Semantics of Declarative Goals in Agent Programming

M. Birna van Riemsdijk
Utrecht University
The Netherlands
birna@cs.uu.nl

Mehdi Dastani
Utrecht University
The Netherlands
mehdi@cs.uu.nl

John-Jules Ch. Meyer
Utrecht University
The Netherlands
jj@cs.uu.nl

ABSTRACT

This paper addresses the notion of declarative goals as used in agent programming. Declarative goals describe desirable states, and semantics of these goals in an agent programming context can be defined in various ways. This paper defines two semantics for goals, with one based on default logic. The semantics are partly motivated by an analysis of other proposals that have been done in the literature. Further, we establish relations between and properties of these semantics.

Categories and Subject Descriptors

F.3.2 [**Logics and Meanings of Programs**]: Semantics of Programming Languages; D.3.1 [**Programming Languages**]: Formal Definitions and Theory; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, languages and structures*; I.2.5 [**Artificial Intelligence**]: Programming Languages and Software; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Representations (procedural and rule-based)*

General Terms

Theory, Languages

Keywords

Agent programming languages, semantics, default logic, declarative goals

1. INTRODUCTION

An important concept in agent theory is the concept of a *goal*. Goals are introduced to explain and specify an agent's (proactive) behavior. Various logics have been introduced to formalize the concept of goals and reasoning about goals [12, 17]. In these logics, a goal is formalized as a set of states and thus has a *declarative* interpretation.

Declarative goals are important not only in agent logics, but also in *agent programming*. They for example provide for the possibility to decouple plan execution and goal achievement [20]. If a plan fails, the goal that was to be achieved by the plan remains in the goal base of the agent. The agent can then for example select a different plan or wait for the circumstances to change for the better. Furthermore, agents can be implemented such that they can communicate about their goals [10]. Also, a representation of goals in agents enables reasoning about goal interaction [14].

This paper is set in a cognitive agent programming context and is thus based on the idea that an agent consists of data structures representing the agent's mental attitudes such as declarative goals and beliefs. To be more specific, an agent in our framework consists of a belief base, a goal base, an intention base and a rule base. Rules are used to represent or manipulate an agent's mental attitudes. Given these data structures, the specification of what it means that the agent has a certain goal, i.e., the specification of the semantics of goals, can be defined in various ways.

The contribution of this paper is twofold. Firstly, we define two semantics for goals, partly motivated by an analysis of other proposals that have been done in the literature. Secondly, we establish relations between and properties of these semantics. Our aim is to provide a more principled and relatively systematic analysis of the kinds of semantics one could consider and of their properties. It is important to have a better understanding of the different possible semantics and their characteristics, because this will help to identify which semantics have the more desirable features, in general or for certain kinds of applications.

The paper is organized as follows. First, we present some preliminaries concerning cognitive agent programming and default logic (section 2). Default logic is used in section 4 to define semantics of goals, based on the goal base and rule base of the agent. Section 3 discusses semantics of goals, based only on the goal base of the agent. Concluding, we remark that although consideration of issues of computational complexity is important, this is not addressed in this paper and remains for future research.

2. PRELIMINARIES

2.1 Cognitive Agent Programming

The notion of an agent configuration as defined below is used to formally define an agent's state. Throughout this paper, we assume a language of propositional logic $\mathcal{L}$ with negation and conjunction, with typical element ϕ. $\top \in \mathcal{L}$ will be used to denote a tautology, $\bot \in \mathcal{L}$ to denote falsum and $\models$ will be used to denote the standard satisfiability relation for $\mathcal{L}$. Further, we assume a language of plans Plan with typical element π. This could for example be a language of sequentially composed actions. An exact specification is not needed for the purpose of this paper and therefore we will not provide one.

DEFINITION 1. *(agent configuration) An agent configuration, typically denoted by c, is a tuple* $\langle \sigma, \gamma, \iota, \mathcal{R} \rangle$ *where* $\sigma \subseteq \mathcal{L}$ *is the belief base,* $\gamma \subseteq \mathcal{L}$ *is the goal base,* $\iota \subseteq (\mathsf{Plan} \times \mathcal{L})$ *is the intention base and* $\mathcal{R}$ *is a tuple of sets of rules. All sets* σ, γ, ι *and sets in* $\mathcal{R}$ *are finite.*

Note that in this definition, we are not specific about which types of rules constitute the rule base. We will gradually define this component in the sequel. The intention base is a set of pairs from $\mathsf{Plan} \times \mathcal{L}$. The idea is, that a pair $\langle \pi, \phi \rangle \in \iota$ represents a selected plan with an associated goal that the plan is to achieve (see also [6]). Beliefs describe the state of the world the agent is in, goals describe the state the agent wants to reach and plans are the means to achieve these goals.

Belief and goal formulas as defined below are used to refer to the beliefs and goals of an agent.

DEFINITION 2. *(belief and goal formulas) The belief formulas* $\mathcal{L}_{\mathsf{B}}$ *with typical element* β *and the goal formulas* $\mathcal{L}_{\mathsf{G}}$ *with typical element* κ *are defined as follows.*

- $\top \in \mathcal{L}_{\mathsf{B}}$ *and* $\top \in \mathcal{L}_{\mathsf{G}}$.
- *If* $\phi \in \mathcal{L}$*, then* $\mathbf{B}\phi \in \mathcal{L}_{\mathsf{B}}$ *and* $\mathbf{G}\phi \in \mathcal{L}_{\mathsf{G}}$.
- *If* $\beta, \beta' \in \mathcal{L}_{\mathsf{B}}$ *and* $\kappa, \kappa' \in \mathcal{L}_{\mathsf{G}}$*, then* $\neg\beta, \beta \wedge \beta' \in \mathcal{L}_{\mathsf{B}}$ *and* $\neg\kappa, \kappa \wedge \kappa' \in \mathcal{L}_{\mathsf{G}}$.

Note that the $\mathbf{B}$ and $\mathbf{G}$ operators cannot be nested, i.e., formulas of the form $\mathbf{BG}\phi$ or $\mathbf{BB}\phi$ are not part of the language. The semantics of belief formulas is defined based on an agent configuration as follows.

