Auton Agent Multi-Agent Syst (2009) 18:471–500
DOI 10.1007/s10458-008-9067-4

Goals in conflict: semantic foundations of goals in agent programming

M. Birna van Riemsdijk · Mehdi Dastani · John-Jules Ch. Meyer

Published online: 16 October 2008

Abstract This paper addresses the notion of (declarative) goals as used in agent programming. Goals describe desirable states, and semantics of these goals in an agent programming context can be defined in various ways. We focus in this paper on the representation of conflicting goals. In particular, we define two semantics for goals, one for unconditional goals and one for conditional goals. The first is based on propositional logic, and the latter is based on default logic. We establish relations between and properties of these semantics.

Keywords Agent programming languages · Goals · Logic

1 Introduction

An important line of research in the agent systems field is research on agent programming languages and frameworks [3]. In this paper, we are in particular interested in those languages and frameworks that focus on the programming of *cognitive agents* (see, e.g., [54]). Cognitive agents are agents endowed with high-level mental attitudes, such as beliefs, goals, desires, intentions, plans, etc.

This title was inspired by the title of the PhD thesis of Harrenstein: *Logic in conflict: logical explorations in strategic equilibrium* [25].

M. B. van Riemsdijk (✉)
Technische Universiteit Delft, Delft, The Netherlands
e-mail: m.b.vanriemsdijk@tudelft.nl; riemsdijk@pst.ifi.lmu.de

M. Dastani · J.-J. Ch. Meyer
Utrecht University, Utrecht, The Netherlands
e-mail: mehdi@cs.uu.nl

J.-J. Ch. Meyer
e-mail: jj@cs.uu.nl

In cognitive agent programming languages, cognitive notions are *first class citizens*. Much research in this area thus investigates which cognitive notions are suitable for use in a programming language, and what kind of language constructs one could use for implementing them. Most cognitive agent programming languages and frameworks have at least an informational component (often called "beliefs") and a procedural component (often called "plans"). In the past several years, there has been an increasing amount of research concerning frameworks that also have a motivational component (often called "goals") [4, 17, 28, 32, 38, 40, 48–50, 55–57, 59, 61]. The idea is then that an agent executes plans in order to reach its goals, depending on what it believes about the world.

An agent may in general pursue multiple goals at the same time. An issue that arises in this context, is that some goals are *conflicting*, in the sense that it is undesirable to pursue these simultaneously. One reason for not pursuing certain goals simultaneously, is that the plans for reaching these goals may interfere. For example, an agent may try executing plans for reaching the goal to be in Paris and reaching the goal to be in Amsterdam simultaneously, while the agent is currently in Brussels. A parallel execution of these plans is not likely to support the effective achievement of both goals. There can also be other reasons that an agent should not pursue certain goals simultaneously. For example, an agent might be hungry, and consequently have the goals to have a pizza and to have sushi. However, it will only need to achieve one of these goals in order not to be hungry anymore.

To prevent the simultaneous pursuit of conflicting goals, it is important to know and to be able to represent which goals are conflicting. One way to do this, is to reason about the plans that may be used to achieve the agent's goals in order to determine conflicts [49, 51]. Another way to do it is to determine conflicts not on the level of plans, but solely on the level of *goals* [40, 56].[1] In this paper, we follow the latter approach.

Such an approach requires that the agent programmer can *represent goals and their mutual conflicts*. The investigation of how to do this has not received much attention in the literature on cognitive agent programming. In [40], the Jadex cognitive agent framework is extended with an inhibition relation between goals which explicitly expresses that two goals are conflicting.

While Jadex supports the programming of cognitive agents using XML and Java, we investigate in this paper the representation of conflicting goals in the context of *logic-based* agent programming languages. Not much work has been done yet in this area, and therefore we believe it is important to study this issue formally and from a foundational perspective in order to get a better understanding of the various techniques that may be used, and of their properties.

The main contribution of this paper is the proposal of two ways of representing conflicting goals, accompanied by the definition of what the goals of the agent are, given these representations. We provide one definition based on a representation of *unconditional* goals, and one that is based on the representation of *conditional* goals. We investigate properties of and relations between these definitions. This paper builds on our earlier work as published in [56]. The main differences are that in the present paper we provide more details regarding the motivations for the definitions, the analysis of the properties, and related work. Further, we investigate not only the relation between various semantics for unconditional goals, but also properties *of* these semantics. Moreover, we slightly change the semantics of conditional goals, and consequently consider different properties.

