In Proceedings of ICCM - 2007- Eighth International Conference on Cognitive Modeling. 235 - 240. Oxford, UK: Taylor & Francis/Psychology Press.

A Belief Framework for Modeling Cognitive Agents

Annerieke Heuvelink (A.Heuvelink@few.vu.nl)
Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
TNO Defence, Security and Safety, Department of Training and Instruction
Kampweg 5, 3769 ZG Soesterberg, the Netherlands

Abstract

Simulation-based training in complex decision-making can be made more effective by using intelligent software agents to play key roles. For successful use in training, these agents should show representative behavior. Representative behavior may reflect expert behavior, but may also be far from optimal, especially under stress conditions. Current agent architectures hardly offer support to model cognitive properties that are essential to human decision-making. The present paper describes a framework in which agents beliefs are extended with additional arguments with which such dynamic cognitive properties can be formalized. An historic military event is used to demonstrate that the resulting framework is capable of modeling representative behavior.

Introduction

Organizations that operate in highly uncertain and dynamic environments, such as the military, require competent staff personal. However, the very nature of their missions makes it hard to setup real-world training. Scenario-based simulator training is considered an appropriate approach for training decision-making in complex environments (Oser, 1999). A main requirement of simulator training is that it correctly represents these aspects of the real world that are necessary to achieve the learning objectives. Perhaps the most important aspect of human decision-making is the interaction with other humans, e.g., team members. In order for simulation-based training to be an alternative of real-world training, simulated entities must be able to respond naturally and validly to any emerging situation. Therefore our goal is to develop agents that are capable of generating behavior that is representative for the human they represent.

There is growing conviction and evidence that we can develop such agents by capturing the human cognitive processes in a cognitive model. The research fields of Artificial Intelligence (AI) and Cognitive Science (CS) have yielded various architectures that can be used to develop cognitive models (Pew & Mavor, 1998).

We start this paper with describing properties of architectures that are currently used for cognitive modeling. We then elaborate on various typical features of human cognition and argue that these architectures lack the mechanisms to model these features. Next, we explain how more human-like behavior can be achieved by formalized reasoning rules on beliefs with additional arguments. We illustrate the strength of this belief framework by implementing a cognitive model of a key player in a historic military incident. Finally we draw conclusions on the significance of our work and propose future research.

Related Research

The potential and benefits of representing human behavior in (training) simulations by cognitive models of key players is generally recognized. As a result several models have been developed that can play such key roles (see e.g., Gluck & Pew, 2005). In general these models are either implemented in a cognitive architecture, like ACT-R or SOAR, or in an agent architecture such as JACK or JADEX. Cognitive architectures embody a theory of cognition, while agent architectures often encapsulate Beliefs, Desires and Intentions (BDI) (Georgeff & Lansky, 1987). For all architectures it holds that they themselves are not a model, but that they offer the constructs to build a model. The most basic construct is a declarative information entity with which the knowledge of the agent can be represented. We will refer to these knowledge entities as *beliefs*.

Since it is important for an agent to have a correct and consistent view of the world, a central issue is how an agent keeps a consistent belief database upon receiving information that is inconsistent with its current beliefs. Within AI the problem is generally solved by throwing away beliefs that cause the inconsistency. By doing so, the agent no longer has access to what it believed before, which is not very human-like. Cognitive architectures tend not to eliminate inconsistent beliefs but deal with them when they retrieve beliefs into working memory (e.g., Anderson & Lebiere, 1998; Paglieri, 2004). Mechanisms differ between architectures, but often the way they revise and retrieve beliefs is fixed. This aspect restrains the agent from having access to his old, possibly currently disbelieved, beliefs.

Research in cognitive science shows that the nature of the beliefs that form the inconsistency influence the way humans solve the inconsistency. For example, the *time* on which information is received has a large influence on the belief formation of humans. Famous temporal order effects in the updating of beliefs are primacy and recency (e.g., Anderson, 1981), which are considered to be typical human biases. Dieussaert et al. (2000) found that when a belief is deduced from a conditional statement (e.g., if A then B), that then upon receiving a categorical statement (not B) the initial *strength* of the belief in the conditional (A) is important for the revision of belief B. Another major finding is that the *source* of the beliefs is very important for how they are treated. Humans are biased to believe information that is obtained by one's own over information communicated by others. The trust of a human in the source of the information is another important factor for its believability (e.g., Mercier & Henst, 2005).