DEFINITION 3. *(semantics of belief formulas) Let* $\phi \in \mathcal{L}$ *and let* $\langle \sigma, \gamma, \iota, \mathcal{R} \rangle$ *be an agent configuration. Let* $\beta \in \mathcal{L}_{\mathsf{B}}$*. The semantics* $\models_{\mathcal{L}_{\mathsf{B}}}$ *of belief formulas is then as defined below.*

$$\begin{array}{lll}
\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \top & & \\
\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \mathbf{B}\phi & \Leftrightarrow & \sigma \models \phi \\
\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \neg\beta & \Leftrightarrow & \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \not\models_{\mathcal{L}_{\mathsf{B}}} \beta \\
\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \beta_1 \wedge \beta_2 & \Leftrightarrow & \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \beta_1 \text{ and} \\
 & & \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_{\mathcal{L}_{\mathsf{B}}} \beta_2
\end{array}$$

Investigating ways to define the semantics of goal formulas is the main research objective of this paper and these analyses will be carried out in sections 3 and 4.

Although the focus of this paper is on the semantics of goals and we do not strive to provide a complete agent programming framework, we do present a way to generate intentions on the basis of certain beliefs and goals. In an agent programming setting, it is common to introduce rules to generate plans or intentions (see for example [19, 6]). We will explain why we introduce intention generation rules after defining their semantics.

DEFINITION 4. *(intention generation rule) The set of intention generation rules* $\mathcal{R}_{\mathsf{IG}}$ *is defined as follows:*

$$\{\beta, \kappa \Rightarrow_{\mathbf{I}} \langle \pi, \phi \rangle \mid \beta \in \mathcal{L}_{\mathsf{B}}, \kappa \in \mathcal{L}_{\mathsf{G}}, \pi \in \mathsf{Plan}, \phi \in \mathcal{L}\}.$$

An intention generation rule $\beta, \kappa \Rightarrow_{\mathbf{I}} \langle \pi, \phi \rangle$ can be applied in a configuration, if the belief and goal conditions β and κ of the rule hold in that configuration. If this rule is applied, the intention $\langle \pi, \phi \rangle$ consisting of a plan π and a goal ϕ that is to be achieved by the plan, is added to the existing set of intentions. These existing intentions should however provide a so-called "screen of admissibility". The idea, as put forward by the philosopher Bratman [2], is that newly adopted intentions should not conflict with already existing ones. This can be incorporated into our framework by requiring that the goal of the intention that is to be adopted is consistent with goals of already existing intentions.

These ideas are formalized below in the definition of the semantics of application of an intention generation rule. The semantics is defined by giving a so-called transition rule [11]. A transition rule is used to derive transitions and a transition is a transformation of one agent configuration into another, which corresponds to a single computation step. We use the notation $c \to c'$ to denote a transition from agent configuration c to c'.

DEFINITION 5. *(semantics of intention generation) Let* $\mathsf{IG} \subseteq \mathcal{R}_{\mathsf{IG}}$ *be a finite set of intention generation rules, let* $\mathcal{R} = \langle \mathsf{IG} \rangle$ *and let* $\beta, \kappa \Rightarrow_{\mathbf{I}} \langle \pi, \phi \rangle \in \mathsf{IG}$ *be an intention generation rule. Further, let* $c = \langle \sigma, \gamma, \iota, \mathcal{R} \rangle$ *be an agent configuration and let* $\models_{\mathcal{L}_{\mathsf{G}}}$ *be a satisfiability relation for goal formulas. The semantics of applying this rule is then as follows, where* $\iota' = \iota \cup \{\langle \pi, \phi \rangle\}$ *and* $\delta = \{\phi \mid \langle \pi, \phi \rangle \in \iota\}$.

$$\frac{c \models_{\mathcal{L}_{\mathsf{B}}} \beta \qquad c \models_{\mathcal{L}_{\mathsf{G}}} \kappa \qquad \{\phi\} \cup \delta \not\models \bot}{\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \to \langle \sigma, \gamma, \iota', \mathcal{R} \rangle}$$

Note that an intention generation rule of the form $\beta, \kappa \Rightarrow_{\mathbf{I}} \langle \pi, \phi \rangle$ for which ϕ is inconsistent, i.e., $\phi \models \bot$, will never be applicable.

We introduce intention generation rules in this paper for two reasons. First, intention generation rules will be used to partly justify the semantics for goal formulas that are proposed in this paper. The applicability of an intention generation rule *depends* on the truth of the goal formula in its antecedent and different choices in defining the semantics for goal formulas will thus influence the applicability of the rule. Second, we will argue in section 3.1 that the goal base of an agent does not have to be consistent, partly because agents in our framework have an intention base that *is* consistent. If the initial intention base of the agent is consistent, our semantics of intention generation rules will maintain this property throughout the execution of the agent.

Concluding, we make the following remark. The check of whether the new intention is not conflicting with existing intentions, is implemented by checking logical consistency of the goal of the new intention, with the goals of existing ones. In general, one could consider other conditions for a new intention to be compatible with existing ones, such as conditions on resources like energy or money that are used by the plans (see for example [15] for a more elaborate treatment of this topic) or conditions on subgoals that should be reached by the plans. Investigating more elaborate definitions of intention compatibility is however not within the scope of this paper and the simple definition as given above will suffice for our purposes.

2.2 Default Logic

In section 4, we will use default logic [13] to define the semantics of declarative goals. In this section, we briefly sketch the ideas of default logic. For more elaborate treatments of this topic, the reader can for example consult [1, 3]. Default logic is generally based on predicate logic, but for this paper it suffices to consider propositional default logic.

Default logic distinguishes facts, representing certain but incomplete information about the world, and default rules or defaults, representing rules of thumb, by means of which conclusions can be drawn that are plausible, but not necessarily true. This means that some conclusions may have to be revised when more information becomes available. Given the propositional language $\mathcal{L}$ which we introduced in section 2.1, a *default rule* has the form $\phi : \psi_1, \ldots, \psi_n/\chi$, where $\phi, \psi_1, \ldots, \psi_n, , \chi \in \mathcal{L}$ and $n > 0$. The intuitive reading of a default rule of this form is the following: if ϕ is provable and for all $1 \leq i \leq n$, $\neg\psi_i$ is not provable, i.e., if it is consistent to assume ψ_i, then derive χ. The formula ϕ is called the prerequisite and the formulas $\psi_1, \ldots, \psi_n$ are called the justifications of the default rule.

A *default theory* [3] is a pair $\langle W, D\rangle$, where $W \subseteq \mathcal{L}$ is the set of facts and D is a set of default rules. The semantics of a default theory $\langle W, D\rangle$ can be defined through so-called *extensions* of the theory. If $E \subseteq \mathcal{L}$ is a set of propositional formulas, then a sequence of sets of formulas $E_0, E_1, \ldots$ is defined as follows, where $\models$ is the standard satisfiability relation for $\mathcal{L}$ and $Th(E_i)$ is the closure under classical logical consequence of E_i.

$$\begin{array}{lll} E_0 & = & W \\ E_{i+1} & = & Th(E_i) \cup \{\chi \mid \phi : \psi_1, \ldots, \psi_n/\chi \in D, \\ & & E_i \models \phi, E \not\models \neg\psi_i\} \end{array}$$

A set $E \subseteq \mathcal{L}$ is then an extension of $\langle W, D\rangle$ iff $E = \bigcup_{i=0}^{\infty} E_i$.

