What is a Normative Goal?

Mehdi Dastani
Dept of Computer Science
Utrecht University
email: mehdi@cs.uu.nl

Leendert van der Torre
Dept of Arti£cial Intelligence
Vrije Universiteit Amsterdam
email: torre@cs.vu.nl

August 5, 2002

Abstract

In this paper we are interested in developing goal-based normative agent architectures. We ask ourselves the question what a normative goal is. To answer this question we introduce a qualitative normative decision theory based on belief (B) and obligation (O) rules. We show that every agent which makes optimal decisions – which we call a BO rational agent – acts *as if* it is maximizing its achieved normative goals. This is the basis of our design of goal-based normative agents.

1 Introduction

Simon [10] interpreted goals as utility aspiration levels, in planning goals have a notion of desirability as well as intentionality, and in the BDI approach [4, 8] goals have been identi£ed with desires. Moreover, recently several approaches to extend decision making and planning with goal generation, such as Thomason's BDP logic [11] and Broersen *et.al.*'s BOID architecture [2]. But what is this thing called goal? Although there are many uses of goals in planning and more recently in agent theory, the ontological status of goals seems to have received little attention.

In this paper we ask ourselves what a normative goal is. We draw inspiration from Savage's classical decision theory [9]. The popularity of this theory is due to the fact that Savage shows that a rational decision maker, which satis£es some innocent looking properties, acts *as if* it is maximizing its expected utility function. This is called a representation theorem. In other words, Savage does not assume that an agent has a utility function and probability distribution which the agent uses to make decisions. However, he shows that if an agent bases his decisions on preferences and some properties of these preferences, then we can assume that the agent bases his decisions on these utilities and probabilities together with the decision rule which maximizes its expected utility. The main advantage is that Savage does not have to explain what a utility function *is*, an ontological problem which had haunted decision theory for ages.

Likewise, we want to develop a qualitative normative decision theory in which a normative agent acts *as if* it is trying to maximize achieved normative goals. This is what we call a goal-based representation theorem. It implies that agents can be

formalized or veri£ed as goal based reasoners even when the agent does not reason with goals at all. In other words, goal based representations do not have to be descriptive. A consequence of this indirect de£nition of goals is that the theory tells us what a goal *is*, such that we do not have to explain its ontological status separately. Our problem is the development of such a normative decision theory.

In this paper we introduce a rule based decision theory, based on belief (B) and obligation (O) rules. Our problem thus breaks down into two subproblems:

- How to develop a normative decision theory based on belief and obligation rules?
- How to de£ne a notion of normative goals in this theory?

We call an agent which minimizes its unreached obligations a BO rational agent, and we de£ne goals as a set of formulas which can be derived by beliefs and obligations in a certain way. The distinction between the decision theory and the goal generation is the way in which the obligation rules are used. In the decision theory obligation rules are only used to evaluate the consequences of decisions, whereas they are *applied* during goal generation. Our central result thus says that *BO rational agents act as if they maximize the set of achieved goals*.

Like classical decision theory but in contrast to several proposals in the BDI approach [4, 8], the theory does not incorporate temporal reasoning and scheduling.

The layout of this paper is as follows. We £rst develops a normative logic of decision. This logic tells us what the optimal decision is, but it does not tell us how to £nd this optimal decision. We then consider the AI solution to this problem [10]: break down the decision problem into goal generation and goal based decisions.

2 A normative decision theory

The qualitative decision theory introduced in this section is based on sets of belief and obligation rules. There are several choices to be made, where our guide is to choose the simplest option available.

2.1 Logic of rules

The starting point of any theory of decision is a distinction between choices made by the decision maker and choices imposed on it by its environment. We therefore assume the two disjoint sets of propositional atoms $A = \{a, b, c, \ldots\}$ (the agent's decision variables [6] or controllable propositions [1]) and $W = \{p, q, r, \ldots\}$ (the world parameters or uncontrollable propositions). We write:

- L_A, L_W and L_{AW} for the propositional languages built up from these atoms in the usual way, and x, y, ... for any sentences of these languages.
- Cn_A, Cn_W and Cn_{AW} for the consequence sets, and $\models_A$, $\models_W$ and $\models_{AW}$ for satis£ability, in any of these propositional logics.
- $x \Rightarrow y$ for an ordered pair of propositional sentences called a rule.

