Logic Journal of IGPL Advance Access published September 9, 2009

Guido Boella, *Università degli Studi di Torino, Dipartimento di Informatica, 10149, Torino, Cso Svizzera 185, Italy.*
E-mail: guido@di.unito.it

Célia da Costa Pereira, *Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione, 26013, Crema, via Bramante 65, Italy.*
E-mail: celia.pereira@unimi.it

Gabriella Pigozzi, *Université du Luxembourg, Computer Science and Communication, L-1359, Luxembourg, rue Richard Coudenhove - Kalergi 6, Luxembourg.*
E-mail: gabriella.pigozzi@uni.lu

Andrea Tettamanzi, *Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione, 26013, Crema, via Bramante 65, Italy.*
E-mail: andrea.tettamanzi@unimi.it

Leendert van der Torre, *Université du Luxembourg, Computer Science and Communication, L-1359, Luxembourg, rue Richard Coudenhove–Kalergi 6, Luxembourg.*
E-mail: leendert@vandertorre.com

Abstract

In this paper we consider the relation between beliefs and goals in agent theory. Beliefs play three roles in reasoning about goals: they play a role in the generation of unconditional desires from conditional ones, they play a role in adoption of desires as goals, and they play a role in the selection of plans to achieve goals. In this paper we consider the role of goals in reasoning about beliefs. Though we assume that goals do not play a role in the belief generation problem, we argue that they play a role in the belief selection problem. We show the rationality of the use of goals in belief selection, in the sense that there are cases in which agents that take their goals into account in selecting a belief set from a set of alternatives outperform agents that do not do so. We also formally distinguish between the rational role of goals in belief selection and irrational wishful thinking.

Keywords: Rational agents, indeterministic belief revision, qualitative decision theory

1 Introduction

Selecting goals from beliefs is an important field in agent theory [2, 7, 27]. The opposite side, that is selecting beliefs from goals is less studied and more controversial. The aim of this paper is to demonstrate that this more controversial way of selecting beliefs can actually be used in common situations. When there is more than one belief alternative and one has

doi:10.1093/jigpal/jzp038

no reason to believe one of them more than the other, i.e., indeterminism, desires and goals may be used to break the tie. Therefore, in addition to the universally accepted fact that beliefs influence which desires are adopted as goals, one must acknowledge that desires may influence which of a number of alternative sets of beliefs is adopted. This latter influence is more controversial, and poses an interesting problem in the framework of rationality or bounded rationality, given the risk of falling into irrational behaviors such as wishful thinking.

The fundamental research problem we investigate is, therefore, how desires can influence beliefs and how that influence relates to rationality and wishful thinking. We approach this problem from the starting point of agent theory, not from belief revision or dynamic epistemic logic, although these latter do certainly play a role in our investigation. Furthermore, the problem is studied within a quite general and neutral abstract agent architecture, in an effort to make our conclusion as much architecture-independent as possible.

The problem can be broken into the following questions:

1. What is the relation between beliefs and goals, and how can goals influence belief selection?
2. What is the problem of wishful thinking for the belief selection problem?
3. How does one formalize the belief selection problem, where goals influence belief selection without wishful thinking?

The paper is organized as follows: Section 2 sets the stage for our investigation on the relationship between goal generation and belief selection; Section 3 introduces a formalization of a very general multi-agent architecture, along with an objective (i.e., external, system-wide) and a subjective (i.e., internal to an agent) view. Section 4 defines a general setting to deal with indeterministic belief revision and analyses three selection criteria for beliefs. Depending on the criterion adopted, we can have different types of agents. Section 5 discusses related work, and Section 6 draws some conclusions and discusses possible directions for further research.

2 Belief selection and goals

2.1 Beliefs, desires and goals

When designing agent architectures based on an explicit representation of mental attitudes such as beliefs, goals, intentions (as in the BDI model [26]), or obligations (as in the BOID [2] and in the B-DOING [15] models), the interactions, dependencies and possible conflicts among these different attitudes must be considered.

A mental attitude conflicts with another mental attitude if both cannot be used to generate a consistent set of goals. Thus, the agent needs to make a choice. For example, if an agent desires to go to the beach and is obliged to work, but it cannot do both at the same time, it has to make a choice: if obligations override desires, it will work.

According to the order in which mental attitudes override each other, different types of agents can be defined. As Thomason observes [28], overriding desires by beliefs reflects the fact that the agent is realistic: beliefs behave as a kind of filter on desires. Only desires consistent with the beliefs are adopted as a goal. Thomason's example is the following: if I think that it is going to rain and I believe that, if it rains, I'll get wet, I should believe that I will get wet despite the fact that I would not like to get wet. A stable agent, instead, will prefer intentions over desires, as in Thomason's BDP [28]: "I'd like to take a nap, but I intend to catch a plane, so I can't take a nap".

The role of beliefs in the generation of goals from desires, as in the decision of maintaining or dropping intentions, is well studied. In the seminal work of Cohen and Levesque [6], intentions are maintained until the agent believes that the intention is achieved or not achievable anymore. More recently, Castelfranchi and Paglieri [5] distinguished several types of beliefs depending on their roles in the generation of intentions: motivating beliefs, assessment beliefs, cost beliefs guiding the deliberation process, precondition and means-end beliefs, etc.

However, the opposite relation between motivational attitudes and beliefs is rarely explored, with the exception of, e.g., Paglieri [24]. This may be due to the fact that beliefs are seen as a logical component whose functioning is not subject to any decision based on motivational attitudes.

Consider the case of an agent who, after a new observation, has to revise its beliefs. This problem, independently of the context of agents, has been widely studied in the field of belief revision. A logical approach to belief revision cannot ensure that a unique revision candidate is selected as the new agent's belief set. As Gärdenfors [18] notices, one problem of belief revision is that logical considerations alone do not tell you which beliefs to give up, but this has to be decided by some other means.

2.2 Wishful thinking in belief selection

When the belief selection problem is addressed from the point of view of agents, the problem becomes how motivational attitudes like desires, goals, intentions and obligations are used in order to select among revision alternatives. When proposing a solution to this question, we must be careful not to fall into the problem of designing wishful thinking agents.

Wishful thinking means deciding what to believe according to what might be desirable to imagine instead of by appealing to evidence or rationality. In rhetoric it is an "argumentum ad consequentiam", where a conclusion is believed because the consequences of a premise are considered to be desirable. As observed by Thomason [28], wishful thinking is a problem in goal generation in the context of conditional beliefs and desires, when desires have the priority over beliefs in the generation of goals. For instance, in the raining example above, the agent should not conclude that it is not going to get wet only because it does not desire so.

When, in the light of a new observation, an agent has to choose among several revision alternatives, the problem of wishful thinking becomes: we should avoid the temptation to select one alternative just because it better fits the view of the world we desire most. Or else, we would lose the opportunity to reach other goals which are not satisfied by other candidate alternatives, but which can still be reached by some action.