Many more cognitive biases are found in the formation of and reasoning on beliefs (Wickens & Flach, 1988). The availability bias denotes the tendency of humans to focus on the most salient outcome, which is *time* related. The confirmation bias functions on two levels; it denotes the tendency to only search for information that confirms the current hypothesis as well as the tendency to give congruent information much more weight than incongruent information. The latter strongly influences the *strengths* of beliefs. The as-if bias denotes the tendency of humans to treat *sources* 'as if' they are the same.

Cognitive biases influence the quality of human decision-making and are found to arise especially under stress conditions (see e.g., Baron, 2000). Since we want our agent to generate representative human behavior under a variety of stress conditions, we need to be able to model the before mentioned processes. Current architectures don't offer support to model (biased) reasoning over beliefs taking their initial time, their source and their certainty into account. In the next section we propose a framework with which such processes can be formalized.

We are not the first to tackle the problem of modeling biased reasoning and belief revision. However, as mentioned above most cognitive belief models adept the strengths of beliefs upon receiving new information and by doing so loose access to what was believed before. Moreover, stress is often not a factor in the revision or retrieval of beliefs.

Belief Framework

We want to develop a decision-making agent that reflects a human in the way it acts and reasons. For this goal we develop a logical framework in which beliefs represent the agent's declarative information entities. We decide to represent beliefs in predicate logic since this format enables formal verification of global properties which is useful for training. To ensure that an agent can have an up-to-date consistent belief set without loosing access to its old beliefs, we propose to *time stamp* each belief at the time it is formed with that time. With this feature it is possible to model (biased) reasoning over beliefs over time. We found only one other paper that proposes to time-stamp beliefs. Sripada (1993) took this approach in search of a more efficient belief revision technique, but only looked at binary beliefs.

We on the other hand want our agent to have graded beliefs like a human and therefore further propose to *certainty stamp* beliefs. The certainty stamp of a belief denotes the strength of the agent's belief in its truth value at the time captured in the time stamp. Last we propose to *source stamp* each belief, by which the origin of the information is captured. Using these three extra arguments various cognitive processes can be formalized as will be shown in the next sections.

Belief Predicate

A belief can be seen as a collection of properties that can be captured with the following *belief* predicate:

$\forall p\, \forall a\, \forall v\, \forall t\, \forall s\, \forall c$ [$belief(p, a, v, t, s, c) \leftrightarrow$

$\exists b\, \exists e$ [$beliefhasterm(b, e) \wedge$
$termhaspredicate(e, p) \wedge$
$termhasattribute(e, a) \wedge$
$termhasvalue(e, v) \wedge$
$beliefhastimestamp(b, t) \wedge$
$beliefhassource(b, s) \wedge$
$beliefhascertainty(b, c)$]]

The core of a belief is a term that denotes the information that is believed, e.g., that the identity (p) of track2 (a) is hostile (v). Besides this term a belief consists of three extra arguments denoting the time it was formed (t), its source (s) and how certain the agent is of that belief (c).

To formalize relations between beliefs over time it is necessary to have a reference to time that specifies the time at which a certain belief was held by the agent. For this we introduce a two-place predicate HoldsAt. When we reify the belief predicate of the object language to a propositional term b, we can state using this meta-language predicate at which time the belief is held: HoldsAt(b, t).

For every belief(p, a, v, t, s, c) that can be found in the agent's database it can stated that HoldsAt (belief (p,a,v,t,s,c), t), since the t of the belief denotes that it was then formed and thus logically holds.

Formation of Beliefs

By using the aforementioned belief system we can model relevant cognitive properties and processes. The first interesting process is the transfer of information from the outside world into a belief. Research in cognitive science mentioned above pointed out that the source of the information as well as the current state of beliefs (confirmation bias) is relevant for this process. These two aspects influence the strength with which an agent ends up believing that information, i.e., the certainty of its belief. We accommodate these aspects by transferring information from the world into a belief in three stages.

