = \{(c, hate), (b, love)\}$$
  and there is $find\text{-}tile(b, (5, 1))$ in D, then the agent a will change the intention to
  $$I = \{find\text{-}tile(b, (5, 1))\}$$
- The plan function generates a sequence of actions based on the intentions. The possible action set in this example is
  $$Ac = \{turn(direction), move(direction)\}$$
  where $direction \in \{west, east, north, south\}$
  For example, if the agent a is currently located at $(6, 1)$, faces east, and the current intention is to reach a tile at $(6, 5)$, then it generates a sequence of actions:
  $$\pi = (turn(south), move(south), move(south), move(south))$$
- For the plan execution function, basically, the agent just follows the sequence of the actions π. Note that every time the agent turns to some direction it can see some new tiles and holes which can trigger the agent's reconsideration.

The above example shows how might build an EBDI agent for the Tileworld. Based on above descriptions about the main components and main functions, the main execution cycle can just follow the interpreter as in Figure

Conclusion

We have presented the EBDI architecture which incorporates emotions into a BDI architecture. Its main features are:

- Our EBDI model takes into account both primary emotions and secondary emotions into decision making process, which reflects human's practical reasoning better. Primary emotions reflect the first response to a situation; Secondary emotions show the result of deeper thinking.
- It uses primary emotions as a first filter to adjust priority of beliefs such that agents with limited resources can speed up decision making. Secondary emotions are used to refine the decision when time is permitted.
- Instead of just considering beliefs from perception as in the BDI model, we also add belief candidates from communication.

We also provided a simple example EBDI agent for the Tileworld domain.

We believe that the EBDI framework will be instrumental in the development of emotional agents. There are many models of human emotion and its effect on decision-making. We hypothesize that our EBDI architecture should be flexible enough to incorporate any one of those models thus changing a traditional BDI agent into an emotional agent. We are currently working on mapping particular emotional models into EBDI so that we might test this hypothesis. More generally, we believe that emotional agents will contribute greatly to research on multiagent systems and agent-based modeling. Namely, by virtue of modeling human actions and emotions, emotional agents will be much more effective at interacting with humans. Also, behavioral economics has shown that people's seemingly irrational behavior often helps the group, even at the expense of the individual. We believe that multiagent systems built with emotional agents will be able to arrive at solutions of higher social welfare than those arrived at by traditional selfish agents.

References

[Bates 1997] Bates, J. 1997. The role of emotion in believable agents. *Communications of the ACM* 37(7):122–125.

[Bechara 2004] Bechara, A. 2004. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition* 55:30C40.

[Bratman 1987] Bratman, M. E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

[Camerer, Loewenstein, & Rabin 2003] Camerer, C. F.; Loewenstein, G.; and Rabin, M., eds. 2003. *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.

[Damasio 1994] Damasio, A. R. 1994. *Descartes' Error : Emotion, Reason, and the Human Brain*. New York: Avon Books.

[Davidson, Scherer, & Goldsmith 2002] Davidson, R. J.; Scherer, K. R.; and Goldsmith, H. H., eds. 2002. *Handbook of Affective Sciences*. New York: Oxford University Press.

[Dyer 1987] Dyer, M. G. 1987. Emotions and their computations: Three computer models. *Cognition and Emotion* 1(3):323–347.

[Ekman & Davidson 1994] Ekman, P., and Davidson, R. J., eds. 1994. *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press.

[Frijda & Swagerman 1987] Frijda, N., and Swagerman, J. 1987. Can computers feel? theory and design of an emotional system. *Cognition and Emotion* 1(3):235–257.

[Frijda, Manstead, & Bem 2000] Frijda, N. H.; Manstead, A. S. R.; and Bem, S., eds. 2000. *Emotions and Beliefs: How Feelings Influence Thoughts*. New York: Cambridge University Press.

[Georgeff & Ingrand 1989] Georgeff, M. P., and Ingrand, F. F. 1989. Decision-making in an embedded reasoning system. In *IJCAI'89: Proceedings of the 11th International Joint Conference on Artificial Intelligence*, 972–978.