Rather than choosing at random, or believing what is common to all the alternative revisions, the agent should choose the alternative which is most promising in terms of its possibility of actions. This is different from choosing an alternative only because it satisfies most of the agent's desires, and it involves planning and goals in the selection process.

Instead of choosing the option that most conforms to one's wishes, we propose to reason by cases and compare the outcomes of the available alternatives with the possible scenarios of the real world. Whether the agent will actually achieve its goals depends on the factual state

Table 1. Decision matrix for Example 1

Reality → ↓ Beliefs	p	$\neg p$
p	Desire $\{p\}$ is achieved	Desire $\{p\}$ is not achieved
$\neg p$	Desire $\{p\}$ is achieved	Desire $\{p\}$ is not achieved

Table 2. Decision matrix for Example 2

Reality → ↓ Beliefs	p	$\neg p$
p	Desire $\{\neg p\}$ is not achieved Desire $\{\neg p \wedge q\}$ is not achieved	Desire $\{\neg p\}$ is achieved Desire $\{\neg p \wedge q\}$ is not achieved
$\neg p$	Desire $\{\neg p\}$ is not achieved Desire $\{\neg p \wedge q\}$ is not achieved	Desire $\{\neg p\}$ is achieved Desire $\{\neg p \wedge q\}$ is achieved

or development of the world. Selecting the revision option that better adapts to its desires would not only be a short sighted decision, but can also turn out to be counterproductive. Instead, the agent should choose the (possibly not unique) revision alternative in which it is better off given all possible factual scenarios.

To illustrate wishful thinking, let us start with a very simple example.

Example 2.1.
Suppose that an agent has a choice between the belief sets $\{p\}$ and $\{\neg p\}$, and it desires $\{p\}$. We thus assume that the empty set, that corresponds to not making a choice, is not a viable alternative.

If p is in fact true, then the agent will achieve its desire p regardless of whether it has chosen to believe $\{p\}$ or $\{\neg p\}$. In other words, p is achieved because it is *factually* achieved. However, when an agent has more complex desires (see Example 2.2), and some of them require a plan to be realized, the agent needs also to be aware of what it has achieved in order to eventually execute a plan to reach additional desires.

If, on the other hand, $\neg p$ is true, the agent will never achieve its desire. Thus, as shown in Table 1, the two revision options are equivalent. To believe $\{p\}$ because p is desired (wishful thinking) is not rational, and our reasoning by cases avoids it.

Example 2.2.
Suppose that an agent has a choice between the belief sets $\{p\}$ and $\{\neg p\}$. The agent's desires are $\{\neg p \wedge q\}$ and $\{\neg p\}$.

As shown in Table 2, when the agent decides to believe $\{p\}$ and yet $\neg p$ is the case, it will achieve $\{\neg p\}$ but not $\{\neg p \wedge q\}$. This is because, since the agent believes $\{p\}$, it will not execute the plan for q. Hence, choosing $\{\neg p\}$ is the only way to generate a goal for q, and it is therefore rational. The best decision for the agent is to believe $\{\neg p\}$, because that is the only case in which it can achieve both its desires. The agent does not decide to believe $\{\neg p\}$ because it wishes $\neg p$ to be true (this would be wishful thinking). The reason for it to believe $\{\neg p\}$ is that of optimization of the possibility to achieve what it wants.

2.3 The running example

We illustrate our idea by a running example. The agent starts from a set of beliefs. Then, new information, which looks trustworthy, must be integrated with the existing beliefs. Since it is in conflict with them, and we assume that the new information should be accepted, one of the previously held beliefs must be dropped. In our example, this leads to two alternatives and the agent has no reason to prefer one over the other.

Consider a politician who wants to be re-elected and believes that:

b_1) A liberal policy leads to decrease of unemployment, and
b_2) Increasing government spending leads to decrease of unemployment,
b_3) A decrease of unemployment leads to re-election.

Therefore he executes a plan based on a liberal policy, or does something else to decrease unemployment, and secure his re-election.

Now, suppose that someone very trustworthy and well-reputed convinces him that:

b_4) A liberal policy does not lead to re-election.

Beliefs b_1, b_3 and b_4 cannot hold together, and the agent has to give up one of them. How should the agent choose among $\{b_1, b_2, b_4\}$ and $\{b_2, b_3, b_4\}$? Assume that the politician desires that his liberal view of the economy is true: he really desires that a liberal policy leads to decrease of unemployment. However, this is not a good reason to choose $\{b_1, b_2, b_4\}$ over $\{b_2, b_3, b_4\}$, and it would be an example of wishful thinking.

If he gives up b_1, then he still has another possibility to reduce unemployment, because he can increase government spending. However, if he gives up b_3, then he does not have any possibility to achieve re-election. Indeed,

1. Let us first assume that b_1 is factually wrong, whereas b_3 is true. If he chooses to retain (wrong) belief b_1 and to reject b_3, then he will do nothing and he will not succeed in being re-elected. But, had he kept his belief in b_3 and rejected b_1, then he could have increased spending in order to reduce unemployment, and therefore he could have satisfied his desire to be re-elected. To conclude, by choosing to maintain b_1, he risks to miss an opportunity to satisfy his desire. Desiring b_1 is not a good reason for choosing it: when he realizes that b_1 is false, not only does he discover that his desire is not satisfied, but also he discovers that he missed the opportunity to achieve his goal to be re-elected by using b_2.
2. Let us now assume that b_1 is actually true and b_3 is wrong. If he chooses to keep (wrong) belief b_3, then he will increase spending to be re-elected. However, even if he had chosen the right revision, i.e., to retain b_1 and reject b_3, there was no way for him to achieve his goal of being re-elected. To conclude, by choosing b_3 (wrongly), he believes that he

could achieve a goal when he could not, so he will be disappointed for trying in vain, but at least he tried.

The moral of the story is that, if our politician is interested only in achieving his goal, choosing to maintain b_3 is the only prescribed choice. This is because, independently of b_3 being right or wrong, by choosing that belief he will be at least as well off. Moreover, in one situation – the former – he will be better off if he chooses b_3 than if he chooses b_1. Summarizing, he should drop b_1, because in that way, he keeps all possibilities to achieve his goal open.

We use the re-election example as a running example throughout the paper.

2.4 A lesson from a well known fallacy

In general, the reasoning pattern based on reasoning by cases and dominance used in the running example is *not* valid. A classic example has been discussed by, amongst others, Jeffrey [21] and Thomason and Horty [29].

“The informal argument: Either there will be a nuclear war or there won’t. If there won’t be a nuclear war, it is better for us to disarm, because armament would be expensive and pointless. If there will be a nuclear war, we will be dead whether or not we arm, so we are better off saving money in the short term by disarming. So we should disarm. …The fallacy, of course, depends on the assumption that whether to arm or disarm will have no effect whether there is war or not.” [29]

For the kind of examples studied in this paper, this fallacy shows that the reasoning pattern assumes that, whether we believe one alternative or another, should not have an effect on which belief alternative factually is the case. This obviously holds in the running example, since our belief whether lowering the unemployment leads to re-election, does not influence whether this is actually the case.