It is important to note that extensions are always *consistent* sets[1] that are *closed* under the application of default rules. A rule $\phi : \psi_1, \ldots, \psi_n/\chi$ is *applicable* to an extension E iff $E \models \phi$ and $E \not\models \neg\psi_i$ for $1 \leq i \leq n$. An extension E of a default theory $\langle W, D\rangle$ is closed under the application of default rules, iff it holds that for all rules $\phi : \psi_1, \ldots, \psi_n/\chi \in D$, that if the rule is applicable to E, then $E \models \chi$.

Example 1. *Let $W = \{a\}$, let $d_1 = a : \neg b/d$ and $d_2 = \top : c/b$ and let $D = \{d_1, d_2\}$. The default theory $\langle W, D\rangle$ then has one extension: $\{a, b\}$. This extension can be generated by applying d_2 to W. The set $\{a, d, b\}$, which might seem to be possible to generate by applying d_1 and then d_2, is not an extension: b is derivable from this set, whereas b should not be derivable because the default rule d_1 with justification $\neg b$ was applied. The set $\{a, d\}$ is neither an extension, because it is not closed under the application of defaults. The rule d_2 is applicable, although application will yield a set that is not an extension.* △

In the so-called credulous semantics for default logic a formula ϕ is said to follow from a default theory iff ϕ is in *one* of the extensions of this theory. The sceptical semantics defines that ϕ follows from a default theory iff ϕ is in *all* of the extensions of this theory.

[1]That is, if W is consistent.

3. SEMANTICS WITHOUT GOAL ADOPTION RULES

In this section, we present a number of semantics of goal formulas, based on the goal base of an agent configuration. In section 4, we will consider a semantics based on the goal base and the rule base, containing a set of so-called goal adoption rules.

3.1 Semantics

A central issue with respect to these semantics, is the issue of *consistency* of goals. In agent theories as well as in agent programming frameworks, goals are often assumed or required to be consistent. The rationale is, that an agent should not simultaneously pursue situations that are mutually logically impossible.

In our framework, this can be modelled by requiring that the goal base of an agent configuration is consistent. Given this requirement, the semantics of goal formulas can be defined in a simple way as follows, where $\mathbf{G}\phi$ is true in a configuration iff ϕ follows from the goal base in this configuration (see also [18]).

Definition 6. *(basic) Let $\gamma \not\models \bot$.*

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models_b \mathbf{G}\phi \Leftrightarrow \gamma \models \phi$$

The semantics of $\top$, negation and conjunction are defined analogously to the way this was done for belief formulas (definition 3), but we omit this here and in definitions in the sequel for reasons of presentation.

Hindriks argues in his definition of the agent programming language GOAL [8], that the goal base of the agent does not need to be consistent. Goals in the goal base do not have to be pursued simultaneously and could be achieved at different times. A goal base $\{p, \neg p\}$ should therefore be allowed in agent configurations. The semantics of goal formulas of definition 6 would in this case however have the undesirable characteristic that the inconsistent goal, i.e., the formula $\mathbf{G}\bot$, can be derived given an inconsistent goal base such as $\{p, \neg p\}$. Moreover, given that $\mathbf{G}\bot$ can be derived, any formula $\mathbf{G}\phi$ can be derived.

To avoid these undesired properties, Hindriks requires that individual goals in the goal base are consistent, rather than the goal base as a whole. He then defines the semantics of goal formulas as follows.

Definition 7. *(Hindriks) Let $\forall\phi \in \gamma : \phi \not\models \bot$.*

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models_h \mathbf{G}\phi \Leftrightarrow \exists\phi' \in \gamma : \phi' \models \phi$$

Although this semantics allows for inconsistent goal bases without the possibility to derive the inconsistent goal, it can be considered too restrictive. Suppose for example that an agent has the (consistent) goal base $\{p, q\}$. In this case, one would most likely want the agent to derive that $p \wedge q$ is also a goal, i.e., that $\mathbf{G}(p \wedge q)$ holds. In particular, if the agent has an intention generation rule $\mathbf{G}(p \wedge q) \Rightarrow_{\mathbf{I}} \langle \pi, p \wedge q\rangle$, which represents the idea that the plan π is supposed to achieve a situation in which $p \wedge q$ holds, this rule should be applicable. If execution of the plan π is successful, both goals p and q of the agent would be achieved.

Moreover, if the agent has the inconsistent goal base $\{p, q, \neg p\}$, the given intention generation rule should *also* be applicable. If the plan π would achieve a situation in which $p \wedge q$ holds, part of the goals in the goal base would be achieved. The agent could then pursue the goal $\neg p$ consecutively.

Given these considerations, we propose the following semantic definition, which specifies that $\mathbf{G}\phi$ holds iff there is a consistent subset of the goal base from which ϕ follows.

DEFINITION 8. *(new)*

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi \Leftrightarrow \exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot \textit{ and } \gamma' \models \phi)$$

Note that, in contrast with Hindriks' semantics, we do not need to require that individual goals in the goal base are consistent, because they are "ignored" by this definition[2]. A goal $p \wedge \neg p$ would for example be ignored, as this goal cannot be used to derive $\mathbf{G}\bot$, or any other goals for that matter. Further, note that a formula $\mathbf{G}p \wedge \mathbf{G}\neg p$ is satisfiable under this semantics, without the formula $\mathbf{G}\bot$ being satisfiable (see also the explanation below proposition 4).

Concluding this section, we make three remarks. First, we revisit the position that we addressed at the beginning of this section, which is that goals are often required to be consistent. For the proposed semantics of definition 8, this is not required and the question now is, whether this can be justified. We believe it can, for the following reason. As explained, the rationale behind the requirement of goal consistency is, that an agent should not simultaneously pursue goals that conflict. In our framework however, the intention base contains the plans of the agent, together with goals that are pursued by those plans. It is suggested by the semantics of intention generation of definition 5, that intentions, i.e., the situations that are actively pursued by the agent, are *not* conflicting. An inconsistent goal base thus does not necessarily imply that inconsistent goals are simultaneously pursued. In a framework in which all goals in the goal base can be or are pursued simultaneously, the consistency requirement for goals would indeed have to be adopted and the semantics of goal formulas of definition 6 could then be used.

