Implementing Commitment in BDI

Bevan Jarvis
PhD Student
bevan.jarvis@postgrads.unisa.edu.au

Prof Lakhmi Jain
Lakhmi.jain@unisa.edu.au

KES Centre
School of Electrical and
Information Engineering
University of South Australia

Objective
To define a model for commitment in a Belief-Desire-Intention (BDI) context, and to demonstrate this model in an implementation.

Methodology
Constructive Research: to build an artefact that solves a domain problem and creates knowledge about how the problem may be solved.

The aim is to implement commitment using a recently developed BDI implementation, solving some limitations inherent in previous implementations.

Intention = Desire + Commitment

Intention is a core component of Bratman's theory of human practical reasoning, which is the philosophical basis of BDI. Intention, in Bratman's theory, is distinguished from desire by the addition of commitment.

Strategies of Commitment

Null Commitment: making a single attempt and giving up if that fails; or delegating to another agent with no follow-up.

Blind or Fanatical Commitment: maintaining an intention until it is achieved.

Open-minded Commitment: maintaining an intention until it is believed impossible.

Single-minded Commitment: maintaining an intention until it is achieved or is believed impossible.

The BDI Model
The Belief-Desire-Intention model is the longest-standing model of artificial intelligence using agents. All software implementations of BDI in principle follow Rao and Georgeff's reference model [Pokahr], which was first implemented as the Procedural Reasoning System (PRS).

In PRS, goals (intentions) are not made explicit but are embodied in the handling of events.

Commitment in BDI
After Rao & Georgeff and later Wooldridge, we identify four types of commitment: null, fanatical, open-minded and single-minded. We must also allow for reconsideration of goals (intentions) and for non-reconsideration, when a goal that might be reconsidered is not.

Our research indicates that in-principle limitations arise in the implementation of reconsideration, and from two areas: firstly the hierarchical (as opposed to heterarchical) nature of structured goals, and secondly the fact that goals are not made explicit in the PRS model.

GORITE – a new Implementation of BDI
GORITE [Rönnquist] differs from the standard PRS model of BDI in two ways. Firstly, it provides explicit specification of goals. Secondly it distinguishes between (possibly shared) data required for the execution of goals and private agent beliefs.

Commitment in GORITE
In current work we seek to develop a proof-of-concept implementation of commitment using GORITE.

GORITE appears to solve some of the limitations of BDI in expressing commitment. In particular, goals are defined explicitly.

Commitment and Trust
In previous work we have developed a BDI model of trust, in which trust is regarded as one agent's beliefs about the abilities, knowledge and motivations of another agent.

One possible future research direction is to explore the connection between commitment and trust. An important part of motivation is the commitment that an agent has to a particular task or role. A model of commitment in BDI could thus be used to inform the trust beliefs of one agent about another.

Previous Achievements

We have developed a BDI model of trust, in which trust is regarded as a set of guarded beliefs about another agent's beliefs (in particular knowledge), abilities, desires and intentions.

We have identified that while commitment is fundamental to the philosophy behind BDI, BDI implementations have in the past not been able to handle all aspects of commitment in a flexible and satisfactory way.

Future Directions

Since commitment is a part of intention in the BDI model, it seems natural to seek to incorporate a model of commitment into our BDI model of trust.
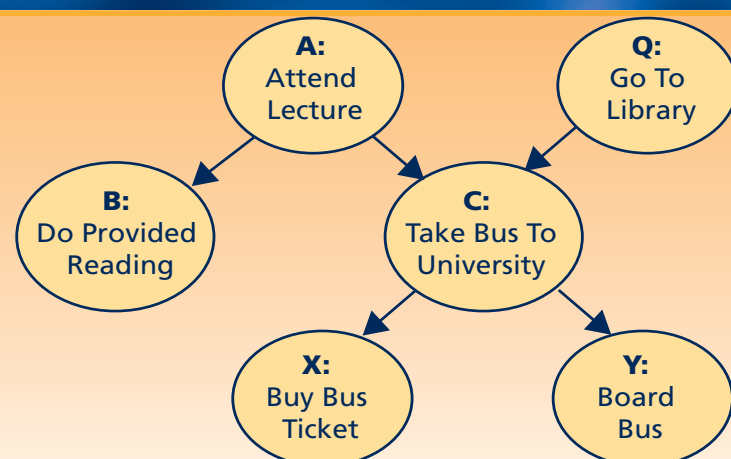

Goals and Commitment
A plan can be represented as a goal and its sub-goals in the form a goal hierarchy.

A goal may be a sub-goal of more than one other goal. In the diagram, if goal Q is dropped, goal A, and therefore sub-goal C, may still continue.

Individual sub-goals may be delegated to other agents. Commitment therefore has two aspects: that associated with the goal in the plan hierarchy, and the private commitment of an individual agent, which may be doing other tasks.
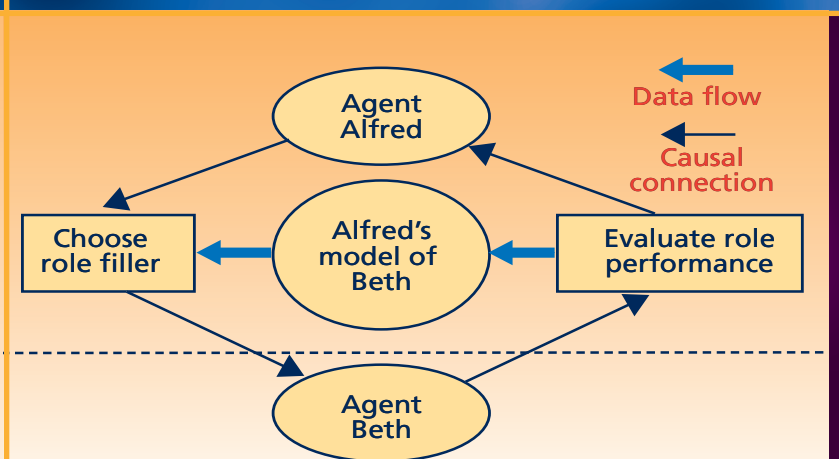
Trust
Alfred delegates (or not) to Beth depending on his trust model of her; the model is updated depending on results.

By trust we mean a set of guarded beliefs (essentially a theory) that one agent has about the knowledge, abilities and motivations of another agent. Alfred's trust model of Beth can be informed by delegating a task and evaluating Beth's performance at it. Beth may also learn about Alfred during this process. Alfred's model can also be developed in other ways, such as formal and informal meetings, observation and evaluation of the artefacts of Beth's work.