3 Formalization

We distinguish between the objective and subjective view on an agent system. The former takes an external viewpoint on the system and describes mental attitudes of the agent over time. It illustrates the three roles of beliefs in reasoning about goals, but it does not assume any role of goals in reasoning about beliefs. The latter takes an internal perspective on the system and considers the decision problems of the individual agent. As illustrated by the running example, the main issue of the latter is to make decisions in the context of possibly false beliefs. For belief generation we still assume that there is no role for goals, but for belief selection the agent needs to reason by cases to find the best alternative belief set.

3.1 Agent architecture assumptions

In general, a multiagent system is a set of agent systems interacting in an environment. When an agent system gets as input an observation (either due to a sensing action or to an interrupting event, for example as feedback on its own actions, and either originating from another agent or from the environment), it processes the input, it may update its internal state, and as output it executes actions or plans. For each agent system, we assume a simple

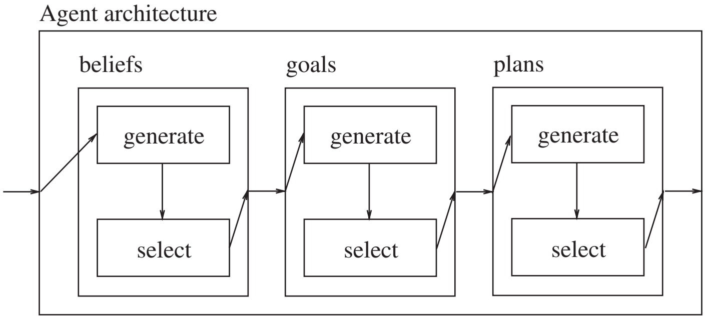

FIG. 1. Agent architecture

and standard abstract agent architecture based on an internal representation of beliefs, goals and plans, visualized in Figure 1:

Three phases. The observations are used to update the agent's beliefs, the beliefs are used to update the agent's goals, and the beliefs and goals together are used to update the agent's plans to achieve the goals. The beliefs of an agent may be wrong.

Generate-select. All three update steps are based on the generation of a set of alternatives, and a procedure to select one of these alternatives (possibly by making a random choice).

In this general framework, we make two additional assumptions.

The goals of an agent are a subset of its derived desires. We assume that the generation of goals is based on a pre-processing step that generates unconditional desires from conditional ones, and that the goals of an agent are a subset of all its derived desires. Note that there may be various reasons why an agent does not adopt a desire as a goal: it may believe the desire has already been achieved (and this belief may be false), it may believe that the desire cannot be achieved (and this belief may be false), it may believe it has more important conflicting desires it can achieve, and so on. For this reason, the distinction between desires and goals plays an important role in the formal analysis of wishful thinking, as illustrated later.

Effect of plans is achieving some goals. Since we are interested in the interaction between mental attitudes like beliefs, desires and goals, we assume that plans are represented by their effects: they partition the set of desires and goals into achieved and unachieved ones (this may be generalized to a probability to achieve the desire or goal). For the same reason, we use an abstract model of multiagent systems which does not detail other aspects of multiagent systems like norms, obligations, intentions, and so on.

In this agent theory, beliefs and goals are bounded reasoning mechanisms. In the ideal case, an agent would collect all its observations over time and do the calculations for the optimal plan at each moment. However, in the bounded reasoning case, it uses internal representations to make decisions faster. However, fast decisions may be imperfect. In particular, the reason to select belief and goal sets instead of considering all belief and goal sets simultaneously at the agent's internal representation is due to efficiency considerations.

3.2 Objective view

The objective view on multiagent systems is based on an external viewpoint, for example for model-checking multiagent systems. We start with the description of the mental attitudes, which we describe using a propositional language.

Definition 3.1 (Mental attitudes description)
Let P be a set of propositional atoms, and let the mental attitudes of an agent system be described by the tuple $\langle \mathcal{B}, \mathcal{O}, \mathcal{M}, \geq, V \rangle$ where:

- $\mathcal{B}$, $\mathcal{O}$, $\mathcal{M}$ are disjoint sets of beliefs, observations, and motivations,
- $\geq \subseteq 2^{\mathcal{M}} \times 2^{\mathcal{M}}$: a partial preorder on sets of motivations, representing an agent's preferences. For $M_1, M_2 \subseteq \mathcal{M}$, if $M_1 \supseteq M_2$, then $M_1 \geq M_2$,
- $V : \mathcal{B} \cup \mathcal{O} \cup \mathcal{M} \rightarrow 2^{2^P}$ is a function that associates with each mental attitude a set of valuations (i.e., a propositional formula).

For observations, beliefs and motivations we refer to the propositional sentence describing them by writing them in bold case, such that we may write for $b, c \in \mathcal{B}$ that $\mathbf{b} \models \mathbf{c}$, where $\models$ is logical entailment in propositional logic.

Example 3.2.
Let $\langle \mathcal{B} = \{b_1, b_2, b_3, b_4\}, \mathcal{O} = \{o_1\}, \mathcal{M} = \{m_1, m_2\}, \geq, V \rangle$ be a mental attitudes description, where $\geq$ (the preference relation) is defined as $\supseteq \cup \{(\{m_1\}, \{m_2\})\}$, i.e., the agent prefers to achieve more motivations ($\supseteq$), and it prefers $\{m_1\}$ over $\{m_2\}$.

Moreover, let $V(b_1) = p \supset u$, $V(b_2) = u \supset r$, $V(b_3) = s \supset u$, $V(b_4) = V(o_1) = \neg(p \supset r)$, $V(m_1) = r$, $V(m_2) = \neg s$, where p stands for following a liberal policy, u for a decrease in unemployment, r for re-election, and s for an increase in public spending.

A state of the system consists of facts and mental attitudes of agents.

Definition 3.3 (State of the system)
Let $\langle \mathcal{B}, \mathcal{O}, \mathcal{M}, \geq, V \rangle$ be a mental attitudes description. A state of the agent system is given by a tuple $\langle B, O, D, G \rangle$ where:

- $B \subseteq \mathcal{B}$, the agent's beliefs in this state,
- $O \subseteq \mathcal{O}$, the agent's observations in this state,
- $D \subseteq \mathcal{M}$, the agent's desires in this state,
- $G \subseteq D$, the agent's goals in this state.

Example 3.4 (Continued)
Let $\langle B = \{b_1, b_2, b_3\}, O = \{o_1\}, D = G = \{m_1, m_2\} \rangle$ be a state of the agent.

The transitions of a multiagent system are partly described by the following description of the agent behavior. It illustrates the three roles of beliefs in reasoning about goals, in desire generation ($\circledast$), goal adoption ($\bullet$) and planning for achieved goals ($\mathfrak{A}$).

Definition 3.5 (Dynamics of the system)
The dynamics of the agent system are described by a tuple $\langle *, \mathfrak{B}, \circledast, \bullet, \mathfrak{G}, \mathfrak{A} \rangle$ where:

- $* : 2^{\mathcal{B}} \times \mathcal{O} \rightarrow 2^{2^{\mathcal{B}}}$: a *belief update* function that for each agent, given an initial belief set and an observation, gives a set of alternative new belief sets.
- $\mathfrak{B} : 2^{2^{\mathcal{B}}} \rightarrow 2^{\mathcal{B}}$, a *belief selection* function from agent's alternative belief sets, to the selected belief set.