The paper is organized as follows. In Sect. 2, we provide some more background on the role of goals in logic-based cognitive agent programming languages, and we discuss how

[1] Thangarajah et al. [49, 51] also involve goals in their approach, but as an integral part of the agent's plans. The reasoning is based on the resources used by the agent's plans, and results of the execution of an agent's plans, respectively.

conflicting goals have been modeled in agent programming frameworks. In Sects. 3 and 4, we present the formal definitions and analysis of unconditional and conditional goals, respectively. Section 5 discusses related work in normative systems, and in Sect. 6 we conclude the paper.

2 Modeling conflicting goals

In this section, we sketch the context for this work by briefly showing how goals are used in logic-based cognitive agent programming languages (Sect. 2.1), discussing ontological aspects of the notion of goal (Sect. 2.2), and discussing which kinds of techniques have been used in other approaches for modeling conflicting goals and which techniques we will use in this paper (Sect. 2.3).

2.1 Logic-based cognitive agent programming

In logic-based cognitive agent programming languages that support the representation of goals [14, 28, 59], goals are typically used to *select plans*. This is done by means of rules of the form $\kappa \mid \beta \Rightarrow \pi$, which intuitively represent that plan π may be executed if the goal condition κ holds, and β is believed to be the case.

The conditions κ and β are *logical formulas*, by means of which one can express that the agent has certain goals or beliefs. The formula $\mathbf{G}(pizza)$, for example, expresses that the agent has the goal to have pizza. The truth of these formulas is evaluated on a representation of the goals of the agent (the goal base γ) and a representation of the beliefs (the belief base σ), respectively.

In such a setting, one thus needs to determine what γ and σ look like, and consecutively it needs to be defined when γ or σ satisfy a formula κ or β, i.e., when $\gamma \models_g \kappa$ or $\sigma \models_b \beta$ hold, respectively.[2] In this paper, we investigate various ways of defining γ and $\models_g$, with a focus on the representation of conflicting goals.

We build on propositional logic and propositional default logic [46] to define γ and $\models_g$. We use these relatively simple logics, as we aim to investigate the *semantic foundations* of (conflicting) goals. For this purpose, it suffices to consider these simple logics. While we thus do not aim to provide a full-fledged representation of goals which can be used directly in a practical agent programming language, we view the work in this paper as providing a solid foundation for such an effort. We have shown in previous work how a cognitive agent programming language based on propositional logic [59] can be extended to a language that builds on a Prolog-like Horn clause logic with variables [14].[3] A platform for the latter language has been implemented in which beliefs and goals are implemented using Prolog.[4]

It is important to note that in this paper we focus on the *definition* of when a goal base satisfies a formula κ, i.e., we *define* when $\gamma \models_g \kappa$ holds. We will not address how to implement the relation $\models_g$, i.e., how to *derive* that κ holds, given a goal base γ. Here, we confine ourselves to remarking that we see in particular answer set programming as a tool for implementing the ideas presented in this paper, as there is a close relation between default logic and answer set programming [21, 22]. Several answer set solvers exist (most notably

[2] Sometimes, the belief base is taken into account as well when defining $\models_g$, in which case one thus defines when $\langle \sigma, \gamma \rangle \models_g \kappa$ holds.

[3] This language does not provide support for the representation of *conflicting* goals.

[4] http://www.cs.uu.nl/3apl/download.html.

Smodels [37] and DLV [33]), and we anticipate that the ideas based on default logic as presented in this paper should be implementable using such solvers.

2.2 Ontological confusion

In the literature on cognitive agents, several different perspectives on goals have appeared. The distinction between goals and related motivational attitudes such as desires and intentions is not always clear (see also [54, Chapter 5] and [58]). It is beyond the scope of this paper to clear up this ontological confusion. Nevertheless, we briefly discuss the different properties that have been attributed to motivational attitudes, and explain what perspective we take in this paper.

In [11, 44], so-called Belief Desire Intention (BDI) logics are proposed in which various motivational attitudes, including goals, are formalized. In both approaches, goals are required to be consistent. Desires are allowed to be inconsistent, but these are not formalized. In [45], the goal operator of [44] is replaced by a desire operator, i.e., desires are required to be consistent in that paper. The characteristic property of intentions in these logics is that an agent is required to be committed to its intentions. That is, an agent may not drop intentions for arbitrary reasons, which means that intentions have a certain persistency. Various levels of commitment are proposed, such as not dropping intentions until they are believed to be achieved, or believed to be impossible.