Secondly, we remark that this paper considers representations of goals without any temporal information on the order in which the goals should be pursued. A representation of goals with a temporal component could be a way of reducing inconsistency, or, more accurately, of reducing what might *appear* to be an inconsistency without the temporal representation. Explorations along these lines are however not within the scope of this paper.

The third remark is about whether semantics of goal formulas should only be based on the goal base, or also on the beliefs of the agent. Generally [4, 17, 8, 19, 20], the view is that an agent cannot have something as a goal that it already believes to be the case. It could be incorporated into our framework as was for example done in [18], by adding the condition $\sigma \not\models \phi$ to the semantics for goal formulas. This would also prevent an agent from having tautologies as goals. We however omit this condition for reasons of simplicity.

3.2 Properties

In this section, we present some propositions relating the definitions of section 3.1.

[2]The requirement could also be omitted from definition 7 if the condition $\phi' \not\models \bot$ would be added to the righthand side of the definition, yielding a close resemblance with definition 8. Definition 7 is however the one provided by Hindriks in [8].

PROPOSITION 1. *Let $\gamma \not\models \bot$. Then the following holds.*

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_b \mathbf{G}\phi \Leftrightarrow \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi \quad (3.1)$$

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_h \mathbf{G}\phi \Rightarrow \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_b \mathbf{G}\phi \quad (3.2)$$

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_b \neg\mathbf{G}\phi \Rightarrow \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_h \neg\mathbf{G}\phi \quad (3.3)$$

Proof: (3.1) If $\gamma \not\models \bot$, we have that $\exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \phi)$ is equivalent with $\exists \gamma' \subseteq \gamma : \gamma' \models \phi$, which is equivalent with $\gamma \models \phi$. (3.2) If $\exists \phi' \in \gamma : \phi' \models \phi$, then $\gamma \models \phi$. (3.3) If $\gamma \not\models \phi$, then $\neg\exists \phi' \in \gamma : \phi' \models \phi$. □

Property (3.1) states that under the assumption of consistency of the goal base, the basic semantics and the new semantics are equivalent. The set of goals, i.e., formulas ϕ for which $\mathbf{G}\phi$ holds in a configuration, is the same for the new semantics as for the basic semantics. Comparing the basic semantics with Hindriks' semantics, we see that the set of goals derivable under Hindriks' semantics is a subset of those derivable under the basic semantics (3.2). The formulas ϕ such that $\neg\mathbf{G}\phi$ is true in a configuration, is the complement of the formulas for which $\mathbf{G}\phi$ holds. We thus have implication (3.3). From the equivalence of basic and new semantics, we can conclude that the set of goals derivable under Hindriks' semantics is a subset of those derivable under the new semantics (corollary 1).

COROLLARY 1. *Let $\gamma \not\models \bot$. Then the following holds.*

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_h \mathbf{G}\phi \Rightarrow \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi$$

The result of this corollary can also be obtained if we relax the constraint of consistency of the entire goal base to consistency of individual goals. This is expressed in the following proposition.

PROPOSITION 2. *Let $\forall \phi \in \gamma : \phi \not\models \bot$. Then the following holds.*

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_h \mathbf{G}\phi \Rightarrow \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi$$

Proof: If $\exists \phi' \in \gamma : \phi' \models \phi$, then $\exists \gamma' \subseteq \gamma : \gamma' \models \phi$ (let $\gamma' = \{\phi'\}$). □

4. SEMANTICS WITH GOAL ADOPTION RULES

In section 3, we presented a number of semantics for goal formulas, based on the goal base of an agent configuration. In this section, we introduce goal adoption rules and present a semantics for goal formulas, based on the goal base *and* these rules.

4.1 Semantics

Below, we define the set of goal adoption rules. A goal adoption rule has a belief and goal condition as the antecedent and a propositional formula as the consequent. Intuitively, it means that if the belief and goal condition in the antecedent hold, the formula in the consequent can be adopted as a goal. As is argued in philosophical logic [7], mental attitudes are conditional by nature.

DEFINITION 9. *(goal adoption rule) The set of goal adoption rules $\mathcal{R}_{\mathsf{GA}}$ is defined as follows:*

$$\{\beta, \kappa \Rightarrow^{+}_{\mathbf{G}} \phi \mid \beta \in \mathcal{L}_{\mathsf{B}}, \kappa \in \mathcal{L}_{\mathsf{G}}, \phi \in \mathcal{L}\}.$$

These goal adoption rules were also proposed in [18]. In that paper, two kinds of semantics of these rules were given. One was based on the *application* of a goal adoption rule over a transition, comparable with the semantics of intention generation rules of definition 5. The second kind, and the one we focus on in this paper, defined the semantics of goal *formulas*, given an agent configuration with a goal base and a set of goal adoption rules. This semantics was however defined for simple forms of goal adoption rules with only a belief condition. Further, it was very restrictive. Basically, a formula $\mathbf{G}\phi$ was specified to be true, if there was a goal adoption rule with true antecedent and a consequent equivalent to ϕ. In this section, we propose a semantics for goal formulas that is based on goal adoption rules with belief *and* goal condition and that is less restrictive than the one provided in [18].

The semantics we propose is based on default logic. The general idea is as follows. Goal adoption rules are transformed into propositional default rules. This set of default rules has a number of (consistent) extensions, which represent the sets of compatible goals that an agent could derive on the basis of its rules. Given an agent configuration and an extension of the default rules generated from the goal adoption rules in this configuration, we define the semantics of goal formulas.

The idea of using default logic to define the semantics of goal adoption rules was taken from the BOID framework by Dastani and Van der Torre [5]. This framework was in turn inspired by Thomason [16], who uses default logic to develop a formalism to integrate reasoning about desires with planning, and Horty [9], who showed how motivational attitudes can be formalized in default logic. A detailed comparison with this work is beyond the scope of this paper. One of the main differences however between our proposal and the BOID framework is that we use default logic to *define* the semantics of goal formulas, whereas Dastani and Van der Torre assume some modal logic satisfiability relation for goal formulas and use default logic to generate consistent sets of goal formulas.

Default logic is designed to handle possibly conflicting defeasible rules. Goal adoption rules could be conflicting and default logic is therefore a natural way to interpret these rules. This can be illustrated by considering an agent with the following rules for deriving goals: if the agent believes that it's raining, it can derive the goal to take the bus, and if it has the goal to be on time, it can derive the goal not to take the bus (but to take a taxi instead, for example). Suppose the agent believes it's raining and wants to be on time, then it has a reason to derive the goal to take the bus *and* it has a reason to derive the goal not to take the bus. Rather than deriving both conflicting goals, default logic gives rise to two extensions or compatible goal sets: one containing the goal to take the bus and one not to take the bus. This thus provides for the possibility to define agents with the desirable characteristic of planning on the basis of a compatible set of goals.