- $\circledast : 2^{\mathcal{B}} \times 2^{\mathcal{M}} \rightarrow 2^{M}$: a *desire update* function that, given an initial desire set and the new belief set, gives a new desire set.
- $\bullet : 2^{\mathcal{M}} \times 2^{\mathcal{B}} \rightarrow 2^{2^{M}}$: a *goal generation* function that, given the agent's desire set and its belief set, gives a set of alternative new goal sets. Each goal set is a subset of its desires.
- $\mathfrak{G} : 2^{2^{M}} \rightarrow 2^{\mathcal{M}}$, a *goal selection* function from the agent's alternative goal sets, to the selected goal set.
- $\mathfrak{A} : 2^{\mathcal{B}} \times 2^{\mathcal{M}} \rightarrow 2^{\mathcal{M}}$: an *achievement* function that determines which desires and goals will be achieved as a consequence of the actions and plans the agent executes.

Example 3.6 (Continued)
Let $\langle *, \mathfrak{B}, \circledast, \bullet, \mathfrak{G}, \mathfrak{A} \rangle$ be a description of the dynamics of the agent system, where

$$\begin{array}{rcl}
\{b_1, b_2, b_3\} * o_1 & = & \{\{b_1, b_3, b_4\}, \{b_2, b_3, b_4\}\}, \\
\mathfrak{B}(\{\{b_1, b_3, b_4\}, \{b_2, b_3, b_4\}\}) & = & \{b_2, b_3, b_4\}, \\
\{m_1, m_2\} \circledast \{b_2, b_3, b_4\} & = & \{m_1, m_2\}, \\
\{m_1, m_2\} \bullet \{b_2, b_3, b_4\} & = & \{\{m_1, m_2\}\}, \\
\mathfrak{G}(\{\{m_1, m_2\}\}) & = & \{m_1, m_2\}, \\
\mathfrak{A}(\{b_2, b_3, b_4\}, \{m_1, m_2\}) & = & \{m_1\}.
\end{array}$$

Given initial state $\langle B = \{b_1\}, O = \{o_1\}, D = G = \{m_1\} \rangle$, the agent ends up with beliefs $B' = \mathfrak{B}(B * O) = \{b_2, b_3, b_4\}$, desires $D' = D \circledast B' = \{m_1, m_2\}$ and goals $G' = \mathfrak{G}(D' \bullet B') = \{m_1, m_2\}$, where the achieved goals are $\{m_1\}$.

Alternatively, the agent could take the priorities between the goals into account during goal generation. The description of the agent could be

$$\begin{array}{rcl}
\{m_1, m_2\} \bullet \{b_2, b_3, b_4\} & = & \{\{m_1\}, \{m_2\}\}, \\
\mathfrak{G}(\{\{m_1\}, \{m_2\}\}) & = & \{m_1\},
\end{array}$$

such that goals $G' = \mathfrak{G}(D' \bullet B') = \{m_1\}$.

We define a simple goal generation procedure by selecting maximal consistent subsets of the desires which have not been achieved yet and which do not conflict with the agent's beliefs. Otherwise the commitment strategy of the agent would drop the goals.

Definition 3.7 (Maximal consistent goal sets)
$\bullet : 2^{\mathcal{M}} \times 2^{\mathcal{B}} \rightarrow 2^{2^{M}}$ is defined as follows.

- $cgs(B, D)$ = candidate-goal-sets(B, D) =

$$\{D' \subseteq D \mid \forall d \in D' : \mathbf{B} \not\models \mathbf{d}, \mathbf{B} \not\models \neg \mathbf{d} \wedge \mathfrak{A}(B, D') = D'\}$$

- $ccgs(B, D)$ = consistent-candidate-goal-sets(B, D) =

$$\{D'' \in cgs(B, D) \mid \mathbf{D}'' \not\models \bot\}$$

- $D \bullet B = mcgs(B, D)$ = max-consistent-goal-sets(B, D) =

$$\{D''' \in ccgs(B, D) \mid \nexists D'''' \in ccgs(B, D) : D'''' \supset D'''\}$$

Instead of selecting the maximal consistent subsets, the agent may also consider only preferred maximal consistent goal sets by replacing $D'''' \supset D'''$ by $D'''' > D'''$ in the final formula. However, this might be opportunistic when there is no plan to achieve the preferred goal set.

Example 3.8.
The goal set $\{m_1, m_2\}$ is consistent, and thus $\{m_1, m_2\} \bullet \{b_2, b_3, b_4\} = \{m_1, m_2\}$.

The generality of our model can be further illustrated by introducing various extensions in our model, such as a more detailed description of the mental attitudes, practical reasoning rules which derive the desires of the agents in each state from their beliefs, planning rules to determine which goals are achieved, and so on. However, since we are primarily interested in the interaction between belief change and goal change, we will not do that here.

3.3 Subjective view on belief selection

The subjective view on multiagent systems is based on an internal viewpoint used to design autonomous agents. In this view, an agent does not know the actual state of the world, and it does not know the beliefs, desires and goals of other agents. To analyze the decision problem formally, we describe the decision problem as concisely as possible. The minimal representation of the decision problem is as follows:

Belief and desire sets are directly represented by sets of propositional sentences, where the desires are partially ordered.

Desires can be adopted as goals, and can be achieved, which leads to a partitioning of the desires of an agent into four sets.

Context of the partitioning of the desires into four sets depends on the belief set of the agent, as well as on the belief set the agent considers to be the case.

Definition 3.9 (Subjective view on multiagent systems)
Let L be a propositional language. The belief selection problem is a pair β, γ where:

- $\beta \subseteq 2^L$ is a set of alternative sets of sentences of L,
- $\gamma : \beta \times \beta \rightarrow 2^L \times 2^L \times 2^L \times 2^L$ is a function from beliefs and facts to four disjoint sets of sentences of L.

The intuitive meaning of $\gamma(B_1, B_2)$ is what happens with the agent's desires if the agent believes B_1 when in fact the set of sentences that are satisfied by the actual state of the world is B_2. Given $\gamma(B_1, B_2) = \langle D^a, D^u, G^a, G^u \rangle$, $\langle D^a, D^u, G^a, G^u \rangle$ is a partition of D, meaning that

1. $D^a \cup D^u \cup G^a \cup G^u = D$, and
2. D^a, D^u, G^a, and G^u are mutually disjoint.