Directly or indirectly inspired by the BDI logics, several cognitive agent programming frameworks have been introduced. Each of these frameworks proposes a particular interpretation of cognitive concepts in a computational context. This has resulted in a wide variety of different notions. For example, in [27], goals are procedural and similar to plans, while in other approaches intentions are equated with plans [43]. In, e.g., [28], goals are declarative, i.e., describe a state that is to be reached, and other approaches propose a notion of goal that incorporates both procedural and declarative aspects (see, e.g., [49, 58, 61]).

Moreover, in [61], goals are required to be non-conflicting. Other approaches, however, maintain that it is very well possible that an agent has conflicting goals [28, 38, 40, 49, 51] (e.g., because the user of the agent endows the agent with conflicting goals), and propose ways of handling this. Also, goals are sometimes persistent in that they are not dropped until they are believed to be achieved (see, e.g., [28, 59, 61]). However, other approaches take a more liberal approach and allow goals to be dropped (or temporarily suspended) also for other reasons (e.g., because a more important but conflicting goal has been adopted, or because the reason for adopting the goal is not valid anymore) [38, 40, 44, 58].

We can thus see that there is no consensus as to what properties the various motivational attitudes should have. Depending on the context and on the issue under investigation, different perspectives are taken. The perspective we take here is the following. We are interested in the modeling of conflicting motivational attitudes. Since motivational attitudes are typically called "goals" in agent programming frameworks, we also use this term in this paper.

Especially in the context of conflicting goals, we believe it is important to make a *distinction* between the goals that the agent is *currently* pursuing, and those that it would in principle *like* to pursue. The former ones should be free of conflict, but the latter ones may well be conflicting, and these are the ones we are concerned with in this paper. We also believe it is important that the behavior of the agent has a certain stability, and that it is thus important that goals have some persistency. However, a relatively high level of stability is particularly important for the goals that the agent is actively pursuing, since those directly influence the agent's behavior. For the goals that the agent would in principle like to pursue, we believe

that it is natural to drop goals also if the reason for adopting the goal is not valid anymore.[5] If the reason for adopting a goal is not valid anymore, and the agent has not started the pursuit of the goal yet, it does not seem to make sense to keep the goal.

2.3 Representing conflicting goals

We identify three important aspects that are relevant for a comprehensive approach for incorporating conflicting goals in agent programming frameworks. First of all, agent programming languages that support dealing with conflicting goals should *allow to identify* that certain goals are conflicting. That is, if an agent is to deal with conflicting goals in an appropriate way, it is important to at least know which goals are conflicting. Second, given that we know which goals are conflicting, agent programming frameworks should provide a mechanism to make sure that conflicting goals are *dealt with* in an appropriate way *during execution*. That is, an agent should take into account potential conflicts when trying to achieve goals. Finally, it will in most cases be useful to be able to represent *priorities* amongst goals, as an agent might have to choose between goals. An agent might, e.g., have to choose between having pizza or having sushi. It is then important that it chooses to pursue the goal that is most important to it or that it values most.

While we believe all of these aspects are important for a comprehensive treatment of conflicting goals, we focus in this paper on the investigation of ways of *representing and defining* that goals are conflicting. That is, we are solely concerned with the representation of goals that the agent would in principle like to pursue. We do not address how an agent then chooses the goals that it will actively pursue such that these are non-conflicting and higher-priority goals are preferred, nor do we address how these goals are pursued. This is left for future research.

In agent programming, there are a number of approaches that allow to represent and deal with conflicting motivational attitudes [9, 15, 28, 40, 49, 59]. One can, broadly speaking, distinguish three approaches for the representation of conflicting goals. In the first approach, the programmer *represents explicitly* that certain goals are conflicting. This approach is taken in the Jadex framework [40], in which an inhibition relation between goals is introduced that explicitly expresses that two goals are conflicting. In the second approach, the agent *reasons* about the goals and the *plans* that may be used for achieving these, in order to determine whether the plans interact in negative ways. This approach is taken in work by Thangarajah et al. [49]. The third approach is used in particular in case goals are represented using logic. In this approach, goals are considered to be conflicting, if they are *logically inconsistent*. This approach is taken in the language GOAL [15, 28], in the language Dribble [59] which is partly based on GOAL, and in the BOID framework [9].

The advantage of the first approach is that it involves less reasoning on the part of the agent, making the approach potentially more computationally efficient. The advantage of the second approach is that it does not put the burden on the programmer to determine whether the plans for achieving goals may interact in negative ways. An advantage of the third approach, i.e., of using logic for the representation of goals, is that this provides for added expressivity, e.g., allowing to represent conjunctive goals. This allows a more flexible mechanism for selecting plans, as the logical structure of goals can be exploited when selecting plans (for example, one may define that a plan for reaching a goal to have pizza can be used if the agent wants to

[5] This is in line with the so-called open-minded commitment strategy of [44].

have pizza and a drink).[6] Moreover, approaches based on logic are particularly amenable to a formal analysis and a comparison with BDI logics, which could potentially lead to a less ad hoc approach with a more solid foundation.