- $E_R(S)$ for the R extension of S, as de£ned in De£nition 1 below.

Belief and obligation rules are interpreted as inference rules. This is formalized in the following de£nition.

De£nition 1 (Extension) *Let $R \subseteq L_{AW} \times L_{AW}$ be a set of rules and $S \subseteq L_{AW}$ a set of sentences. The consequents of the S-applicable rules are:*

$$R(S) = \{y \mid x \Rightarrow y \in R, x \in S\}$$

and the R extension of S is the set of the consequents of the iteratively S-applicable rules:

$$E_R(S) = \cap_{S \subseteq X, R(Cn_{AW}(X)) \subseteq X} X$$

The following proposition shows that $E_R(S)$ is the smallest superset of S closed under the rules R interpreted as inference rules.

Proposition 1 *Let*

- $E^0_R(S) = S$
- $E^i_R(S) = E^{i-1}_R(S) \cup R(Cn_{AW}(E^{i-1}_R(S)))$ *for* $i > 0$

We have $E_R(S) = \cup^{\infty}_0 E^i_R(S)$.

The following proposition shows that $E_R(S)$ is monotonic.

Proposition 2 *We have $R(S) \subseteq R(S \cup T)$ and $E_R(S) \subseteq E_R(S \cup T)$.*

Monotonicity is illustrated by the following example.

Example 1 *Let $R = \{\top \Rightarrow p, a \Rightarrow \neg p\}$ and $S = \{a\}$, where $\top$ stands for any tautology like $p \vee \neg p$. We have $E_R(S) = \{a, p, \neg p\}$, i.e. the R extension of S is inconsistent. We do* not *have that for example the speci£c rule overrides the more general one such that $E_R(S) = \{a, \neg p\}$.*

2.2 Decision speci£cation

A decision speci£cation given in De£nition 2 is a description of a decision problem. It contains a set of belief and obligation rules, as well as a set of facts and an initial decision (or prior intentions).

A belief rule 'the agent believes y in context x' is an ordered pair $x \Rightarrow y$ with $x \in L_{AW}$ and $y \in L_W$, and a obligation rule 'the agent is ought y in context x' is an ordered pair $x \Rightarrow y$ with $x \in L_{AW}$ and $y \in L_{AW}$. It implies that the agent's beliefs are about the world ($x \Rightarrow p$), and not about the agent's decisions. These beliefs can be about the effects of decisions made by the agent ($a \Rightarrow p$) as well as beliefs about the effects of parameters set by the world ($p \Rightarrow q$). Moreover, the agent's obligations can be about the world ($x \Rightarrow p$, obligation-to-be), but also about the agent's decisions ($x \Rightarrow a$, obligation-to-do). These obligations can be triggered by parameters set by the world ($p \Rightarrow y$) as well as by decisions made by the agent ($a \Rightarrow y$).

De£nition 2 (Decision speci£cation) *A decision speci£cation is a tuple $DS = \langle F, B, O, d_0 \rangle$ that contains a consistent set of facts $F \subseteq L_W$, £nite set of belief rules $B \subseteq L_{AW} \times L_W$, £nite set of obligation rules $O \subseteq L_{AW} \times L_{AW}$ and an initial decision $d_0 \subseteq L_A$.*

2.3 Decisions

The belief rules are used to determine the expected consequences of a decision, where a decision d is any subset of L_A that implies the initial decision d_0, and the set of expected consequences of this decision d is the belief extension of $F \cup d$. A decision does not imply a contradiction.

De£nition 3 (Decisions) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation. The set of DS decisions is*

$$\Delta = \{d \mid d_0 \subseteq d \subseteq L_A, E_B(F \cup d) \text{ is consistent } \}$$

A decision $d \in \Delta$ is a DS decision.

The following example illustrates decisions.