There are various ways to define these sets in the abstract agent architecture, depending on additional assumptions. For example, if the agent believes B_1 (and acts accordingly), but in fact it is B_2, then we can define γ as follows:

D^a is the set of desires not adopted as goals but achieved,

$$D^a = (D \setminus \mathfrak{G}(D \bullet B_1)) \cap \mathfrak{A}(B_2, \mathfrak{G}(D \bullet B_1));$$

D^u is the set of desires not adopted as goals and unachieved,

$$D^u = D \setminus \mathfrak{G}(D \bullet B_1) \setminus D^a;$$

G^a is the set of adopted goals that will be achieved,

$$G^a = \mathfrak{G}(D \bullet B_1) \cap \mathfrak{A}(B_2, \mathfrak{G}(D \bullet B_1));$$

G^u is the set of adopted goals that will not be achieved, $G^u = \mathfrak{G}(D \bullet B_1) \setminus G^a$.

Notice that, based on this definition, when the agent is "correct'' all the adopted goals will be achieved and none of the desires not adopted will be achieved accidentally: for all $B \in \beta$,

$$\gamma(B, B) = \langle \emptyset, D \setminus G, G, \emptyset \rangle.$$

Indeed, $G^a = \mathfrak{G}(D \bullet B) \cap \mathfrak{A}(B, \mathfrak{G}(D \bullet B)) = \mathfrak{G}(D \bullet B)$ and $G^u = \mathfrak{G}(D \bullet B) \setminus G^a = \mathfrak{G}(D \bullet B) \setminus \mathfrak{G}(D \bullet B) = \emptyset$.

The four sets of desires encode the three roles of beliefs in reasoning about goals. If $\gamma(B_1, B_2) = \langle D^a, D^u, G^a, G^u \rangle$ then:

Desire generation is the process that determines the set of $D^a \cup D^u \cup G^a \cup G^u$, which depends on the set of beliefs.

Goal adoption is the process which determines the set of adopted goals $\mathfrak{G}(D \bullet B_1) = G^a \cup G^u$.

Planning is the process which determines the desires and goals which are believed to be achieved $D^a \cup G^a$.

The running example is relatively simple, since the set of generated desires remains constant. The following example illustrates the definitions.

Example 3.10.
In this example, we show the phases of the agent cycle starting from the generation of desires and arriving to the goals which can be achieved by the actions of the agent. The different beliefs, desires, goals and plans, after generation and selection, are illustrated in Table 3, together with the specification of which desires and goals the agent believes that remain unsatisfied after the action of the agent. The generation of beliefs from goals will be discussed in Example 4.5.

We assume that the agent desires are not consistent (for readability we use their representation as formulae): $D = \{q, \neg q, a \wedge r, \neg a\}$, and, conditionally, in case a is true it desires also p while in case a is false it desires $\neg p$.

The inconsistencies among the desires in the two different situations a and $\neg a$ are resolved using preference ordering $\geq$.

Even if plans are not explicitly represented in the formalism, the reader can interpret the example as if the agent had a plan to achieve $p \wedge r$ if a is true, and $\neg p \wedge q$ otherwise. It can unconditionally reach $p \wedge q$ with a third plan.

Given the belief that a, the agent generates the set of desires taking into account the applicable conditional ones: $\{p, q, \neg q, a \wedge r, \neg a\}$. Then, to generate the goals it considers the possible consistent sets of desires $D \bullet \{a\} = \{\{p, a \wedge r, q\}, \{p, a \wedge r, \neg q\}\}$ and selects the one maximizing the preference ordering, $\mathfrak{G}(D \bullet \{a\}) = \{p, a \wedge r, q\}$, since $\{p, a \wedge r, q\} \geq \{p, a \wedge r, \neg q\}$.

TABLE 3. An example desire, goal and plan selection and generation process of an agent.

beliefs	a	$\neg a$
desires	$p, q, \neg q, a \wedge r, \neg a$	$\neg p, q, \neg q, a \wedge r, \neg a$
goals	$\{p, a \wedge r, q\}, \{p, a \wedge r, \neg q\}$	$\{\neg p, q\}, \{\neg p, \neg q\}$
selected	$p, a \wedge r, q$	$\neg p, q$
plans	$\{p \wedge r\}, \{p, q\}$	$\{\neg p, q\}$
selected	$p \wedge r$	$\neg p, q$
D^a		$\neg a$
D^u	$\neg a$	$\neg q, a \wedge r$
G^a	$p, a \wedge r$	$\neg p, q$
G^u	q	

TABLE 4. The example taking into account the objective view of the world.

world	beliefs	effect	D^a	D^u	G^a	G^u
$\neg a$	a	$\neg p, q, r$	$\neg a$	$\neg q$	q	$\neg p, a \wedge r$
a	$\neg a$	p, q, r	$a \wedge r$	$\neg a, \neg q$	q	$\neg p$

Given the set of selected goals, the agent considers which plans can be executed in the believed situation a. The goals which can be achieved by the different plans are respectively: $\{\{p \wedge r\}, \{p, q\}\}$.

Using again the preference ordering, the plan for $\mathfrak{A}(\{a\}, \mathfrak{G}(D \bullet \{a\})) = \{p \wedge r\}$ is selected.

Table 3 reports $\gamma(\{a\}, \{a\}) = \langle D^a, D^u, G^a, G^u \rangle$.

Note that this plan achieves a subset of the selected goals, but not all of them (q remains unachieved). Among the desires which were not selected, $\neg q$ is satisfied as a side effect of the actions of the agent.

A similar line of reasoning can be followed for the belief $\neg a$, as illustrated in the second row of Table 3.

Note that the goals and desires which are achieved or unachieved in reality depend not on the agent beliefs, but on the objective state of the world, which the agent eventually comes to know (for example, as a feedback from the actions it performs): for example, $\gamma(\{a\}, \{\neg a\}) = \langle D^a, D^u, G^a, G^u \rangle$

In Table 4, we illustrate the situation from the point of view of reality. Note that the generated goals and desires don't change according to the real view of the world (thus we do not repeat them), while, given the conditional character of plans, the performed actions may have not reached the effect believed by the agent.

We now turn to the role of goals in belief selection.

4 Choosing Beliefs

> "Most models of belief change are deterministic in the sense that given a belief set and an input, the resulting belief set is well-determined. There is no scope for chance in selecting the new belief set. Clearly, this is not a realistic feature, but it makes the models much simpler and easier to handle, not least from a computational point of view.

In indeterministic belief change, the subjection of a specified belief set to a specified input has more than one admissible outcome.
Indeterministic operators can be constructed as sets of deterministic operations. Hence, given n deterministic revision operators $*_1, *_2, \ldots, *_n$, $* = \{*_1, *_2, \ldots, *_n\}$ can be used as an indeterministic operator.''[19]

Let us consider an agent whose belief set is B and an observation $o \in O$. The revision of B in light of observation o is simply:

$$B * o = \{B \underset{1}{*} o, B \underset{2}{*} o, \ldots, B \underset{n}{*} o\}. \tag{4.1}$$

More precisely, revising the belief set B with the indeterministic operator $*$ in light of new observation o leads to a set of possible belief revision results

$$B * o = \{B_1, B_2, \ldots B_n\}, \tag{4.2}$$

where $B_i = B *_i o$ is the ith possible belief revision result. In general, it is possible that distinct belief revision operators yield the same result in some cases; moreover, we assume n distinct virtual operators just for the sake of illustration — in fact, a countable infinity of virtual operators may be assumed. Therefore, $\|B * o\| \geq 1$.