The approaches we propose in this paper are based on logic. Our first modeling of conflicting goals (based on propositional logic) takes goals to be conflicting iff they are logically inconsistent. The second approach (based on propositional default logic) *additionally* allows to define that goals are conflicting even though they are logically consistent. We think it is important to also allow the latter possibility, as not all conflicting goals are necessarily logically inconsistent, although, conversely, logical inconsistency does imply that goals are conflicting. An example of goals that are conflicting but not logically inconsistent, is the example mentioned above where an agent has the goal to have a pizza and to have sushi. Eventually, one might want to incorporate also the possibility to let the agent reason about plans in order to establish whether goals are conflicting, yielding a framework that combines the three approaches for representing conflicting goals. Investigations along these lines are, however, not carried out in this paper.

We view the proposal for the representation of conditional goals (Sect. 4) as the main contribution of this paper when it comes purely to modeling conflicting goals, as it is a more comprehensive approach than the approach based on unconditional goals (Sect. 3). Nevertheless, the proposal of Sect. 3 also has two important purposes. First, the proposal is relevant in its own right, as it provides a definition of unconditional goals that has some desirable properties. Second, we clarify the relation between our proposal of Sect. 4 and the semantics of unconditional goals as provided in GOAL, by showing that our semantics of unconditional goals is a special case of the semantics of conditional goals of Sect. 4, and showing how our semantics of unconditional goals is related to that of GOAL.

3 Unconditional goals

We are interested in a logic-based representation of goals. In the field of logic, one typically introduces some kind representation of the underlying system that one wants to model, together with a *logical language* in which *properties* of this representational structure can be expressed. The semantics of sentences in this language is defined on the structure.

In propositional logic, for example, the representational structure is formed by a valuation of propositional variables, and the language of propositional logic can be used for expressing properties of this structure using the standard logical connectives. In (modal) BDI logics such as [45], the structure is a kind of Kripke structure. The logical language combines the language of branching time temporal logic CTL* [18] with modal operators for beliefs, desires and intentions.

Given that we aim to investigate the semantic foundations of goals *in agent programming*, we use in this section a structure for representing goals which is based on a structure that has been used for the representation of goals in logic-based agent programming languages [28, 59]. To be more specific, the structure is a set of propositional formulas, which is called the goal base. We define a logical language that has a goal operator, in order to be able to express what the goals of the agent are, given a particular goal base. This logical language is defined below.

[6] Although this approach is usable in many cases, it does not always work. For example, in case there is no plan which reaches the goal of having a drink, or if the plan for having a drink "undoes" the goal of having pizza.

Definition 1 (*goal formulas*) Throughout this paper, we assume a language of propositional logic $\mathcal{L}$ with negation and conjunction, with typical element ϕ. We will use $\top \in \mathcal{L}$ to denote a tautology, $\bot \in \mathcal{L}$ to denote falsum and $\models$ will be used to denote the standard entailment relation for $\mathcal{L}$. The goal formulas $\mathcal{L}_{\mathsf{G}}$ with typical element κ are defined as follows, where $\phi \in \mathcal{L}$.

$$\kappa ::= \top \mid \mathbf{G}\phi \mid \neg\kappa \mid \kappa \wedge \kappa'$$

Note that the **G** operator cannot be nested, i.e., formulas of the form $\mathbf{GG}\phi$ are not part of the language. This is common in agent programming languages in which logical languages are used for specifying goals (see [26] for an exception). The reason is that it is difficult to establish what the meaning and practical use of such formulas would be in the context of agent programming languages in which the goal base is a set of propositional formulas.

3.1 Semantics

Given a goal base, the semantics of goal formulas can be defined in various ways. One way of defining the semantics would be to focus on the syntactic representation of the goal base, and define that $\mathbf{G}\phi$ holds, given a goal base $\gamma \subseteq \mathcal{L}$, iff $\phi \in \gamma$. Such syntactical approaches have been considered in so-called awareness logics for the representation of explicit beliefs [19]. While such syntactic approaches may be suitable in some contexts, in this paper we aim for semantic definitions that *exploit the logical structure of the goal base*.