Example 2 *Let $A = \{a, b, c\}$, $W = \{p, q, r\}$ and $DS = \langle F, B, O, d_0 \rangle$ with $F = \{p \rightarrow r\}$, $B = \{p \Rightarrow q, b \Rightarrow \neg q, c \Rightarrow p\}$, $O = \{\top \Rightarrow r, \top \Rightarrow a, a \Rightarrow b\}$ and $d_0 = \{a\}$. The initial decision d_0 re¤ects that the agent has already decided in an earlier stage to reach the obligation $\top \Rightarrow a$. Note that the consequents of all B rules are sentences of L_W, whereas the antecedents of the B rules as well as the antecedents and consequents of the O rules are sentences of L_{AW}. We have due to the de£nition of $E_R(S)$:*
$E_B(F \cup \{a\}) = \{p \rightarrow r, a\}$
$E_B(F \cup \{a, b\}) = \{p \rightarrow r, a, b, \neg q\}$
$E_B(F \cup \{a, c\}) = \{p \rightarrow r, a, c, p, q\}$
$E_B(F \cup \{a, b, c\}) = \{p \rightarrow r, a, b, c, p, q, \neg q\}$
Therefore $\{a, b, c\}$ is not a DS decision, because its extension is inconsistent.

There are two ways in which we continue. In the following we introduce a normative theory, which determines the interpretation of the elements of the decision speci£cation. Thereafter we introduce a new element, called goals, in the decision theory. The distinction between the normative decision theory and the goal-based decision theory is how the obligation rules are used. In the normative theory obligation rules are only used to evaluate the consequences of decisions, they are never *applied*. In the goal based decision theory the obligation rules are applied during the generation of goals.

2.4 Optimal decisions

The obligation rules are used to compare the decisions. The comparison is based on the set of unreached obligations and not on the set of violated or reached obligations, where a obligation $x \Rightarrow y$ is unreached by a decision if the expected consequences of this decision imply x but not y, and it is violated or reached if these consequences imply respectively $x \wedge \neg y$ or $x \wedge y$. Note that the set of unreached desires is a superset of the set of violated desires.[1]

[1] In earlier work such as [12] we used the set of violated and reached obligations to order states, in the sense that we minimized violations and maximized reached obligations. The present de£nition has the advantage that it is simpler because it is based on a single minimization process only. Note that in the present circumstances we cannot minimize violations only, because it would lead to the counterintuitive situation that the minimal decision $d = d_0$ is always optimal.

De£nition 4 (Comparing decisions) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation and d be a DS decision. The unreached obligations of decision d are:*

$$U(d) = \{x \Rightarrow y \in O \mid E_B(F \cup d) \models x \text{ and } E_B(F \cup d) \not\models y\}$$

Decision d_1 is at least as good as decision d_2, written as $d_1 \geq_U d_2$, iff

$$U(d_1) \subseteq U(d_2)$$

Decision d_1 dominates decision d_2, written as $d_1 >_U d_2$, iff

$$d_1 \geq_U d_2 \text{ and } d_2 \not\geq_U d_1$$

Decision d_1 is as good as decision d_2, written as $d_1 \sim_U d_2$, iff

$$d_1 \geq_U d_2 \text{ and } d_2 \geq_U d_1$$

The following continuation of Example 2 illustrates the comparison of decisions.

Example 3 (Continued) *We have:*
$U(\{a\}) = \{\top \Rightarrow r, a \Rightarrow b\}$,
$U(\{a, b\}) = \{\top \Rightarrow r\}$,
$U(\{a, c\}) = \{a \Rightarrow b\}$.
We thus have that the decisions $\{a, b\}$ and $\{a, c\}$ both dominate the initial decision $\{a\}$, i.e. $\{a, b\} >_U \{a\}$ and $\{a, c\} >_U \{a\}$, but the decisions $\{a, b\}$ and $\{a, c\}$ do not dominate each other nor are they as good as each other, i.e. $\{a, b\} \not\geq_U \{a, c\}$ and $\{a, c\} \not\geq_U \{a, b\}$.

The following proposition shows that the binary relation on decisions is transitive and we can thus interpret it as a preference relation.

Proposition 3 *The binary relation $\geq_U$ is transitive.*

The following proposition shows that obligations only matter as long as they are different from beliefs.