Applying operator $*$ is then equivalent to applying one of the virtual operators $*_i$ contained in its definition. While the rationality of an agent does not suggest any criterion to prefer one revision over the others, our framework defines an agent that will choose which revision to adopt based on the consequence of that choice. One important consequence is the set of goals the agent will decide to pursue.

By considering an indeterministic belief revision, we admit $\beta = B * o$ to contain more than one possible result. In this case, the agent must select, by means of function $\mathfrak{B}$, (possibly) one among all possible revisions, $\mathfrak{B}(\beta)$. Among the possible criteria for selection, one is to choose the belief revision operator for which the goal set selection function returns the most preferred goal set. In other words, selecting the revision amounts to solving an optimization problem.

To choose what to believe, an agent has to compare all possible combinations of beliefs and facts. Given a set $\beta = \{B_1, B_2, \ldots, B_n\}$ of alternatives, it is convenient to consider the choice matrix

$$M = \begin{pmatrix} \gamma(B_1, B_1) & \gamma(B_1, B_2) & \ldots & \gamma(B_1, B_n) \\ \gamma(B_2, B_1) & \gamma(B_2, B_2) & \ldots & \gamma(B_2, B_n) \\ \vdots & & \ddots & \vdots \\ \gamma(B_n, B_1) & \gamma(B_n, B_2) & \ldots & \gamma(B_n, B_n) \end{pmatrix}. \tag{4.3}$$

The cell on the ith row and jth columns contains the 4-tuple $M_{ij} = \gamma(B_i, B_j) = \langle D^a_{ij}, D^u_{ij}, G^a_{ij}, G^u_{ij} \rangle$. Each row of the matrix corresponds to an alternative set of beliefs.

Different types of agents can be defined by defining different criteria to compare the rows of matrix M. Every decision criterion must consist of two components: an order relation between 4-tuples, based on the preferences of the agent as formalized by the $\geq$ preorder, and an extension thereof on the rows of matrix M. A natural extension of $\geq$ to 4-tuples $\langle D^a, D^u, G^a, G^u \rangle$ is the following.

Definition 4.1.
$\langle D_1^a, D_1^u, G_1^a, G_1^u\rangle \geq \langle D_2^a, D_2^u, G_2^a, G_2^u\rangle$ iff $D_1^a \cup G_1^a \geq D_2^a \cup G_2^a$, i.e., iff the set of the achieved desires and goals in the first 4-tuple is preferred to the set of the achieved desires in the second.

We can now define a number of possible decision criteria, all based on Definition 4.1, that would enable an agent to choose the most preferred belief alternative among the elements of β.

Definition 4.2 (Wald's Criterion)
Choose the row whose least preferred cell is most preferred:

$$i^* = \arg\max_i \min_j M_{ij}.$$

Definition 4.3 (Wishful Thinking)
Choose the row which contains the most preferred of all the cells in the matrix:

$$i^* = \arg\max_i \max_j M_{ij}.$$

The third criterion, which we will call *neutral*, because it treats all alternative beliefs as equally plausible, considers all pairwise comparisons among the elements of two rows to decide which is most preferred.

Definition 4.4 (Neutral Criterion)
Given two rows R^1 and R^2, let

$$c = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij},$$

where

$$c_{ij} = \begin{cases} -1, & \text{if } R_i^1 < R_j^2; \\ 0, & \text{if } R_i^1 = R_j^2; \\ 1, & \text{if } R_i^1 > R_j^2. \end{cases}$$

Then, define $R^1 \geq R^2$ iff $c \geq 0$ and choose the most preferred row according to such definition.

Example 4.5.
The belief selection problem introduced in Section 2.3 can be formalized by defining the following atomic propositions:

p the politician is following a liberal policy;
u unemployment decreases;
r the politician will be re-elected;
s the politician is increasing public spending.

The belief set before being told that a liberal policy does not lead to re-election ($\neg(p \supset r)$) would contain the three formulas $p \supset u$, $u \supset r$, and $s \supset u$. The politician desires, first of all, to be re-elected, r, and, if possible, not to increase public spending, $\neg s$. Adding $\neg(p \supset r)$ to his beliefs would make them inconsistent. Therefore, the politician has to revise his beliefs by giving up either $p \supset u$ or $u \supset r$. His choice may depend on the goals he can achieve in the alternatives: If it gives up $p \supset u$, his plan will be to increase public spending, so he will not achieve $\neg s$, but might succeed in achieving r; if he gives up $u \supset r$, his plan will be to do nothing, so he will certainly not achieve r, but he will fulfill $\neg s$. Depending on his preference of r over $\neg s$, he could prefer one or the other alternative.

Let us first assume that the politician only desires to be elected. He has to decide between two alternatives $\beta = \{B_1 = \{p \supset u, s \supset u, \neg(p \supset r)\}$ and $B_2 = \{u \supset r, s \supset u, \neg(p \supset r)\}\}$. Therefore,

$$M = \begin{pmatrix} \gamma(B_1, B_1) & \gamma(B_1, B_2) \\ \gamma(B_2, B_1) & \gamma(B_2, B_2) \end{pmatrix} = \begin{pmatrix} \langle \emptyset, \{r\}, \emptyset, \emptyset \rangle & \langle \emptyset, \{r\}, \emptyset, \emptyset \rangle \\ \langle \emptyset, \emptyset, \emptyset, \{r\} \rangle & \langle \emptyset, \emptyset, \{r\}, \emptyset \rangle \end{pmatrix}.$$

Now,

$$\begin{aligned} \min_j M_{1j} &= \langle \emptyset, \{r\}, \emptyset, \emptyset \rangle, \\ \min_j M_{2j} &= \langle \emptyset, \emptyset, \emptyset, \{r\} \rangle, \end{aligned}$$

and, according to Definition 4.1, $\langle \emptyset, \{r\}, \emptyset, \emptyset \rangle \geq \langle \emptyset, \emptyset, \emptyset, \{r\} \rangle$. Therefore, according to Wald's Criterion, B_1 is to be chosen.

A wishful thinking agent, instead, would select the beliefs corresponding to the row of M where the most preferred cell lies. The most preferred cell here is $M_{22} = \langle \emptyset, \emptyset, \{r\}, \emptyset \rangle$. Therefore, according to Wishful Thinking, B_2 would be chosen.

Finally, by adopting the Neutral Criterion, we would have to compare Row 1 against Row 2:

$$\begin{array}{ccccccccccc} c & = & c_{11} & + & c_{12} & + & c_{21} & + & c_{22} & = & \\ & = & 1 & + & (-1) & + & 1 & + & (-1) & = & 0. \end{array}$$

The comparison of Row 2 against Row 1 is analogous, but with signs reversed, and yields the same zero result. Therefore, according to the Neutral Criterion, the two alternatives are on the same level, and neither is preferred over the other. This would be a case when indeterminism is not resolved and another method would have to be adopted to break the tie.