3.1.1 Basic semantics

A semantics that exploits the logical structure of the goal base is the following, which we call the basic semantics. This semantics expresses that ϕ is a goal iff ϕ follows from the goal base (see also [55]).

Definition 2 ($basic$ ($\models_b$)) Let $\gamma \subseteq \mathcal{L}$ be the agent's goal base. Then the basic semantics for goal formulas $\models_b$ is defined as follows.

$$\begin{aligned}
&\gamma \models_b \top \\
&\gamma \models_b \mathbf{G}\phi \quad &&\Leftrightarrow \gamma \models \phi \\
&\gamma \models_b \neg\kappa \quad &&\Leftrightarrow \gamma \not\models_b \kappa \\
&\gamma \models_b \kappa \wedge \kappa' &&\Leftrightarrow \gamma \models_b \kappa \text{ and } \gamma \models_b \kappa'
\end{aligned}$$

There are several things to note about this semantics. First, we remark that this semantics does not take into account the beliefs of the agent. Often, this kind of semantics defines that the agent can have a goal for ϕ only if it does not believe ϕ to be reached [28, 59]. This would also prevent an agent from having tautologies as goals. We, however, omit this for reasons of simplicity.

A second observation about this semantics is that if, e.g., the formula $pizza \wedge sushi$ is in the goal base, the agent will be able to derive the goal $\mathbf{G}(pizza \wedge sushi)$, as well as the goals $\mathbf{G}(pizza)$ and $\mathbf{G}(sushi)$. This stems from the fact that we have defined the semantics such that all logical consequences of the goal base are goals. We feel that the fact that goals are closed under propositional logical consequence is in principle an intuitive property. Moreover, this property is generally attributed to motivational attitudes in BDI logics, which makes this semantics in line with these logics (see Sect. 3.2 for a more elaborate discussion). Finally, such a semantics facilitates a more flexible use of plans. For example, if the agent has the

goal *cleanRoom* $\wedge$ *washedDishes*, it will often be the case that it has two plans, i.e., one for cleaning rooms and one for washing dishes. A semantics in which the agent can derive the two goals *cleanRoom* and *washedDishes* from the one conjunctive goal, facilitates the selection of two separate plans. Naturally, this does not always have the desired result, e.g., if washing the dishes makes the room dirty again. However, the technique is often used in agent programming frameworks that have a logical representation of goals [14, 28, 59].

Further, note that this semantics, although it defines that logical consequences of goals are also goals, does not suffer from the well-known "dentist problem". The dentist problem is the issue that if an agent has the goal to go to the dentist and believes that going to the dentist implies feeling pain, the agent should intuitively not derive the goal to feel pain. Modal BDI logics typically suffer from this problem, i.e., the goal of feeling pain is derived in these logics. In our case, it is not a problem, as the implication *dentist* $\rightarrow$ *pain* will typically be part of the *belief base*, and not of the goal base. That is, the agent does not have the goal that if it goes to the dentist, it will feel pain, and will therefore not be able to derive the goal to feel pain from the goal to go to the dentist.

A third characteristic of this semantics is that the formulas $\mathbf{G}\neg\phi$ and $\neg\mathbf{G}\phi$ are not equivalent. Intuitively, the first formula expresses the *presence* of a goal $\neg\phi$, while the second formula expresses the *absence* of a goal ϕ. Formally, it could be the case that ϕ is not derivable from the goal base, in which case $\neg\mathbf{G}\phi$ would hold, while $\neg\phi$ is not derivable, in which case $\mathbf{G}\neg\phi$ would *not* hold. That is, it does not hold in general that $\neg\mathbf{G}\phi$ implies $\mathbf{G}\neg\phi$. Note that this observation also explains that this semantics does not adopt the "closed world assumption". The closed world assumption is the presumption that what cannot be derived should be interpreted as being false. Under a closed world assumption, we therefore *would* have that $\neg\mathbf{G}\phi$ implies $\mathbf{G}\neg\phi$. If the goal base is consistent, then the converse of this implication does hold, as expressed in the following proposition.

Proposition 1 *Let* $\gamma \subseteq \mathcal{L}$ *where* $\gamma \not\models \bot$. *Then the following holds.*

$$\gamma \models_b \mathbf{G}\neg\phi \Rightarrow \gamma \models_b \neg\mathbf{G}\phi$$

Proof Assume $\gamma \models_b \mathbf{G}\neg\phi$. This means that $\gamma \models \neg\phi$ (Definition 2). Given that γ is consistent, we have that $\gamma \not\models \phi$. This means that $\gamma \models_b \neg\mathbf{G}\phi$, yielding the desired result. □

Another issue worth discussing regarding Definition 2, is the structure of the goal base. One might consider to let the goal base be a set of formulas from $\mathcal{L}_{\mathsf{G}}$, rather than from $\mathcal{L}$. Such a goal base would represent explicitly which goals the agent has. Nevertheless, one would still need a definition of what the goals of the agent are, given such a goal base.