[Gershenson 1999] Gershenson, C. 1999. Modelling emotions with multidimensional logic. In *NAFIPS '99: Proceedings of 18th International Conference of the North American Fuzzy Information Processing Society*, 42–46.

[Gettier 1963] Gettier, E. L. 1963. Is justified true belief knowledge? *Analysis* 23:121–123.

[Hustadt *et al.* 2001] Hustadt, U.; Dixon, C.; Schmidt, R. A.; Fisher, M.; Meyer, J.-J. C.; and van der Hoek, W. 2001. Verification within the KARO agent theory. In Rash, J. L.; Rouff, C. A.; Truszkowski, W.; Gordon, D.; and Hinchey, M. G., eds., *Proceedings of the First International Workshop on Formal Approaches to Agent-Based Systems (FAABS 2000)*, volume 1871 of *LNAI*, 33–47. Springer.

[Kleinginna & Kleinginna 1981] Kleinginna, P., and Kleinginna, A. 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion* 5:345–379.

[LeDoux 1994] LeDoux, J. E. 1994. Emotion, memory and the brain. *Scientific American* 50–57.

[Mehrabian 1998] Mehrabian, A. 1998. Correlations of the pad emotion scales with self-reported satisfaction in marriage and work. *Genetic, Social, and General Psychology Monographs* 124:311–334.

[Meyer 2004] Meyer, J.-J. C. 2004. Reasoning about emotional agents. In *Proceedings of ECAI 2004*. IOS Press.

[Morgado & Gaspar 2005] Morgado, L., and Gaspar, G. 2005. Emotion based adaptive reasoning for resource bounded agents. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 921–928. New York, NY, USA: ACM Press.

[Ortony, Clore, & Collins 1988] Ortony, A.; Clore, G. L.; and Collins, A., eds. 1988. *The Cognitive Structure of Emotions*. Cambridge, UK: Cambridge University Press.

[Padgham & Taylor 1996] Padgham, L., and Taylor, G. 1996. A system for modelling agents having emotion and personality. In *PRICAI Workshop on Intelligent Agent Systems*, 59–71.

[Parunak *et al.* 2006] Parunak, H. V. D.; Bisson, R.; Brueckner, S.; Matthews, R.; and Sauter, J. 2006. A model of emotions for situated agents. In *AAMAS'06: Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (submitted)*.

[Pereira *et al.* 2005] Pereira, D.; Oliveira, E.; Moreira, N.; and L.Sarmento. 2005. Towards an architecture for emotional bdi agents. In *EPIA '05: Proceedings of 12th Portuguese Conference on Artificial Intelligence*, 40–47. Springer.

[Picard 2000] Picard, R. W. 2000. *Affective Computing*. Cambridge, MA: MIT Press.

[Pollack & Ringuette 1990] Pollack, M., and Ringuette, M. 1990. Introducing the tileworld: experimentally evaluating agent architectures. In Dietterich, T., and Swartout, W., eds., *Proceedings of the Eighth National Conference on Artificial Intelligence*, 183–189. Menlo Park, CA: AAAI Press.

[Sousa 2003] Sousa, R. D. 2003. Stanford encyclopedia of philosophy: Emotion. `http://plato.stanford.edu/entries/emotion/`.

[Toda 1982] Toda, M., ed. 1982. *Man, Robot and Society*. Boston, MA: Martinus Nijhoff Publishing.

[Wegner & Wheatley 1999] Wegner, D. M., and Wheatley, T. 1999. Apparent mental causation: Sources of the experience of will. *American Psychologist* 54(7):480–492.

[Weiss 1999] Weiss, G., ed. 1999. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, Massachusetts: MIT Press.

[Wooldridge 2001] Wooldridge, M. J. 2001. *An introduction to multiagent systems*. Baffins Lane, Chichester, West Sussex, England: John Wiley & Sons, LTD.

[Wright *et al.* 1996] Wright; Ian; Sloman, A.; and Beaudoin, L. 1996. Towards a design-based analysis of emotional episodes. *Philosophy, Psychiatry, and Psychology* 3:101–126.