First, a *presourceexpectancybelief* is formed:

$\forall p\, \forall a\, \forall v\, \forall t\, \forall s\, \forall c$ [
$HoldsAt(input\_from\_world(p, a, v, s, c), t)$
$\rightarrow$
$HoldsAt(presourceexpectancybelief(p, a, v, t, s, c), t)$]

Secondly, the influence of the source on the believability of the given information is determined, by using the agent's trust level in that source. In how far this bias occurs, i.e., how much this process moves the perceived certainty away from the actual certainty, is influenced by the current stress level of the agent.

$\forall p\, \forall a\, \forall v\, \forall t\, \forall s\, \forall c\, \forall tr\, \forall st$ [
$HoldsAt(presourceexpectancybelief(p, a, v, t, s, c), t) \wedge$
$HoldsAt(trust\_in\_source(s, tr), t) \wedge$ $(-1 \leq tr \leq 1)$
$HoldsAt(stress(st), t)$ $(0 \leq st \leq 1)$
$\rightarrow$
$HoldsAt(preexpectancybelief(p, a, v, t, s, c + tr*c*st), t)$]

Thirdly, the current state of beliefs is taken into account. This is not done directly, but through the notion of expectancies. The *expectancy* predicate has 4 arguments, denoting the expected term (p, a, v) as well as a certainty. Expectancies differ from beliefs in that they are formed automatically and can be considered unconscious.

Expectancies are formed in two ways; each term that is currently believed gets transferred to an expectancy that will hold the next time step. The strength of the expectancy is a function of the strength of the belief and the persistence of the predicate; we will elaborate on the latter later on. Secondly, certain (combinations of) beliefs can yield new expectancies. The certainty of expectancies decays over time and the expectancy ceases to exist when its certainty becomes equal to zero.

To determine the final certainty of the belief existing congruent and incongruent expectancies are taken into account. The extent to which these expectancies bias the certainty of the agent in the final belief is influenced by the current stress level. Since multiple situations are possible multiple rules are needed to formalize this process:

$\forall p \forall a \forall v \forall t \forall s \forall c$ [
HoldsAt(preexpectancybelief(p, a, v, t, s, c), t) $\wedge$
$\neg\exists w \exists d$ [*HoldsAt(expectancy(p, a, w, d), t)*]
$\rightarrow$
HoldsAt(belief(p, a, v, t + 1, s, c), t + 1)]

$\forall p \forall a \forall v \forall t \forall s \forall c \forall d \forall st$ [
HoldsAt(preexpectancybelief(p, a, v, t, s, c), t) $\wedge$
HoldsAt(expectancy(p, a, v, d), t) $\wedge$
HoldsAt(stress(st), t)
$\rightarrow$
*HoldsAt(belief(p, a, v, t + 1, s, c + d * st), t + 1)*]

$\forall p \forall a \forall v \forall t \forall s \forall c \forall u \forall d \forall st$ [
HoldsAt(preexpectancybelief(p, a, v, t, s, c), t) $\wedge$
HoldsAt(expectancy(p, a, u, d), t) $\wedge$
$u \neq v \wedge$
$\neg \exists e$ [*HoldsAt(expectancy(p, a, v, e), t)*] $\wedge$
HoldsAt(stress(st), t)
$\rightarrow$
*HoldsAt(belief(p, a, v, t + 1, s, c – d * st), t + 1)*]

Intermediate rules (not denoted) handle new (pre)beliefs whose certainties lie outside the certainty range.

An agent can also form new beliefs using *conditional statements* and its current beliefs. These rules, together with believed *categorical statements*, make up the task specific knowledge of an agent. The formation of a new belief by a conditional statement happens in two stages. First a *preexpectancybelief* is formed, which is than transferred into a belief using the mechanisms described above. A belief formed by a reasoning rule receives that rule's name as it source. An example rule is the following:

$\forall c$ [*HoldsAt(belief(weather, local, raining, t, integratedsources, c), t)*
$\rightarrow$
HoldsAt(preexpectancybelief(status, street, wet, t, deduce_wet_from_raining, c), t)]

Note that this rule requests as input a just formed belief (denoted by t) whose source is equal to *integratedsources*.