Let us now complicate the example a little, by assuming that, in addition, the politician desires not to increase public spending. We obtain the new choice matrix

$$M = \begin{pmatrix} \langle \emptyset, \{r\}, \{\neg s\}, \emptyset \rangle & \langle \emptyset, \{r\}, \{\neg s\}, \emptyset \rangle \\ \langle \emptyset, \emptyset, \emptyset, \{r, \neg s\} \rangle & \langle \emptyset, \{\neg s\}, \{r\}, \emptyset \rangle \end{pmatrix}.$$

Let us assume, furthermore, that $\{r\} > \{\neg s\}$, i.e., that the politician desires to be re-elected more than not to increase public spending. We would thus have

$$\begin{aligned} \min_j M_{1j} &= \langle \emptyset, \{r\}, \{\neg s\}, \emptyset \rangle, \\ \min_j M_{2j} &= \langle \emptyset, \emptyset, \emptyset, \{r, \neg s\} \rangle. \end{aligned}$$

Therefore, according to Wald's Criterion, B_1 is to be chosen, whereas, according to Wishful Thinking, B_2 would be preferred and the Neutral Criterion would again end in a tie.

However, if $\{r\} < \{\neg s\}$, i.e., if the politician desired not to increase public spending more than to be re-elected, Wald's Criterion and Wishful Thinking would agree in choosing B_1. Furthermore, comparing Row 1 against Row 2 with the Neutral Criterion would yield

$$\begin{array}{cccccccccc} c & = & c_{11} & + & c_{12} & + & c_{21} & + & c_{22} & = \\ & = & 1 & + & 1 & + & 1 & + & 1 & = & 4. \end{array}$$

Therefore, the Neutral Criterion too would choose B_1.

5 Related work

5.1 Goal change

In this paper we do not explain the process of goal generation from rules nor revision, i.e., we are not interested in how goals change in the light of new beliefs or desires. That aspect is considered, for example, in [7, 9], where an approach has been proposed to dynamically construct the goal set to be pursued by a rational agent, by considering changes in its mental state. More precisely, the authors propose a general framework based on classical propositional logic, to represent changes in the mental state of the agent after the acquisition of new information and/or after the arising of new desires. That framework was subsequently extended to take gradual trust in the sources of information into account [8]. Such models of goal generation may be regarded as complementary to the account of belief selection we have discussed above.

5.2 BOID

The BOID architecture [2] extends a classical planner with a component for goal generation. In this goal generation component, there are subcomponents for beliefs, obligations, intentions and desires [4]. The interaction among these subcomponents is studied using a qualitative decision theory [3, 13] and qualitative game theory [14] based on extensions of input/output logic [1, 22, 23]. Using merging operators [10], as an extension of the 3APL programming language [11], and using defeasible logic [12]. Though in all of these approaches the relation between beliefs and goals plays a central role, in these papers the impact of goals on the choice among belief sets has not been studied.

5.3 Data oriented belief revision

Concerning the relation between belief revision and motivational attitudes, Paglieri and Castelfranchi [25] study the gap between belief revision and argumentation. They argue that the AGM model does not take into account the reasons to believe and choose among beliefs. An alternative model of belief revision is presented by the authors, called Data-oriented Belief Revision (DBR) which has two informational categories: data and beliefs to account for the distinction between pieces of information that are simply gathered and stored by the agent (data), and pieces of information that the agent considers (possibly up to a certain degree) truthful representations of states of the world (beliefs).

In this paper we do not study belief revision, but there is a similarity in the criteria used to select data in their model: relevance, credibility, importance and likeability. Likeability, a measure of the motivational appeal of the datum, i.e. the number and values of the (pursued) goals that are directly fulfilled by that datum, directly relates to our notion of appeal. The difference is that our notion is more future-oriented to avoid the risk of wishful thinking by the agents.

5.4 Achieving higher impact

The role of motivational attitudes in the choice among belief is studied by Hunter [20] in the context of argumentation theory. He addresses the problem of selecting which arguments to convey to the audience. The criterion to select arguments is the impact they can have.

The impact of argumentation depends on what an agent regards as important. Different agents think different things are important. He assumes a desiderata base for capturing what an agent in the intended audience thinks is important, and then uses this to measure how arguments resonate with the agent. Intuitively, a desideratum (a formula) represents what an agent would like to be true in the world. There is no constraint that it has to be something that the agent can actually make true. It may be something unattainable such as “there will be no more wars" or “bread and milk is free for everyone". There is also no constraint that the desiderata for an agent are consistent.

With respect to our work there are two differences: first, the notion of impact is an heuristic used to maximize the possibility to persuade an audience. In contrast, in our framework, we propose a rationality criterion to decide what to believe. Second, since the notion of desideratum is detached from the notion of planning, the notion of impact risks to overlap with wishful thinking.

5.5 Preference over beliefs

Doyle suggests to have a preference order over belief sets [16]. We have however an indirect link from belief sets to feasible goals, and a preference order over these goals; and from these preferences over goals, we again derive the preferences over belief sets. Therefore, if one wanted to accept Doyle’s suggestion, our work could be regarded as a method for deriving a rationally justified preference order over belief sets.

5.6 Conventional wisdom

The reader could wonder whether the tendency of humans to prefer beliefs which are convenient for them has been identified in other areas. The relation between beliefs and goals described in our framework could be related with the notion of conventional wisdom of the economist John Kenneth Galbraith:

> “We associate truth with convenience, with what most closely accords with self-interest and personal well-being” [17].

That is, conventional wisdom consists of “ideas that are convenient, appealing” [17]. This is the rationale for keeping them.

Note that Galbraith distinguishes conventional wisdom from wishful thinking in the same way as we do in this paper. Conventional wisdom is related to “promises to avoid awkward efforts or unwelcome dislocation of life'' [17], i.e., with the dimension of eventual action rather than with the satisfaction of goals in the current situation.

More specifically conventional wisdom is articulated in at least three facets:

- "Associating truth with convenience — with what most closely accords with self-interest and personal well-being or promises to avoid awkward efforts or unwelcome dislocation of life'',
- "What contributes most to self esteem'',
- "Most important …people approve most of what they best understand''.

(All the above quotations are from [17].)

The model proposed in this paper relates to the first issue. Other works in the Artificial Intelligence field have some relation with conventional wisdom, like the one discussed in Section 5.4.

6 Summary and conclusions

In this paper we consider the relation between beliefs and goals in agent theory. Beliefs play three roles in reasoning about goals: they play a role in the generation of unconditional desires from conditional ones, they play a role in adoption of desires as goals, and they play a role in the selection of plans to achieve goals. In this paper we consider the role of goals in reasoning about beliefs. Though we assume that goals do not play a role in the belief generation problem, we argue that they do play a role in the belief selection problem. Using a running example where a politician has to give up either his belief that a liberal policy decreases unemployment or that decreasing unemployment leads to his reelection, we argue that in the context of an alternative plan to decrease unemployment and the goal to be reelected, it is rational to give up the belief that liberal policy leads to a decrease in unemployment.