Proposition 4 *The decision set of decision speci£cation $DS = \langle F, B, O, d_0 \rangle$ is exactly the decision set of $DS' = \langle F, B, O \setminus B, d_0 \rangle$. Moreover, we have $d_1 \geq_U d_2$ in DS if and only if $d_1 \geq_U d_2$ in DS'.*

The decision theory prescribes a decision maker to select the optimal or best decision, which is de£ned as a decision that is not dominated.

De£nition 5 (Optimal decision) *Let DS be a decision speci£cation. A DS decision d is U-optimal iff there is no DS decision d' that dominates it, i.e. $d' >_U d$.*

The following example illustrates that the minimal decision d_0 is not necessarily an optimal decision.

Example 4 *Let $A = \{a\}$, $W = \emptyset$ and $DS = \langle \emptyset, \emptyset, \{\top \Rightarrow a\}, \emptyset \rangle$. We have $U(\emptyset) = \{\top \Rightarrow a\}$ and $U(\{a\}) = \emptyset$. Hence, doing a is better than doing nothing.*

The following example illustrates optimal decisions.

Example 5 *Let $A = \{a, b\}$, $W = \emptyset$ and $DS = \langle \emptyset, \emptyset, \{a \Rightarrow b\}, \emptyset \rangle$. We have:*
$U(d) = \{a \Rightarrow b\}$ if $d \models_{AW} a$ and $d \not\models_{AW} b$, $U(d) = \emptyset$ otherwise
The U-optimal decisions are the decisions d that either do not imply a or that imply $a \wedge b$.

The following proposition shows that for each decision speci£cation, there is at least one optimal decision. This is important, because agents have to act in some way.

Proposition 5 *Let DS be a decision speci£cation. There is at least one U-optimal DS decision.*

Proof. Since the facts F are consistent, there exists at least one DS decision. Since the set of desire rules is £nite there do not exist in£nite ascending chains in $\geq_U$, and thus there is an optimal D decision.

An alternative to our notion of optimality is to introduce a notion of minimality in the de£nition of optimal decisions. The following De£nition 6 introduces a distinction between smaller and larger decisions. A smaller decision implies that the agent commits itself to less choices. A minimal optimal decision is an optimal decision such that there is no smaller optimal decision.

De£nition 6 (Minimal optimal decision) *A decision d is a minimal U-optimal DS decision iff it is an U-optimal DS decision and there is no DS decision d' such that $d \models d'$ and $d' \not\models d$.*

The following example illustrates the distinction between optimal and minimal optimal decisions.

Example 6 *Let $DS = \langle F, B, O, d_0 \rangle$ with $F = \emptyset$, $B = \{a \Rightarrow x, b \Rightarrow x\}$, $O = \{\top \Rightarrow x\}$, $d_0 = \emptyset$. Optimal decisions are $\{a\}$, $\{b\}$ and $\{a, b\}$, of which the former two are minimal.*

The following proposition illustrates in what sense a decision theory based on optimal decisions and one based on minimal optimal decisions are equivalent.

Proposition 6 *For every U-optimal DS decision d, there is a minimal U-optimal DS decision d' such that $d \sim_U d'$.*

We £nally de£ne what a BO rational agent is.

De£nition 7 *A BO rational agent is an agent that, confronted with a decision speci£cation DS, selects an U-optimal DS decision. A BO parsimoneous agent is a BO rational agent that selects a minimal U-optimal DS decision.*

Thus far, we have not considered the notion of goals. This concept is introduced in the following section.

3 Goal-based decision theory

In this section we show that every rational agent, in the sense of De£nition 7, can be understood as a goal planning agent [7].

3.1 Goal-based optimal decisions

Goal-based decisions in De£nition 8 combine decisions in De£nition 3 and the notion of goal, which is a set of propositional sentences. Note that a goal set can contain decision variables (which we call to-do goals) as well as parameters (which we call to-be goals).

De£nition 8 (Goal-based decision) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation and the goal set $G \subseteq L_{AW}$ a set of sentences. A decision d is a G decision if $E_B(F \cup d) \models_{AW} G$.*

What is the goal set of an optimal decision? One way to start is to consider all derivable goals from an initial decision and a *maximal* set of obligations.