Belief Integration

An important aspect of the belief framework is that reasoning rules request beliefs as input that have as time argument the *current time* (t) and as source argument *integratedsources* (s). The requested time argument denotes the claim that the belief should just be formed and thus holds (present in working memory) while the source denotes the claim by which rule it should be formed. The reasoning rule that produces beliefs with *integratedsources* as source argument deduces what exactly is currently believed by the agent. This rule deals with any inconsistencies in the belief set formed by beliefs from different sources or at different times. The retrieval of a belief into working memory can be seen as its human equivalent.

To implement this process we first implement the agent's memory by the following simple rule, which assumes that beliefs are never forgotten.

$\forall p \forall a \forall v \forall t' \forall s \forall c \forall t$ [
HoldsAt(belief(p, a, v, t', s, c), t)
$\rightarrow$
HoldsAt(belief(p, a, v, t', s, c), t+1)]

To facilitate the formalization of reasoning rules that use the agent's memory we introduce the *lastbelief* predicate, which denotes the most recent belief in the agent's memory for given specifications. Its definition is:

$\forall p \forall a \forall v \forall t \forall s \forall c \forall n$ [
HoldsAt(lastbelief(p,a,v,t,s,c),n)
$\leftrightarrow$
[*HoldsAt(belief(p,a,v,t,s,c),t)* $\wedge$ $t \leq n \wedge$
$\neg\exists t'$ [*HoldsAt(belief(p,a,v,t',s,c),t')* $\wedge$
$t' \geq t \wedge t' \leq n$]]]

To determine what exactly is believed by the agent, it is relevant to consider that a belief's validity over time is strongly influenced by its predicate. Values of certain predicates are much more persistent than others; consider the chance that a person's sex, marital status or mood changes over time. An agent's certainty level in a belief whose predicate is very persistent does not change much over time. However, beliefs about predicates of which the values are likely to change will quickly loose certainty. The persistence level of a predicate also influences the decaying factor of expectancies about it. The rule that determines what exactly is believed, so that is responsible of deducing the current belief from old beliefs, is formalized as:

given (p, a)
$\forall v1 \forall t1 \forall s1 \forall c1 \forall t \forall pd \forall c'$ [
HoldsAt(lastbelief(p, a, v1, t1, s1, c1), t) $\wedge$
HoldsAt(persistence_decay(p, pd), t) $\wedge$ $(0 \leq pd \leq 1)$
$\neg\exists c'' \exists v2 \exists t2 \exists s2 \exists c2$
[*HoldsAt(lastbelief(p, a, v2, t2, s2, c2)* $\wedge$
*c2 – pd * (t – t2) > c1 – pd * (t – t1)*]
$\rightarrow$
HoldsAt(belief(p, a, v1, t+1, integratedsources,
*... c1 – pd * (t – t1)), t+1)*
HoldsAt(belief(p, a, v1, t+10, integratedsources,
*c1 – pd * (t – t1)), t+10)*]

Also in this case there is an intermediate rule that handles beliefs whose certainties lie outside the certainty range.

Following this rule, the agent ends up believing the value of the belief whose certainty is the greatest after taking into account the time passed since it was formed and the persistence of the predicate. This might entail that an older belief with a higher certainty is believed over a newer belief from a different source or the other way around, it depends on the nature of predicate. The determination of the new certainty is currently kept straightforward; it is equal to the

highest one after taking the time into account. Other sources that claim the same do not contribute to its certainty.

A belief that is consciously deduced using this rule is stated to hold for ten following time points. This reflects the fact that items retrieved by humans also stay a while in working memory. The above rule takes many aspects into account and is cognitive expensive. As mentioned on page 2 humans display a bias to treat all sources as equally likely. With this simplification a decision can be made much cheaper, for example, by simply taking the most recent one. In such cases the antecedent becomes:

HoldsAt(lastbelief(p, a, v1, t1, s1, c1), t) ∧
HoldsAt(persistence_decay(p, pd), t) ∧
¬∃*v2* ∃*t2* ∃ *s2* ∃ *c2*
[*HoldsAt(lastbelief(p, a, v2, t2, s2, c2)* ∧ *t2 > t1*]

Which rule is applied is influenced by the agent's stress level and should be determined at the control level.