Below, we define the function f that takes a set of goal adoption rules with only a goal condition and yields a set of propositional default rules. It will become clear later on why we define this function for rules without a belief condition. In the definition, we use CL to denote the set of conjunctions of goal literals. A goal literal is a formula of the form $\mathbf{G}\phi$ or $\neg\mathbf{G}\phi$, where $\phi \in \mathcal{L}$. Formulas of the former kind are called positive goal literals and formulas of the latter type negative goal literals. The formula $\top$ is treated as a positive goal literal. We use a function pl that takes a conjunction of goal literals and yields a set containing the propositional parts of the positive goal literals of this conjunction[3]. The function nl similarly yields the set of propositional parts of negative goal literals of a conjunction. Further, we use a function dnf that takes a set of goal adoption rules of the form $\kappa \Rightarrow^+_\mathbf{G} \phi$ and yields these rules with the antecedent transformed into disjunctive normal form. We map goal adoption rules to disjunctive normal form, because rules of this form can be intuitively mapped to default rules. $\mathcal{R}_{\mathsf{GA}_{\mathsf{DNF}}}$ is the set of goal adoption rules with the antecedent in disjunctive normal form, i.e., $\mathcal{R}_{\mathsf{GA}_{\mathsf{DNF}}} = \{\bigvee_{1 \leq i \leq n} cl_i \Rightarrow^+_\mathbf{G} \chi \mid n > 0, cl_i \in CL, \chi \in \mathcal{L}\}$. Finally, the number of elements in a set S is denoted by $|S|$.

Definition 10. *(goal adoption rules to default rules) Let* DR *denote the set of propositional default rules. Let* $cl, cl_1, \ldots, cl_k \in CL$*. The function* $t : \mathcal{R}_{\mathsf{GA}_{\mathsf{DNF}}} \to \wp(\mathsf{DR})$*, taking a goal adoption rule and yielding a set of default rules, is then defined as follows, where* $\phi_i \in pl(cl)$ *for* $1 \leq i \leq m$ *and* $\psi_j \in nl(cl)$ *for* $1 \leq j \leq n$ *with* $|pl(cl)| = m$ *and* $|nl(cl)| = n$*. If* $n = 0$*, the sequence* $\psi_1, \ldots, \psi_n$ *is empty.*

$$\begin{array}{lcl} t(cl \Rightarrow^+_\mathbf{G} \chi) & = & \{\phi_1 \wedge \ldots \wedge \phi_m : \\ & & \neg\psi_1, \ldots, \neg\psi_n, \chi/\chi\} \\ t(cl_1 \vee \ldots \vee cl_k \Rightarrow^+_\mathbf{G} \chi) & = & \bigcup_{1 \leq i \leq k} t(cl_i \Rightarrow^+_\mathbf{G} \chi) \end{array}$$

The function $f : \wp(\mathcal{R}_{\mathsf{GA}}) \to \wp(\mathsf{DR})$ *taking a set of goal adoption rules of the form* $\kappa \Rightarrow^+_\mathbf{G} \phi$ *and yielding a set of default rules, is then defined as follows.*

$$f(\mathsf{GA}) = \bigcup_{r \in dnf(\mathsf{GA})} t(r)$$

We explain this definition using an example. Consider the goal adoption rules $g_1 = \mathbf{G}p \wedge \neg\mathbf{G}q \Rightarrow^+_\mathbf{G} r$, $g_2 = \top \Rightarrow^+_\mathbf{G} p$ and $g_3 = \mathbf{G}r \Rightarrow^+_\mathbf{G} q$, corresponding with the default rules $d_1 = p : \neg q, r/r$, $d_2 = \top : p/p$ and $d_3 = r : q/q$ respectively.

When transforming a goal adoption rule with a conjunction as the antecedent, the propositional parts of positive goal literals are mapped onto the prerequisite of a default rule, whereas the propositional parts of negative goal literals are negated and mapped onto the justification of the default rule. This reflects the difference between for example the formulas $\mathbf{G}\neg q$ and $\neg\mathbf{G}q$: the former represents the *presence* of a goal $\neg q$, whereas the latter represents the *absence* of the goal q. Considering goal adoption rules g_1 and g_2, the set $\{p, r\}$ is an extension of the default rules d_1 and d_2. This reflects our intuition about goal adoption rules: p can be derived on the basis of the second rule and if p is a goal and q is not, we can derive goal r.

If we consider the default rules d_1, d_2 and d_3, we have that the set $\{p, r, q\}$ is *not* an extension of these rules. This is due to the fact that q, which was derived using rule d_3, is inconsistent with the justification $\neg q$ of rule d_1. This corresponds with our intuition about goal adoption rules: given rule g_1, r can only be a goal if q is not. The goals r and q thus cannot be part of the same extension.

Negative goal literals are mapped to a sequence of justifications, rather than to one conjunctive justification. The

[3]The propositional part of the positive goal literal $\top$ is the propositional formula $\top$. Also, if the number of positive goal literals is 0, the function pl yields the set $\{\top\}$.

reason is, that we want to allow goal adoption rules such as $\neg \mathbf{G}p \wedge \neg \mathbf{G}\neg p \Rightarrow^{+}_{\mathbf{G}} q$, specifying that goal q can be adopted if neither p nor $\neg p$ is a goal. If we would map this rule to the default rule $\top : p \wedge \neg p \wedge q/q$, we would get an inconsistent justification and the rule would never be applicable. The rule $\top : p, \neg p, q/q$ on the other hand does the job.

The consequent χ of a goal adoption rule is added to the justification, because we only want to derive a new goal if it is consistent with the already derived ones. Further, goal adoption rules without negative goal literals then yield so-called normal default rules, i.e., rules of the form $\phi : \chi/\chi$. Normal default rules have a number of desirable characteristics, such as the fact that normal default theories always have extensions [3].

Moreover, a goal adoption rule such as $\mathbf{G}p \vee \mathbf{G}q \Rightarrow^{+}_{\mathbf{G}} r$ with a disjunctive goal formula in the antecedent is transformed into the set of *multiple* defaults $\{p : r/r, q : r/r\}$. The rationale is, that the goal r can be derived if either p or q is a goal. This is established through this set of default rules, because if p has been derived as a goal, the first rule can be applied to derive r. Alternatively, r can also be derived using the second rule if q has been derived as a goal.