De£nition 9 (Derivable goal set) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation. A set of formulas $G \subseteq L_{AW}$ is a derivable goal set of DS if*

$$G = E_{B \cup O'}(F \cup d_0) \setminus Cn_{AW}(E_B(F \cup d_0))$$

where $O' \subseteq O$ is a maximal *(with respect to set inclusion) set such that*

1. *$E_{B \cup O'}(F \cup d_0)$ is consistent and*
2. *there is a DS decision d that is a G decision.*

However, the following proposition shows that for some derivable goal set G, not all G decisions are optimal.

Proposition 7 *For a derivable goal set G of DS, a G decision does not have to be an U-optimal decision.*

Proof. Consider the decision speci£cation $DS = \langle \emptyset, \{a \Rightarrow p\}, \{\top \Rightarrow p, a \Rightarrow b\}, \emptyset \rangle$. The set $G = \{p\}$ is the only derivable goal set (based on $O' = \{\top \Rightarrow p, a \Rightarrow b\}$). The DS decisions $d_1 = \{a\}$ and $d_2 = \{a, b\}$ are both G decisions, but only d_2 is an U-optimal decision.

We therefore de£ne goals with respect to an optimal decision.

De£nition 10 (Achievable goal set) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation. A set of formulas $G \subseteq L_{AW}$ is an achievable goal set of DS if there is an U-optimal DS decision d such that*

$$G = \{x \wedge y \mid x \Rightarrow y \in O', E_{B \cup O'}(F \cup d) \models_{AW} x \wedge y\}$$

where

$$O' = \{x \Rightarrow y \in O \mid E_B(F \cup d) \not\models_{AW} x \text{ or } E_B(F \cup d) \models_{AW} x \wedge y\}$$

Proposition 8 *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation and let a set of formulas $G \subseteq L_{AW}$ be an achievable goal set of DS. There exists an U-optimal DS decision d such that*

$$G = \{x \wedge y \mid x \Rightarrow y \in O, E_B(F \cup d) \models_{AW} x \wedge y\}$$

Proof. Follows directly from $E_{B \cup O'}(F \cup d) = E_B(F \cup d)$, where O' is as de£ned in De£nition 10.

The following proposition shows that we can de£ne one half of the representation theorem for achievable goal sets.

Proposition 9 *For an U-optimal decision d of DS there is an achievable goal set G of DS such that d is a G decision.*

Proof. Follows directly from $E_{B \cup O'}(F \cup d) = E_B(F \cup d)$.

However, the following proposition shows that the other half of the representation theorem still fails.

Proposition 10 *For an achievable goal set G of DS, a G decision does not have to be an U-optimal decision.*

Proof. Consider the decision speci£cation $DS = \langle \{\neg q\}, \{a \Rightarrow p, b \Rightarrow p\}, \{\top \Rightarrow p, b \Rightarrow q\}, \emptyset \rangle$. The set $G = \{p\}$ is the only achievable goal set (based on $O' = \{\top \Rightarrow p, b \Rightarrow q\}$). The DS decisions $d_1 = \{a\}$ and $d_2 = \{b\}$ are both (minimal) G decisions, but only d_1 is an optimal decision.

The counterexample in Proposition 10 also shows that we cannot prove the second half of the representation theorem, because we only consider positive goals (states the agent wants to reach) and not negative goals (states the agents wants to evade). The theory is extended with positive and negative goals in the following subsection.

3.2 Positive and negative goals

In this section we show that the representation theorem works both ways if we add negative goals, which are de£ned in the following de£nition as states the agent has to avoid. They function as constraints on the search process of goal-based decisions.

De£nition 11 (Goal-based decision) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation, and the so-called positive goal set G^+ and negative goal set G^- subsets of L_{AW}. A decision d is a $\langle G^+, G^- \rangle$ decision if $E_B(F \cup d) \models_{AW} G^+$ and for each $g \in G^-$ we have $E_B(F \cup d) \not\models_{AW} g$.*

The de£nition of achievable goal set is extended with negative goals.