Reasoning over Beliefs over Time

With the given belief predicate we can deduce whether an agent believes something for a longer period of time. The *timecertaintyintegratedbelief* predicate denotes the time when the term of the current integratedsources-belief was believed for the first time. Furthermore it should hold that no other value was believed in the mean time and that it did not become unknown caused by the time passed and the decay of certainty:

given(p, a, pd)
∀*n* ∀*c* ∀*t* ∀*d* [
HoldsAt(belief(p, a, v, n, integratedsources, c), n) ∧
HoldsAt(belief(p, a, v, t, integratedsources, d), t) ∧
∀*v'* ∀*t'* ∀*c'* [
HoldsAt(belief(p, a, v', t', integratedsources, c'), t') ∧
v ≠ *v'* ∧ *t'* < *n* ∧ *t* > *t'* ∧
¬∃*t''* ∃*e* [
HoldsAt(belief(p, a, v, t'', integratedsources, e), t'')
∧ *t''* > *t'* ∧ *t''* < *t*]]
∀*t'* ∀*c'* [
HoldsAt(belief(p, a, v, t', integratedsources, c'), t') ∧
t ≤ *t'* ∧ *t'* < *n* ∧
∃*t''* ∃*e* [
HoldsAt(belief(p, a, v, t'', integratedsources, e), t'') ∧
t'' > *t'* ∧ *c'* – *pd* * (*t'*– *t''*) > 0]]
↔
HoldsAt(timecertaintyintegratedbelief(p,a,v,t),n)]

Note that this rule can be made executable by replacing the HoldsAt(b, tx) statements with HoldsAt(b, n), given that a memory system is in place. This should obviously hold for an implemented model, as presented in the next section.

This extra object predicate is useful for modeling the deduction of a belief based on the persistence of another, e.g., position stays equal → speed = 0. The predicate is also very useful to model the reasoning over belief patterns over time. E.g., to determine whether a ship zigzags the beliefs over time concerning its headings have to be integrated. The following rule depicts the principle, but should be filled with more domain specific knowledge.

given(p, a, v1, v2)
∀*t1* ∀*t2* ∀*t3* ∀*n* [
HoldsAt(timecertaintyintegratedbelief(p,a,v1,t1),n) ∧
HoldsAt(timecertaintyintegratedbelief(p,a,v2,t2),t1') ∧
t1' < *t1* ∧ ¬∃*t1''* ∃*v* ∃*t* [
HoldsAt(timecertaintyintegratedbelief(p,a,v,t),t1'') ∧
t1'' < *t1* ∧ *t1''* > *t1'*]
HoldsAt(timecertaintyintegratedbelief(p,a,v1,t3),t2') ∧
t2' < *t2* ∧ ¬∃*t2''* ∃*v* ∃*t* [
HoldsAt(timecertaintyintegratedbelief(p,a,v,t),t2'') ∧
t2'' < *t2* ∧ *t2''* > *t2'*]
→
HoldsAt(preexpectancybelief(pp,a,vp,n,this_rule,c),n)]

Case Study – Iran Air Flight 655

To illustrate our approach we present an historic case for which we developed and implemented a cognitive model of a human decision maker. It concerns the Identification Designation Supervisor (IDS) aboard the combat information center of the *USS Vincennes* cruiser that in 1988 erroneously shot down an Iranian Airbus (Fogarty, 1988). This accident has been widely referred to as an example of faulty decision-making under stress (Klein, 1998). Using this case, we want to investigate whether our approach can be used to model the behavior of the IDS-officer.

We now give a short description of the sequence of most relevant events that led to the wrong identification of the airbus by the IDS, which contributed to it being shot down. This description mixes facts about the behavior of the IDS with assumptions about his reasoning. We deduced both from the formal investigation rapport (Fogarty, 1988).