The following function transforms goal bases to goal adoption rules and will be used in definition 12.

Definition 11. *(goal base to goal adoption rules) The function $g : \wp(\mathcal{L}) \to \wp(\mathcal{R}_{\mathsf{GA}})$, taking a goal base and yielding a set of goal adoption rules, is defined as follows: $g(\gamma) = \{\top \Rightarrow^{+}_{\mathbf{G}} \phi \mid \phi \in \gamma\}$.*

In the definition of the semantics of goals below, we transform the goal adoption rules generated from the goal base, as well as the goal adoption rules in the rule base of the configuration, to default rules. That is, we only take those goal adoption rules for which the belief condition holds in the given configuration. These rules can be transformed into default rules by means of the function of definition 10, if we remove the (true) belief condition. Given an extension of the generated default rules, we define that $\mathbf{G}\phi$ holds iff ϕ follows from this extension.

Definition 12. *(semantics of goals) Let $\mathcal{R} = \langle \mathsf{IG}, \mathsf{GA}\rangle$, where $\mathsf{GA} \subseteq \mathcal{R}_{\mathsf{GA}}$ is a finite set of goal adoption rules. Let GA' be defined as*

$$\{\kappa \Rightarrow^{+}_{\mathbf{G}} \phi \mid \exists(\beta, \kappa \Rightarrow^{+}_{\mathbf{G}} \phi) \in \mathsf{GA} : \langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models_{\mathcal{L}_{\mathsf{B}}} \beta\}.$$

Let E be an extension of $\langle \emptyset, f(g(\gamma)) \cup f(\mathsf{GA}')\rangle$. The default semantics $\models^{E}_{d}$ for goal formulas in the presence of these goal adoption rules, given the extension E, is then as follows.

$$\begin{array}{lcl}
\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \mathbf{G}\phi & \Leftrightarrow & E \models \phi \\
\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \neg\kappa & \Leftrightarrow & \langle \sigma, \gamma, \iota, \mathcal{R}\rangle \not\models^{E}_{d} \kappa \\
\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \kappa \wedge \kappa' & \Leftrightarrow & \langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \kappa \textit{ and} \\
 & & \langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \kappa'
\end{array}$$

Note that the set of facts, i.e., the first component of a default theory, is empty in our case. In the sequel, we will therefore omit the set of facts and speak of extensions of a set of default rules. We chose to transform the goal base into default rules, rather than taking these as facts and considering extensions of $\langle \gamma, f(\mathsf{GA}')\rangle$. The reason is, that we want to allow γ to be inconsistent. A default theory with an inconsistent set of facts only has one extension, i.e., the inconsistent extension. The following definition will be used in the next section.

Definition 13. *(extension of a configuration) Let $\mathcal{R} = \langle \mathsf{IG}, \mathsf{GA}\rangle$ and let $c = \langle \sigma, \gamma, \iota, \mathcal{R}\rangle$ be an agent configuration. Let GA' be $\{\kappa \Rightarrow^{+}_{\mathbf{G}} \phi \mid \exists(\beta, \kappa \Rightarrow^{+}_{\mathbf{G}} \phi) \in \mathsf{GA} : c \models_{\mathcal{L}_{\mathsf{B}}} \beta\}$. E is then said to be an extension of the configuration c, iff E is an extension of $f(g(\gamma)) \cup f(\mathsf{GA}')$.*

4.2 Properties

In this section, we investigate some properties of the default semantics of goals.

The first theorem specifies the following: if a configuration contains a goal adoption rule of which the antecedent is true given an extension of the defaults generated on the basis of this configuration, and the consequent of this rule is consistent with this extension, then the consequent is a goal in this configuration. This theorem formalizes our main desired characteristic of the semantics of goals, being that if the antecedent of a goal adoption rule holds, the consequent is a goal.

Theorem 1. *Let $c = \langle \sigma, \gamma, \iota, \mathcal{R}\rangle$ be a configuration, let $\mathcal{R} = \langle \mathsf{IG}, \mathsf{GA}\rangle$ and let E be an extension of c. Then the following holds.*

If $\exists(\beta, \kappa \Rightarrow^{+}_{\mathbf{G}} \phi) \in \mathsf{GA}$:

$$c \models_{\mathcal{L}_{\mathsf{B}}} \beta \textit{ and } c \models^{E}_{d} \kappa \textit{ and } E \not\models \neg\phi$$

$$\textit{then } c \models^{E}_{d} \mathbf{G}\phi.$$

Proof: Let $\kappa = \bigvee_{1 \le k \le o} cl_k$ with $o > 0, cl_k \in CL$. Since $c \models^{E}_{d} \kappa$ by assumption, it must be the case by definition 12 that $c \models^{E}_{d} cl_k$, for some $1 \le k \le o$. Assume that $c \models^{E}_{d} cl_k$ and let $\phi_i \in pl(cl_k)$ for $1 \le i \le m$ and $\psi_j \in nl(cl_k)$ for $1 \le j \le n$ with $|pl(cl_k)| = m$ and $|nl(cl_k)| = n$. If $c \models^{E}_{d} cl_k$, it must be the case by definition 12, that $E \models \bigwedge_{1 \le i \le m} \phi_i$ and $E \not\models \psi_j$ for $1 \le j \le n$. We also have that $E \not\models \neg\phi$. The default $\phi_1 \wedge \ldots \wedge \phi_m : \neg\psi_1, \ldots, \neg\psi_n, \phi/\phi$ is thus applicable to E. As E is closed under the application of applicable defaults, it must be the case that $E \models \phi$ and thus by definition 12, we can conclude that $c \models^{E}_{d} \mathbf{G}\phi$. □

The following proposition specifies a property of the translation of goal adoption rules into default rules, with respect to conjunctions of positive goal literals in the antecedent.

Proposition 3. *A goal adoption rule of the form $r = \bigwedge_{1 \le i \le m} \mathbf{G}\phi_i \Rightarrow^{+}_{\mathbf{G}} \chi$ is equivalent with the rule $r' = \mathbf{G}(\bigwedge_{1 \le i \le m} \phi_i) \Rightarrow^{+}_{\mathbf{G}} \chi$, i.e., $f(\{r\}) = f(\{r'\})$.*

Proof: Immediate from definition 10. □

The following proposition establishes a number of properties of the $\mathbf{G}$ operator under the default semantics for goal formulas.

Proposition 4.