De£nition 12 (Positive and negative achievable goal set) *Let $DS = \langle F, B, O, d_0 \rangle$ be a decision speci£cation. The two sets of formulas $G^+, G^- \subseteq L_{AW}$ are respectively a positive and negative achievable goal sets of DS if there is an optimal DS decision d such that*

$$G^+ = \{x \wedge y \mid x \Rightarrow y \in O', E_B(F \cup d) \models_{AW} x \wedge y\}$$

$$G^- = \{x \mid x \Rightarrow y \in O', E_{B \cup O'}(F \cup d) \not\models_{AW} x\}$$

where

$$O' = \{x \Rightarrow y \in O \mid E_B(F \cup d) \not\models_{AW} x \text{ or } E_B(F \cup d) \models_{AW} x \wedge y\}$$

The following example illustrates the distinction between optimal decisions and minimal optimal ones.

Example 7 *Let $A = \{a, b\}$, $W = \emptyset$ and $DS = \langle \emptyset, \emptyset, \{a \Rightarrow b\}, \emptyset \rangle$. The optimal decision is $\emptyset$ or $\{a, b\}$, and the related goal sets are $\langle G^+, G^- \rangle = \langle \emptyset, \{a\} \rangle$ and $\langle G^+, G^- \rangle = \langle \{a \wedge b\}, \emptyset \rangle$. The only minimal optimal decision is the former.*

The following example illustrates a con¤ict.

Example 8 *Let $W = \{p\}$, $A = \{a\}$, $DS = \langle F, B, O, d_0 \rangle$ with $F = \emptyset$, $B = \emptyset$, $O = \{\top \Rightarrow a \wedge p, \top \Rightarrow \neg a\}$, $d_0 = \emptyset$. We have optimal decision $\{\neg a\}$ with goal set $\langle G^+, G^- \rangle = \langle \{\neg a\}, \emptyset \rangle$. The decision $\{a\}$ does not derive goal set $\langle G^+, G^- \rangle = \langle \{a \wedge p\}, \emptyset \rangle$. One of the possible choices is $\{a\}$, which is however sub-optimal since we cannot guarantee that the £rst obligation is reached.*

The £rst part of the representation theorem is analogous to Proposition 9.

Proposition 11 *For an U-optimal decision d of DS there is an achievable goal set $\langle G^+, G^- \rangle$ of DS such that d is a $\langle G^+, G^- \rangle$ decision.*

Proof. See Proposition 9.

Proposition 12 *For an achievable goal set $\langle G^+, G^- \rangle$ of DS, a $\langle G^+, G^- \rangle$ decision is an U-optimal decision.*

Proof. $\langle G^+, G^- \rangle$ is achievable and thus there there is an U-optimal DS decision such that $E_B(F \cup d) \models_{AW} G^+$ and for all $g \in G^-$ we have $E_B(F \cup d) \not\models_{AW} g$. Let d be any decision d such that $E_B(F \cup d) \models_{AW} G^+$ and for all $g \in G^-$ we have $E_B(F \cup d) \not\models_{AW} g$. Suppose d is not U-optimal. This means that there exists a d' such that $d' >_U d$, i.e. such that there exists an obligation $x \Rightarrow y \in O$ with $E_B(F \cup d) \models_{AW} x$, $E_B(F \cup d) \not\models_{AW} y$ and either:

- *$E_B(F \cup d') \not\models_{AW} x \wedge y$;*
- *$E_B(F \cup d') \models_{AW} y$;*

However, the £rst option is not possible due to the negative goals and the second option is not possible due to the positive goals. Contradiction, so d has to be U-optimal.

The representation theorem is a combination of Proposition 11 and 12.

Theorem 1 *A decision d is an U-optimal decision if and only if there is an achievable goal set $\langle G^+, G^- \rangle$ of DS such that d is a $\langle G^+, G^- \rangle$ decision.*

4 Agent speci£cation and design

In this section we discuss how the proposed qualitative normative decision and goal theory can be used to guide the design and speci£cation of rational BO agents in a compositional way. The general idea of compositional design and speci£cation is to build agents using components. They may be either primitive components or composed of other components, such that the speci£cation of agents can be broken down into the speci£cation of components and their relations. Here we give some preliminary ideas and explain how the proposed qualitative normative decision and goal theory supports a speci£c compositional design for rational BO agent.