Providing such a definition for a goal base $\gamma \subseteq \mathcal{L}_{\mathsf{G}}$ in a similar way as was done for the basic semantics of Definition 2, however, leads to problems. This has to do with the fact that if the semantics of negation is defined for γ in the way as was done in that definition, this would be a definition of closed world assumption. The distinction with the semantics of Definition 2 which does *not* define a closed world assumption, is somewhat subtle.

Given a goal base $\{\mathbf{G}\phi \vee \mathbf{G}\psi\}$, one would be able to derive $\neg\mathbf{G}\phi$ and $\neg\mathbf{G}\psi$, which is inconsistent with this goal base. This is the same issue that arises when closed world assumption is adopted in case of, e.g., a knowledge base $\{p \vee q\}$. This unwanted behavior does not occur in the case of Definition 2.

An alternative to defining the semantics for a goal base $\gamma \subseteq \mathcal{L}_{\mathsf{G}}$ in the way just discussed, would be to regard the formulas in the goal base as modal formulas, and to use a modal logic consequence relation for deriving goals on the basis of the goal base. Such a definition would, however, imply the need of using modal logic theorem provers or similar tools if it

is going to be used in an agent programming language. Moreover, the underlying Kripke semantics of modal logic is (arguably) relatively complex. The use of modal formulas for representing goals would thus not necessarily make the language more practically usable. Moreover, (normal) modal logics are not particularly suitable for dealing with conflicting goals, as will be explained in Sect. 3.2. We thus argue that the way we define the goal base, i.e., as a set of propositional formulas, nicely circumvents these problems. Moreover, such a structure is closely related to existing approaches to the modeling of goals in logic-based agent programming languages.

Our final observation regarding the basic semantics, is one that will lead us up to our next definition of semantics of goals. This observation concerns the case where γ is inconsistent, e.g., if there is a formula $\phi \in \gamma$ and a formula $\neg\phi \in \gamma$. The definition of the basic semantics for goals is such that any (propositional) logical consequence of the goal base is a goal. If the goal base is inconsistent, this then means that the agent has the goal falsum, i.e., we have $\mathbf{G}\bot$. Worse still, anything becomes a goal, i.e., any formula $\mathbf{G}\phi$ holds on such a goal base.

If the goal base is inconsistent, the semantics is thus such that the logic is *trivialized*. This essentially means that the agent cannot deal with inconsistent goals. We, however, want our agent to be able to deal with inconsistent goals, which is why we are interested in alternative definitions.

As an aside, we remark that the issue of trivialization of the logic in case of inconsistency is also discussed in the context of *paraconsistent* logics [2]. In that work, it is argued that being able to reason with inconsistent information without trivializing the logic is central to practical reasoning. In [20], Gabbay and Hunter also argue that inconsistency should be viewed as a "good" thing, rather than as a "bad" thing.

3.1.2 Semantics of Hindriks et al.

Hindriks et al. have proposed a semantics of goals that allows the goal base to be inconsistent, without trivializing the logic [15, 28]. To be more accurate, their semantics does not trivialize the logic if individual formulas in the goal base are consistent. Their semantics of goals defines that ϕ is a goal, iff there is a formula in the goal base from which ϕ follows.

Definition 3 (*Hindriks et al.* ($\models_h$))

$$\gamma \models_h \mathbf{G}\phi \Leftrightarrow \exists \phi' \in \gamma : \phi' \models \phi$$

The semantics of $\top$, negation, and conjunction are as in Definition 2, but we leave them out here and in definitions in the sequel for reasons of presentation.

Intuitively, this semantics does not trivialize the logic in case of an inconsistent goal base (assuming that individual formulas in the goal base *are* consistent), as it does not combine multiple (possibly inconsistent) formulas from the goal base to derive a goal. Formally, if each formula in the goal base is consistent, the goal $\mathbf{G}\bot$ cannot be derived.

While this is a desired property, this semantics does not fully exploit the logical structure of the goal base. In particular, goals are not closed under propositional logical consequence under this semantics. We have argued in the context of the basic semantics for goals that we consider this to be in principle a desired property. As we have seen, however, this leads to problems if the goal base is inconsistent.