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} \mathbf{G}(\phi \to \psi) \to (\mathbf{G}\phi \to \mathbf{G}\psi) \quad (4.1)$$

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \models^{E}_{d} (\mathbf{G}\phi \wedge \mathbf{G}\psi) \leftrightarrow \mathbf{G}(\phi \wedge \psi) \quad (4.2)$$

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \not\models^{E}_{d} \mathbf{G}\phi \wedge \mathbf{G}\neg\phi \quad (4.3)$$

$$\langle \sigma, \gamma, \iota, \mathcal{R}\rangle \not\models^{E}_{d} \mathbf{G}\bot \quad (4.4)$$

Proof: Let $c = \langle \sigma, \gamma, \iota, \mathcal{R} \rangle$.
(4.1) We have to show that $c \models_d^E \mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi)$. This means we have to show that $c \models_d^E \mathbf{G}(\phi \rightarrow \psi) \Rightarrow (c \models_d^E \mathbf{G}\phi \Rightarrow c \models_d^E \mathbf{G}\psi)$. Assume that $c \models_d^E \mathbf{G}(\phi \rightarrow \psi)$ and $c \models_d^E \mathbf{G}\phi$. This means that $E \models \phi \rightarrow \psi$ and $E \models \phi$ (definition 12). From this we can conclude that $E \models \psi$, which is equivalent with $c \models_d^E \mathbf{G}\psi$, yielding the desired result.
(4.2) We have to show that $c \models_d^E \mathbf{G}\phi \wedge \mathbf{G}\psi \Leftrightarrow c \models_d^E \mathbf{G}(\phi \wedge \psi)$. This means we have to show that ($c \models_d^E \mathbf{G}\phi$ and $c \models_d^E \mathbf{G}\psi$) $\Leftrightarrow c \models_d^E \mathbf{G}(\phi \wedge \psi)$, which is equivalent with ($E \models \phi$ and $E \models \psi$) $\Leftrightarrow E \models \phi \wedge \psi$ (definition 12). This is obviously the case.
(4.3) We have to show that $c \not\models_d^E \mathbf{G}\phi \wedge \mathbf{G}\neg\phi$. This means we have to show that it is not the case that $c \models_d^E \mathbf{G}\phi$ and $c \models_d^E \mathbf{G}\neg\phi$, which is equivalent with $E \models \phi$ and $E \models \neg\phi$. We have that E is consistent (section 2.2), which means that this is a contradiction, yielding the desired result.
(4.4) We have to show that $c \not\models_d^E \mathbf{G}\bot$, i.e., that $E \not\models \bot$. This follows immediately from the fact that E is consistent. □

(4.1) specifies that goals are closed under classical logical consequence. This property is satisfied by all semantics presented in this paper, except Hindriks' semantics (definition 7). (4.2) says that separate goals can be combined into one. This property does not hold for the semantics of definitions 7 and 8. A formula $\mathbf{G}p \wedge \mathbf{G}\neg p$ is for example satisfiable under Hindriks' semantics. The goals p and $\neg p$ can however not be combined into one, as this would lead to the derivation of the inconsistent goal. Combining goals under the default semantics will never lead to the derivation of the inconsistent goal, for the following reason. A goal formula evaluated under the default semantics is evaluated given an extension. An extension is always consistent and therefore a formula $\mathbf{G}\phi \wedge \mathbf{G}\neg\phi$ is not satisfiable (4.3). Property (4.4) is a desirable characteristic that is satisfied by all semantics discussed in this paper.

The last result of this section relates the semantics of definition 8 to the default semantics, given that the set of goal adoption rules is empty.

THEOREM 2. *Let $\mathcal{R} = \langle \mathsf{IG}, \emptyset \rangle$. Then the following holds.*

$$\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi \Leftrightarrow \exists E : \langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_d^E \mathbf{G}\phi$$

A set of propositional formulas γ' is a maximal consistent subset of a set of formulas γ iff $\gamma' \subseteq \gamma$, $\gamma' \not\models \bot$ and $\neg\exists\phi \in \gamma : \phi \notin \gamma'$ and $\{\phi\} \cup \gamma' \not\models \bot$.

LEMMA 1. *There is a consistent subset γ' of γ such that $\gamma' \models \phi$ iff there is a maximal consistent subset γ' of γ such that $\gamma' \models \phi$. Further, γ' is a maximal consistent subset of γ iff γ' is an extension of $\{\top : \phi/\phi \mid \phi \in \gamma\}$ [3].*

Proof of theorem 2: By definition 12 and the fact that $\mathsf{GA} = \emptyset$, E must be an extension of $f(g(\gamma))$. $\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \models_n \mathbf{G}\phi$ means that there is a consistent subset γ' of γ such that $\gamma' \models \phi$. By lemma 1, this is equivalent to there being a maximal consistent subset γ' of γ such that $\gamma' \models \phi$. We thus have to show that there is a maximal consistent subset γ' of γ such that $\gamma' \models \phi$ iff there is an extension E of $f(g(\gamma))$ such that $E \models \phi$. By definition 11, we have that $g(\gamma) = \{\top \Rightarrow_{\mathbf{G}}^{+} \phi \mid \phi \in \gamma\}$ and therefore $f(g(\gamma)) = \{\top : \phi/\phi \mid \phi \in \gamma\}$. By lemma 1, we then have that γ' is a maximal consistent subset of γ iff γ' is an extension of $f(g(\gamma))$, yielding the desired result. □

Theorem 2 does not hold for arbitrarily composed goal formulas, because a formula $\mathbf{G}\phi \wedge \mathbf{G}\neg\phi$ is satisfiable under $\models_n$, but not under $\models_d^E$ (proposition 4). Goal formulas evaluated under $\models_d^E$ are evaluated under *one* extension, whereas each conjunct of a conjunction evaluated under $\models_n$, can be evaluated with respect to a *different* consistent subset. A goal formula $\mathbf{G}p \wedge \mathbf{G}q$ for example is only satisfiable under $\models_d^E$ if there is an extension of a relevant set of goal adoption rules, in which both p and q occur, or in other words, if p and q are compatible.

The evaluation of goal formulas with respect to one compatible set of goals rather than providing for the possibility to evaluate conjuncts of a goal formula with respect to different sets, was introduced to yield the validity of theorem 1. If a goal formula would not be evaluated with respect to one and the same extension, it could be the case that the antecedent of a goal rule holds, while the goal in the consequent is not a goal, thereby invalidating theorem 1.[4]

The fact that a goal formula is evaluated with respect to a compatible set of goals yields desirable behavior, also if we consider intention generation.