The qualitative decision theory, as proposed in section 2, speci£es the decision making of an agent in terms of its observations and its mental attitudes such as beliefs and obligations. The speci£ed agent can therefore be considered as consisting of components that represent agent's beliefs and obligations and a reasoning component that generates agent's decisions based on its observations and mental attitudes. The abstract design of such a BO agent is illustrated in Figure 1. For this design of BO agents, notions such as optimal decisions and minimal optimal decisions can be used to specify the reasoning component and thus the decision making mechanism of the agent.

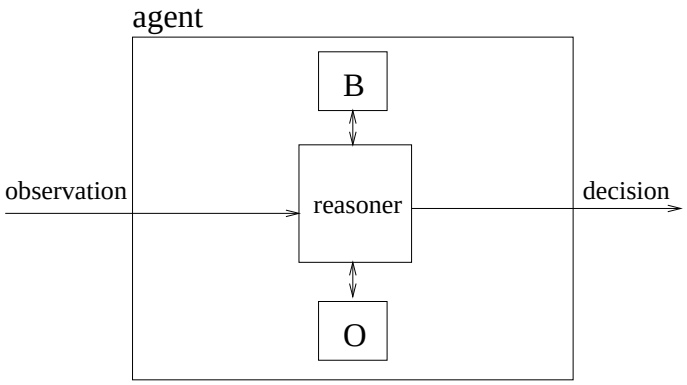

Figure 1: Agent

The following example illustrates an agent with certain beliefs and obligations, the possible decisions that the agent can make, and how the notions from qualitative normative decision theory can be used to specify the subset of decisions that the agent can make.

Example 9 *Consider an agent who believes that he works and that if he sets an alarm clock he can wake up early to arrive in time at his work, i.e.*

$$B = \{\top \Rightarrow Work, SetAlarm \Rightarrow InTime\}$$

The agent has also the obligation to arrive early at his work and he has to inform his boss when he does not work, i.e.

$$O = \{Work \Rightarrow InTime, \neg Work \Rightarrow InformBoss\}$$

In this example, the propositions $SetAlarm$ and $InformBoss$ are assumed to be decision variables (the agent has control on setting the alarm clock and informing his boss), while $Work$ and $Intime$ are assumed to be world parameters (the agent has no control on its working status and the starting time). Moreover, we assume that the agent has no observation and no intentions. One can specify the agent as a rational BO agent in the sense that it makes optimal decisions. Being speci£ed as a rational BO agent, he will decide to use the alarm clock though he has in principle many possible decisions including $\emptyset, \{SetAlarm\}, \{InformBoss\}$, and $\{SetAlarm, InformBoss\}$.

The goal based decision theory, as proposed in section 3, explains the decision making of a rational BO agent as if it aims at maximizing achieved normative goals. In particular, the goal based decision theory explains how normative goals of an agent can be speci£ed based on its decision speci£cation. The speci£ed reasoning component of the rational BO agent can therefore be decomposed and designed as consisting of two reasoning components: one which generates normative goals and one which generate decisions to achieve those goals. This decomposition suggests an agent design as illustrated in Figure 2. According to this agent design, a BO agent generates £rst its normative goals based on its observation, its beliefs, obligations and its intentions. The generated goals are subsequently the input of the decision generation component.