3.1.3 Consistent subset semantics

Our next semantics now aims to combine the features of the semantics of Hindriks et al. when it comes to handling an inconsistent goal base, with the desired characteristic that goals are closed under logical consequence. The intuitive idea is that formulas in the goal base should be combined to derive a goal on their basis, if these formulas *can* be combined, i.e., if these formulas are consistent. For example, if we have a goal base $\{p, q\}$, we would like to be able to derive the goal $\mathbf{G}(p \wedge q)$, which would not be possible in the semantics of Hindriks et al. Moreover, if we have a goal base $\{p, q, \neg q\}$, we would *also* like to derive this goal, but we do *not* want to derive the goal $\mathbf{G}\bot$.

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. A nice feature of this semantics is that it is equivalent with the basic semantics if the goal base is consistent (see Proposition 3).

Definition 4 (*consistent subset* ($\models_s$))

$$\gamma \models_s \mathbf{G}\phi \Leftrightarrow \exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot \text{ and } \gamma' \models \phi)$$

Note that, in contrast with the semantics of Hindriks et al., the consistent subset semantics does not trivialize the logic, even if there are inconsistent formulas in the goal base. Inconsistent formulas in the goal base are "ignored" by this definition, because we only consider subsets γ' of γ which are consistent ($\gamma' \not\models \bot$).[7] An inconsistent formula such as $p \wedge \neg p$, for example, cannot be used to derive $\mathbf{G}\bot$, or any other goal for that matter.

This semantics can be viewed as related to a proposal by Poole [41] in the area of non-monotonic reasoning. He proposes to reason on the basis of a theory consisting of a set of facts, and a set of hypotheses (both being sets of first order formulas). A formula is then explainable on the basis of this theory, if it follows from the set of facts, and a consistent subset of the hypotheses.

3.1.4 Alternative structures for the goal base

One might consider to "wrap" (mutually inconsistent) formulas in the goal base such that a consistent set results, in order to get rid of the issues that come with an inconsistent goal base. One could, e.g., consider to have a goal base $\{\mathbf{G}(p), \mathbf{G}(\neg p)\}$, instead of a goal base $\{p, \neg p\}$. However, this also leads to problems as was already discussed above. In particular, if a normal modal logic consequence relation is used to derive goals, it will be the case that $\mathbf{G}\bot$ can be derived on the basis of these formulas. Wrapping the propositional formulas in this way thus does not immediately solve the question of how to deal with inconsistencies.

Alternatively, one could consider the use of temporal logic in a similar manner, resulting in a goal base $\{\Diamond p, \Diamond(\neg p)\}$ which expresses that the agent has the goal to reach p eventually, and to reach $\neg p$ eventually. These formulas are not inconsistent, and would normally not lead to the derivation of $\Diamond\bot$. However, "hiding" the inconsistency in this way is not necessarily beneficial when goals are incorporated into an agent programming framework. The representation of conflicting goals should facilitate the agent to refrain from pursuing conflicting goals simultaneously. In order to do this appropriately, it is important to have a notion of

[7] A similar behavior could be obtained in Definition 3 if the condition $\phi' \not\models \bot$ would be added to the righthand side of the definition, yielding a close resemblance with Definition 4. Definition 3 is however the one provided by Hindriks et al. [28].

when goals are conflicting. Presumably, this can be defined more clearly if inconsistencies are not hidden. Moreover, using temporal logic would most likely also require to perform temporal logic reasoning.

3.2 Properties

In this section, we investigate properties of the semantics of Sect. 3.1, and compare these semantics to one another.

3.2.1 Properties of the Semantics

The kind of properties we are interested in, are axioms of modal logics. That is, we want to compare the goal operator of our logic with modal operators. The particular axioms that we consider are listed below (see [10, 36] for more details on modal logics).

$$\begin{aligned}
\mathbf{K} &: \mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi)\\
\mathbf{D_1} &: \neg\mathbf{G}\bot\\
\mathbf{D_2} &: \neg(\mathbf{G}\phi \wedge \mathbf{G}\neg\phi)\\
\mathbf{M} &: \mathbf{G}(\phi \wedge \psi) \rightarrow (\mathbf{G}\phi \wedge \mathbf{G}\psi)\\
\mathbf{C} &: (\mathbf{G}\phi \wedge \mathbf{G}\psi) \rightarrow \mathbf{G}(\phi \wedge \psi)
\end{aligned}$$

In modal logics, ϕ and ψ are modal logic formulas, possibly containing modal operators. As we do not have a nesting of goal operators in this paper, we consider ϕ and ψ to be propositional here.