DEFINITION 14. *(semantics of intention generation) Let $c = \langle \sigma, \gamma, \iota, \mathcal{R} \rangle$ be a configuration, let $\mathcal{R} = \langle \mathsf{IG}, \mathsf{GA} \rangle$ and let $\beta, \kappa \Rightarrow_{\mathbf{I}} \langle \phi, \pi \rangle \in \mathsf{IG}$ be an intention generation rule. The semantics of applying this rule is then as follows, where $\iota' = \iota \cup \{\langle \phi, \pi \rangle\}$ and $\delta = \{\phi \mid \langle \pi, \phi \rangle \in \iota\}$.*

$$\frac{c \models_{\mathcal{L}_\mathsf{B}} \beta \qquad \exists E : c \models_d^E \kappa \qquad \{\phi\} \cup \delta \not\models \bot}{\langle \sigma, \gamma, \iota, \mathcal{R} \rangle \rightarrow \langle \sigma, \gamma, \iota', \mathcal{R} \rangle}$$

This semantics of intention generation specifies that an intention generation rule is applicable if an extension of the relevant configuration exists under which the goal condition holds.[5] An intention is thus generated on the basis of compatible goals, which could be considered desirable.

We conclude with a remark about persistency of goals. Under the default semantics, goals may vary during execution of the agent as goals depend on goal adoption rules, which depend on beliefs and beliefs may change. One could argue that this is not the appropriate level of persistency for goals. However, in our framework goals are used to generate intentions and these intentions can be defined such that they have a higher level of persistency than goals. A related issue that we do not consider in this paper, is the question of whether a goal, or even a goal adoption rule, should be removed from the goal base or rule base if the goal is achieved. If goals are not removed, they will be of the maintenance type.

5. CONCLUSION

We have presented two semantics for declarative goals in an agent programming setting. One was based only on the goal base of the agent and the other was based on the goal base

[4]A formula $\mathbf{G}p \wedge \mathbf{G}q$ could for example hold, because $\mathbf{G}p$ is derived on the basis of one extension and $\mathbf{G}q$ on the basis of another. A default rule $p \wedge q : r/r$, corresponding with a goal adoption rule $\mathbf{G}p \wedge \mathbf{G}q \Rightarrow_{\mathbf{G}}^{+} r$, is however only applicable and therefore able to derive r, if $p \wedge q$ follows from *one* extension.

[5]Note that this corresponds with the credulous semantics of default logic.

and a set of goal adoption rules. The former was motivated by an analysis of other proposals that have been done in the literature and was compared to them. The latter was defined by specifying a mapping of goal adoption rules to propositional default rules. We have shown that this default semantics has a number of desirable characteristics and that it is moreover closely related to the semantics that does not take into account goal adoption rules.

Concluding, we maintain that a systematic analysis of semantics of declarative goals in agent programming is essential, in order to be able to understand how we can best incorporate these in agent programming languages. This paper contributes to this effort.

Acknowledgements

We would like to thank Henry Prakken for his helpful comments on an earlier version of this paper.

6. REFERENCES

[1] G. Antoniou. *Nonmonotonic Reasoning.* Artificial Intelligence. The MIT Press, Cambridge, Massachusetts, 1997.

[2] M. E. Bratman. *Intention, plans, and practical reason.* Harvard University Press, Massachusetts, 1987.

[3] G. Brewka, J. Dix, and K. Konolige. *Nonmonotonic reasoning: an overview.* CSLI Publications, Stanford, 1997.

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

[5] M. Dastani and L. van der Torre. Programming BOID-Plan agents: deliberating about conflicts among defeasible mental attitudes and plans. In *Proceedings of the Third Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04)*, pages 706–713, New York, USA, 2004.

[6] M. Dastani, M. B. van Riemsdijk, F. Dignum, and J.-J. Ch. Meyer. A programming language for cognitive agents: goal directed 3APL. In *Programming multiagent systems, first international workshop (ProMAS'03)*, volume 3067 of *LNAI*, pages 111–130. Springer, Berlin, 2004.

[7] B. Hansson. An analysis of some deontic logics. In *Nous 3*, 1969.

[8] K. V. Hindriks, F. S. de Boer, W. van der Hoek, and J.-J. Ch. Meyer. Agent programming with declarative goals. In *Intelligent Agents VI - Proceedings of the 7th International Workshop on Agent Theories, Architectures, and Languages (ATAL'2000)*, Lecture Notes in AI. Springer, Berlin, 2001.

[9] J. F. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23(1):35–65, 1994.

[10] A. F. Moreira, R. Vieira, and R. H. Bordini. Extending the operational semantics of a BDI agent-oriented programming language for introducing speech-act based communication. In *Proceedings of the first International Workshop on Declarative Agent Languages and Technologies (DALT03)*, pages 129–145, 2003.

[11] G. D. Plotkin. A Structural Approach to Operational Semantics. Technical Report DAIMI FN-19, University of Aarhus, 1981.

[12] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann, 1991.

[13] R. Reiter. A logic for default-reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[14] J. Thangarajah, L. Padgham, and M. Winikoff. Detecting and exploiting positive goal interaction in intelligent agents. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems (AAMAS'03)*, pages 401–408, Melbourne, 2003.

[15] J. Thangarajah, M. Winikoff, L. Padgham, and K. Fischer. Avoiding resource conflicts in intelligent agents. In F. van Harmelen, editor, *Proceedings of the 15th European Conference on Artifical Intelligence 2002 (ECAI 2002)*, Lyon, France, 2002.

[16] R. H. Thomason. Desires and defaults: A framework for planning with inferred goals. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 702–713, San Francisco, 2000. Morgan Kaufmann.

[17] W. van der Hoek, B. van Linder, and J.-J. Ch. Meyer. An integrated modal approach to rational agents. In M. Wooldridge and A. S. Rao, editors, *Foundations of Rational Agency*, Applied Logic Series 14, pages 133–168. Kluwer, Dordrecht, 1998.

[18] M. B. van Riemsdijk, M. Dastani, F. Dignum, and J.-J. Ch. Meyer. Dynamics of declarative goals in agent programming. In J. A. Leite, A. Omicini, P. Torroni, and P. Yolum, editors, *Proceedings of the second international workshop on Declarative agent languages and technologies (DALT'04)*, pages 17–32, 2004.

[19] M. B. van Riemsdijk, W. van der Hoek, and J.-J. Ch. Meyer. Agent programming in Dribble: from beliefs to goals using plans. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems (AAMAS'03)*, pages 393–400, Melbourne, 2003.

[20] M. Winikoff, L. Padgham, J. Harland, and J. Thangarajah. Declarative and procedural goals in intelligent agent systems. In *Proceedings of the eighth international conference on principles of knowledge respresentation and reasoning (KR2002)*, Toulouse, 2002.