At first sight, it may seem counterintuitive that goals play a role in reasoning about beliefs, because it may lead to wishful thinking. We show the rationality of our approach, in the sense that there are cases in which agents that take their goals into account in selecting belief sets from a set of alternatives outperform agents that do not do so. In the running example, the agent is always better off when he gives up the belief that liberal policy leads to a decrease of unemployment, because it is the only way to achieve his goal. Even if the alternative plan of the agent to decrease unemployment is based on an increase in public spending, and the agent has the desire not to increase public spending, giving up the belief that liberal policy decreases unemployment may be rational when the goal to be reelected is more important than the goal not to increase public spending.

We define an objective view on multiagent systems based on an external viewpoint to explain the three roles of beliefs in reasoning about goals. We then focus on the belief selection problem from a subjective view based on an internal viewpoint.

Moreover, we formally study the role of goals in belief selection. We first relate our subjective view to indeterministic belief revision models. We formally characterize wishful thinking in the belief selection problem, and we characterize the borderline between the rational role of goals in belief selection and irrational wishful thinking. Whereas wishful thinking maximizes the set of achieved goals (and minimizes the set of unachieved goals), we argue that the set of achievable goals should be maximized.

In future research, besides investigating the formal properties of the proposed framework, we plan on developing heuristics for belief selection in our general model, and to further investigate the notion of conventional wisdom.

References

[1] G. Boella, J. Hulstijn, and L. van der Torre. Interaction in normative multi-agent systems. *Electronic Notes in Theoretical Computer Science*, 141(5):135–162, 2005.

[2] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly Journal*, 2(3–4):428–447, 2002.

[3] J. Broersen, M. Dastani, and L. van der Torre. Realistic desires. *Journal of Applied Non-Classical Logics*, 12(2):287–308, 2002.

[4] J. Broersen, M. Dastani, and L. van der Torre. Beliefs, obligations, intentions and desires as components in an agent architecture. *International Journal of Intelligent Systems*, 20:9:893–919, 2005.

[5] C. Castelfranchi and F. Paglieri. The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. *Synthese*, 155:237–263, 2007.

[6] P. Cohen and H. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, 1990.

[7] C. da Costa Pereira and A. Tettamanzi. Towards a framework for goal revision. In: W. Vanhoof, P.-Y. Schobbens and G. Schwanen, editors, *BNAIC-06, Proceedings of the 18th Belgium-Netherlands Conference on Artificial Intelligence*, pages 99–106, Namur, Belgium, October 5-6, 2006. University of Namur.

[8] C. da Costa Pereira and A. Tettamanzi. Goal generation with relevant and trusted beliefs. In: *Proceedings of AAMAS'08*, pages 397–404. IFAAMAS, 2008.

[9] C. da Costa Pereira, A. Tettamanzi, and L. Amgoud. Goal revision for a rational agent. In: G. Brewka, S. Coradeschi, A. Perini, and P. Traverso, editors, *ECAI 2006, Proceedings of the 17th European Conference on Artificial Intelligence*, pages 747–748, Riva del Garda, Italy, August 29–September 1 2006. IOS Press.

[10] M. Dastani and L. van der Torre. Specifying the merging of desires into goals in the context of beliefs. In: *Proceedings of The First Eurasian Conference on Advances in Information and Communication Technology (EurAsia ICT 2002)*, LNCS 2510, pages 824–831. Springer, 2002.

[11] M. Dastani and L. van der Torre. Programming BOID agents: a deliberation language for conflicts between mental attitudes and plans. In: *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04)*, pages 706–713, 2004.

[12] M. Dastani, G. Governatori, A. Rotolo, and L. W. N. van der Torre. Programming cognitive agents in defeasible logic. In: G. Sutcliffe and A. Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning, 12th International Conference, LPAR 2005*, volume 3835 of *Lecture Notes in Computer Science*, pages 621–636. Springer, 2005.

[13] M. Dastani and L. W. N. van der Torre. Regulated agent-based social systems, First International Workshop, RASTA 2002, Bologna, Italy, july 16, 2002, revised selected and invited papers. In: G. Lindemann, D. Moldt, and M. Paolucci, editors, *What Is a Normative Goal?: Towards Goal-Based Normative Agent Architectures*, volume 2934 of *Lecture Notes in Computer Science*, pages 210–227. Springer, 2002.

[14] M. Dastani and L. W. N. van der Torre. Games for cognitive agents. In J. J. Alferes and J. A. Leite, editors, *Logics in Artificial Intelligence, 9th European Conference, JELIA 2004, Lisbon, Portugal, September 27-30, 2004, Proceedings*, volume 3229 of *Lecture Notes in Computer Science*, pages 5–17. Springer, 2004.

[15] F. Dignum, D. N. Kinny, and E. A. Sonenberg. From desires, obligations and norms to goals. *Cognitive Science Quarterly*, 2(3–4):407–427, 2002.

[16] J. Doyle. Rational belief revision. In: J. F. Allen, R. Fikes, and E. Sandewall, editors, *KR'91: Principles of Knowledge Representation and Reasoning*, pages 163–174, San Mateo, California, 1991. Morgan Kaufmann.

[17] J. K. Galbraith. *The Affluent Society*. Houghton Mifflin, Boston, 1958.

[18] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, 1988.

[19] S. O. Hansson. Logic of belief revision. In: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, pages URL http://plato.stanford.edu/archives/sum2006/entries/logic–belief–revision/. Published on the internet, 2006.

[20] A. Hunter. Towards higher impact argumentation. In: D. L. McGuinness and G. Ferguson, editors, *AAAI*, pages 275–280. AAAI Press / The MIT Press, 2004.

[21] R. C. Jeffrey. *The logic of decision*. University of Chicago Press, 1965.

[22] D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[23] D. Makinson and L. van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.

[24] F. Paglieri. *Belief dynamics: From formal models to cognitive architectures, and back again*. PhD thesis, University of Siena, 2006.

[25] F. Paglieri and C. Castelfranchi. Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in MAS. In: I. Rahwan, P. Moraitis, and C. Reed, editors, *Argumentation in Multi-Agent Systems, First International Workshop, ArgMAS 2004*, volume 3366 of *Lecture Notes in Computer Science*, pages 78–94. Springer, 2005.

[26] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann, 1991.

[27] S. Shapiro, Y. Lespérance, and H. J. Levesque. Goal change. In: L. Pack Kaelbling and A. Saffiotti, editors, *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 582–588, Edinburgh, Scotland, UK, July 30–August 5 2005. Professional Book Center.

[28] R. H. Thomason. Desires and defaults: A framework for planning with inferred goals. In *KR 2000, Principles of Knowledge Representation and Reasoning Proceedings of the Seventh International Conference*, pages 702–713. Morgan Kaufmann, 2000.

[29] R. H. Thomason and J. F. Horty. Nondeterministic action and dominance: Foundations for planning and qualitative decision. In: *Proceedings of the TARK'96*, pages 229–250. Morgan Kaufmann, 1996.

Received January 15, 2008