Time: 10.47 AM

• The IDS is focused on an Iranian P-3. Since the P-3 belongs to hostile country Iran and is a patrol aircraft that can guide other aircraft on hostile missions, the IDS expects hostile aircrafts.

• The radar reports a new track of interest (track2) at a range of 47nm and bearing 025, which corresponds to the runway of Iranian airport Bandar Abbas. The IDS observes the new track and based on the fact that the track's origin is an Iranian airport also used for military aircrafts, he believes it might be hostile.

• In order to determine whether the track represents a commercial aircraft, the IDS checks the Bandar Abbas commercial airline departure times schedule. However, the time of departure and scheduled time differ too much to make the neutral identification.

• In order to obtain more information the IDS sets its remote control indicator (RCI) challenge gate at the track, so it can pick up the track's Identify Friend or Foe (IFF) Mode, a system all planes are equipped with. Based on his hostile assumption he expects to receive mode II or mode III.

• The IDS picks up the neutral IFF Mode III-6675. However, all aircrafts can emit Mode III and therefore this information is not conclusive for a neutral identification.

Time: 10.48 AM

• The IDS observes from its Large Screen Display (LSD) that track2 is locked on by the USS Sides, however does not react. When military aircrafts are locked on to, they tend to change behavior. Non-military aircrafts do not notice when they are locked-on and therefore is unchanged behavior an indicator of a neutral aircraft. However, the IDS keeps believing the track might be hostile.

Time: 10.50 AM

• The IDS sees a Mode II-1100 on its RCI-display. He expected this response from the last track he queried and simply assumes that the signal comes from that track.

• Since the IDS knows that a modeII-11XX block is used by Iranian F-14's he reports track2 as 'possible F-14'.

Cognitive Model of the IDS

Our approach currently focuses on formalizing belief predicates and processes on beliefs with which we can model *how* humans process information. The formalization of w*hen* they do that has not yet been tackled. However, to simulate a cognitive model that demonstrates the former, an implementation of the latter is needed. To simulate human control we use a simple goal-directed reasoning strategy. For this strategy to work we abstracted the necessary in- and output of each rule, added the goal it contributes to and specified what satisfies that goal. For the example rule on page 3 two of these constructs would be:

Input_of_rule_goal(deduce_wet_from_raining,determine_status (street), belief_tc(local,weather,raining,integratedsources))
satisfies_goal(determine_status(street), belief_vtsc(status,stree t))

Furthermore we added backwards-reasoning rules as:

$\forall g1\ \forall p1\ \forall a1\ \forall r1\ \forall p2\ \forall a2\ \forall s\ \forall r2\ \forall g2\ \forall t$ [
HoldsAt(goal(g1), t)
HoldsAt(satisfies_goal(g1, belief_vtsc(p1, a1)), t)
$\neg\exists v \exists s \exists c$ [*HoldsAt(belief(p1, a1, v, t, s, c), t)*]
HoldsAt(output_of_rule_goal(r1, g1, belief_tsc(p1, a1)), t)
HoldsAt(input_of_rule_goal(r1, g1, belief_vtc(p2, a2, s)), t)
HoldsAt(output_of_rule_goal(r2, g2, belief_vtc(p2, a2, s)), t)
$\neg\exists v \exists c$ [*HoldsAt(belief(p2, a, v, t, integratedsources, c), t)*]
$\rightarrow$
HoldsAt(goal(g2), t)]

The main goal of the IDS-officer is to identify each track in the environment as quickly as possible in terms of *hostile*, *neutral* or *friend*. From this main goal all other relevant sub goals are determined each time step by backtracking, using the agent's task knowledge as well as its current belief state.

The model is implemented using the LEADSTO language with which temporal dependencies between two state properties can be modeled and depicted graphically (Bosse et al. 2007). The modeled dynamic properties have the following executable format: Let α and β be state properties of the form 'conjunction of atoms or negations of atoms', and *e*, *f*, *g*, *h* non-negative real numbers. In the LEADSTO language $\alpha \rightarrow_{e,f,g,h} \beta$, means:

If state property α holds for a certain time interval with duration *g*,
Then after some delay (between *e* and *f*) state property β will hold for a certain time interval of length *h*.