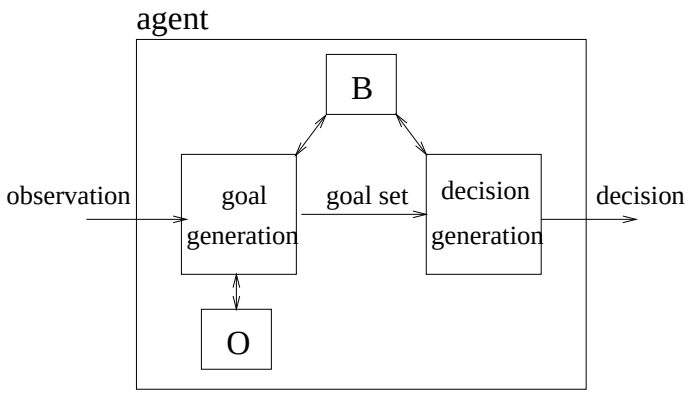

Figure 2: Goal-based agent

Following the design decomposition, the speci£cation of a BO agent can now also be decomposed and de£ned in terms of the speci£cation of its goal and decision generation mechanisms. In particular, the goal generation mechanism can be speci£ed in terms of agent's observations and its mental state on the one hand and its goals on the other hand. The decision generation component can then be speci£ed in terms of agent's goals and mental state on the one hand and its decisions on the other hand.

For example, consider again the working agent that may have in principle many goal sets consisting of $\emptyset, Work, Intime, SetAlarm,$ and $InformBoss$. This implies that the goal generation component may generate one of these possible goal sets. Using the notions from goal based decision theory one may specify the goal generation mechanism in order to generates achievable goal sets which when planned by the decision

generation component will result optimal decisions.

In summary, we believe that the qualitative normative decision theory and goal based decision theory can be used to provide compositional speci£cation and design of rational BO agents. This leads to a transparent agent speci£cation and design structure. Moreover, it leads to support for reuse and maintainability of components and generic models. The compositional speci£cation and design of agents enable us to specify and design agents at various levels of abstraction leaving out many details such as representation issues and reasoning schemes. For our rational BO agents we did not to explain how decisions are generated; we only speci£ed what decisions should be generated. At one lower level we decomposed the reasoning mechanism and speci£ed goal and decision generation mechanisms. We also did not discuss the representation of individual components such as the belief or the obligation components. The conditional rules in these components specify the input/output relation.

5 Related research

The theories in Thomason's BDP [11] and Broersen et al.'s BOID [2] are different, because they allow multiple belief sets. This introduces the new problem of blocking wishful thinking discussed extensively in [3].

6 Concluding remarks

In this paper we have given an interpretation for goals in a qualitative decision theory based on beliefs and obligation rules, and we have shown that any agent which makes optimal decisions acts as if it is maximizing its achieved goals.

Our motivation comes from the analysis of goal-based architectures, which have recently been introduced. However, the results of this paper may be relevant for a much wider audience. For example, Dennett argues that automated systems can be analyzed using concepts from folk psychology like beliefs, obligations, and goals. Our work may be used in the formal foundations of this 'intentional stance' [5].

There are several topics for further research. The most interesting question is whether belief and obligation rules are fundamental, or whether they in turn can be represented by some other construct. Other topics for further research are a generalization of our representation theorem to other choices in our theory, the development of an incremental approach to goals, and the development of computationally attractive fragments of the logic, and heuristics of the optimization problem.

References

[1] C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the KR'94*, pages 75–86, 1994.

[2] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, to appear.

[3] J. Broersen, M. Dastani, and L. van der Torre. Realistic desires. *Journal of Applied Non-Classical Logics*, to appear.

[4] P.R. Cohen and H.J. Levesque. Intention is choice with commitment. *Arti£cial Intelligence*, 42:213–261, 1990.

[5] D. Dennett. *The intentional stance*. MIT Press, Cambridge, MA, 1987.

[6] J. Lang. Conditional desires and utilities - an alternative approach to qualitative decision theory. In *In Proceedings of the European Conference on Arti£cial Intelligence (ECAI'96)*, pages 318–322, 1996.

[7] A. Newell. The knowledge level. *Arti£cial Intelligence*, 1982.

[8] A. Rao and M. Georgeff. Modeling rational agents within a BDI architecture. In *Proceedings of the KR91*, 1991.

[9] L. Savage. *The foundations of statistics*. 1954.

[10] H. A. Simon. *The Sciences of the Arti£cial*. MIT Press, Cambridge, MA, second edition, 1981.

[11] R. Thomason. Desires and defaults: A framework for planning with inferred goals. In *Proceedings KR 2000*, pages 702–713, 2000.

[12] L. van der Torre and Y. Tan. Diagnosis and decision making in normative reasoning. *Arti£cial Intelligence and Law*, 7(1):51–67, 1999.