In the following figures traces are shown that visualize the IDS properties (on the vertical axes) over time (horizontal axes). Dark boxes on top of a line denote that the property *HoldsAt* that time, light boxes below that it does not. In all traces the certainty and persistence decay parameters range from 0-10 instead as proposed in the text from 0-1. For displaying purposes the *integratedsources* beliefs that hold for 10 timestamps are summed up in one predicate belief_t.

We lack the space to show all the reasoning steps of the IDS model, so we focus on the events of bullet 2. Figure 1 shows a trace depicting that the IDS actively observes the altitude of the track from its screen (ownCROD) and forms a belief about its value. This trace shows how the IDS's trust in his CROD (0.8) given his stress level (0.5) influences the certainty of the final belief (7 instead of 5).

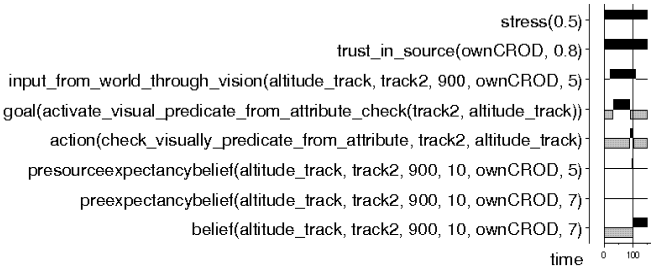

Figure 1: Observation of World and Formation of Belief

Next he reasons about the track's origin taking into account the track's position and altitude he just observed. The outcome, a belief about the airport it departed from, leads together with beliefs about the nature of that airport to a belief about the track's identity which is biased by the existing expectancy of hostile tracks (bullet 1), see figure 2.

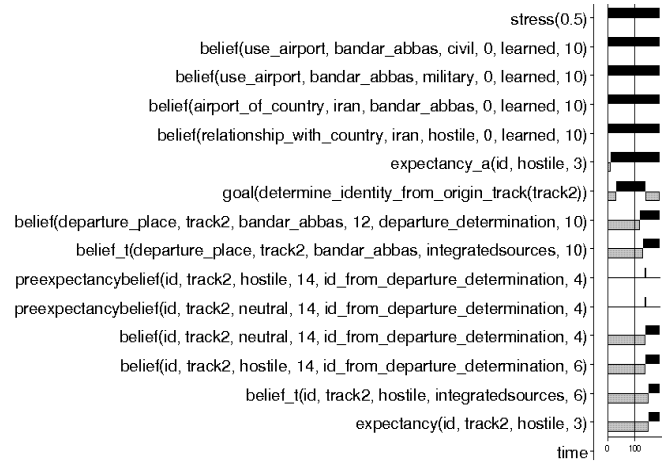

Figure 2: Formation of New Belief and Expectancy

In the following time steps the IDS performs various actions that lead to new beliefs that contribute to the reasoning about the track's identity. Unfortunately the IDS biased reasoning caused by his stress level causes him to belief he is dealing with a hostile F-16.

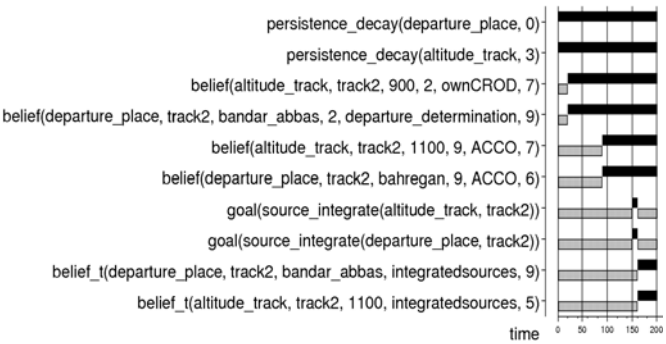

Figure 3: Source Integration on Two Predicate Types

To illustrate one important aspect of our framework a bit further we made a trace that displays the source-integration process on two different types of belief predicates: see figure 3. It can be seen that based on the nature of their predicate the beliefs are treated differently.

Discussion and Conclusion

We developed a framework for cognitive modeling based on beliefs with a *time*, *source* and *certainty* label attached. These extra labels enable the formalization of various processes on beliefs that lie at the basis of human cognition. Interactions between the time, source and certainty of beliefs has been made explicit, which is not possible in other architectures. Moreover, the influence of these parameters on each other is made tunable by the introduction of a stress level parameter.

The model of the IDS-officer shows that the framework is capable of generating human-like behavior. Agents modeled with this framework will be capable of generating more human-like behavior than, e.g., standard BDI agents. The fact that they are able to show behavior that is more representative for humans will make the agents more believable to the trainee that interacts with them. Since the believability of a training environment influences the effectiveness of the training, the modeling of agents using our framework will contribute to the effectiveness of the training and achievement of training objectives.

Our research doesn't stop here. The current framework will be extended by adding formal specifications of other relevant cognitive processes, such as attention and trust. Although the latter is already represented in the framework the current trust of an agent in sources is static. In reality however, trust is a dynamic property which is strongly influenced by experience. An agent capable of reasoning over its experiences with sources would be able to adapt its trust in sources. Stress level is another parameter that is currently fixed and that we would like to formalize as a dynamic property. Also the persistence values of properties are currently given and static, which is reasonable assuming that humans have learned them during their lifetime. However, when an agent would be capable to determine these values based on experiences with the environment, it would be much more adaptable to new environments.

As the next research step we will tackle the control of the agent. The simple control implemented in this paper was sufficient for demonstrating the reasoning rules. However, real humans have to deal with a limited amount of attention and processing power and therefore make many decisions on the control level. We like to develop a control framework in which we can capture these, probably biased, processes.

The cognitive validity of the model is debatable. However, by incorporating more outcomes of cognitive science research in our approach, we hope to approach our goal: the modeling of agents that can correctly represent human behavior in specific task training environments.

Acknowledgements

The author likes to thank Jan Treur for many fruitful discussions during this research and assisting on formal details, Tibor Bosse for clarifying aspects of the LEADSTO language and Karel van den Bosch for commenting an earlier draft.

References

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, N. H. (1981). *Foundations of Information Integration Theory*. New York, NY: Academic Press.

Baron, J. (2000). *Thinking and Deciding* (3d. edition). New York, NY; Cambridge University Press.

Bosse, T., Jonker, C. M., Meij, L. van der, & Treur, J. (2007). A Language and Environment for Analysis of Dynamics by SimulaTiOn. *International Journal of Artificial Intelligence Tools, 16(3),* to appear.

Dieussaert, K., Schaeken, W., Neys, W. de, & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Current Psychology of Cognition, 19(3),* pp. 277 – 286.

Fogarty, W. M. (1988). *Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988* (Invest. Rep. 93-FOI-0184). Department of Defense, USA.

Georgeff, M. P., & Lansky, A. L. (1987). Reactive Reasoning and Planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence* (pp. 677 – 682). Menlo Park, Ca: AAAI Press.

Gluck, K. A., & Pew, R. W. (Eds.) (2005). *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

Mercier, H. & der Henst, J.-B. V. (2005). The source of beliefs in conflicting and non-conflicting situations. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1495 – 1500). Mahwah, NJ: Lawrence Erlbaum Associates.

Oser, R. L. (1999). A structured approach for scenario-based training. In *Proceedings of the 43rd Annual Meeting of the HFES* (pp. 1138 – 1142). Santa Monica, CA: Human Factors and Ergonomics Society.

Paglieri, F. (2004) Data-oriented Belief Revision: Towards a Unified Theory of Epistemic Processing. In *STAIRS 2004: Proceedings of the Second Starting AI Researchers' Symposium* (pp. 179 – 190). Amsterdam: IOS Press.

Pew, R. W., & Mavor, A. S. (1998). *Modeling Human and Organizational Behavior*. Washington, DC: National Academy Press.

Sripada, S. M. (1993). A Temporal Approach to Belief Revision in Knowledge Bases. In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications* (pp. 56 – 62). Orlando, FL; IEEE Computer Society Press.

Wickens, C. & Flach, J. (1988). Information Processing. *Human Factors in Aviation*. San Diego, CA: Academic Press.