The **K** axiom is the basic axiom that all normal modal logics adhere to, and expresses that goals are closed under propositional logical consequence. That is, if $\mathbf{G}\phi$ holds, than $\mathbf{G}\psi$ holds if ψ follows from ϕ under the standard consequence relation of proposition logic. The **D** axiom expresses that goals are consistent. It comes in two forms ($\mathbf{D_1}$ and $\mathbf{D_2}$), which are equivalent if the **K** axiom is adopted. Axiom **M** expresses that conjunctive goals can be "taken apart" to yield goals for the separate conjuncts. Axiom **C** expresses the opposite, i.e., it says that separate goals can be combined into one. Axioms **M** and **C** together are equivalent with the **K** axiom. The axioms **M** and **C** are used to axiomatize so-called non-normal modal logics, which are logics that do not adopt the **K** axiom but a weaker axiom such as **M** or **C** instead [10].

The following proposition expresses which of these axioms of modal logics hold for our various semantics for goals. We distinguish three cases, i.e., the case in which the goal base is an arbitrary one, the case in which each formula in the goal base is consistent, and the case in which the entire goal base is consistent. Note that these conditions on the goal base get increasingly restrictive. If a semantics satisfies an axiom in a particular case, we can immediately conclude that the axiom is satisfied in a more restrictive case. This is not explicitly incorporated in the proposition. We use, e.g., the notation $\gamma \models_b \mathbf{K}, \mathbf{M}$ to abbreviate $\gamma \models_b \mathbf{K}$ and $\gamma \models_b \mathbf{M}$.

Proposition 2 *Let γ be an arbitrary goal base.*

$$\gamma \models_b \mathbf{K}, \mathbf{M}, \mathbf{C} \qquad \gamma \models_h \mathbf{M} \qquad \gamma \models_s \mathbf{D_1}, \mathbf{M} \tag{3.1}$$

Let $\forall\phi \in \gamma : \phi \not\models \bot$.

$$\gamma \models_h \mathbf{D_1} \tag{3.2}$$

Let $\gamma \not\models \bot$.

$$\gamma \models_b \mathbf{D_1}, \mathbf{D_2} \qquad \gamma \models_h \mathbf{D_2} \qquad \gamma \models_s \mathbf{K}, \mathbf{D_2}, \mathbf{C} \tag{3.3}$$

Proof (3.1) We show that $\gamma \models_b \mathbf{K}$. We have to show that $\gamma \models_b \mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi)$. This means we have to show that $\gamma \models_b \mathbf{G}(\phi \rightarrow \psi) \Rightarrow (\gamma \models_b \mathbf{G}\phi \Rightarrow \gamma \models_b \mathbf{G}\psi)$. Assume that $\gamma \models_b \mathbf{G}(\phi \rightarrow \psi)$ and $\gamma \models_b \mathbf{G}\phi$. This means that $\gamma \models \phi \rightarrow \psi$ and $\gamma \models \phi$. From this we can conclude that $\gamma \models \psi$, which is the definition of $\gamma \models_b \mathbf{G}\psi$, yielding the desired result.

We show that $\gamma \models_b \mathbf{M}$. We have to show that $\gamma \models_b \mathbf{G}(\phi \wedge \psi) \Rightarrow \gamma \models_b \mathbf{G}\phi \wedge \mathbf{G}\psi$. This means we have to show that $\gamma \models_b \mathbf{G}(\phi \wedge \psi) \Rightarrow (\gamma \models_b \mathbf{G}\phi$ and $\gamma \models_b \mathbf{G}\psi)$, which is defined as $\gamma \models \phi \wedge \psi \Rightarrow (\gamma \models \phi$ and $\gamma \models \psi)$. The latter is obviously the case.

We have to show that $\gamma \models_b \mathbf{C}$. The proof is analogous to the proof of $\gamma \models_b \mathbf{M}$.

We show that $\gamma \models_h \mathbf{M}$. We have to show that $\gamma \models_h \mathbf{G}(\phi \wedge \psi) \Rightarrow (\gamma \models_h \mathbf{G}\phi$ and $\gamma \models_h \mathbf{G}\psi)$. Assume that $\gamma \models_h \mathbf{G}(\phi \wedge \psi)$, which is defined as: $\exists \phi' \in \gamma : \phi' \models \phi \wedge \psi$. From the latter we can conclude that $\exists \phi' \in \gamma : \phi' \models \phi$ and $\exists \phi' \in \gamma : \phi' \models \psi$, which is the definition of $\gamma \models_h \mathbf{G}\phi$ and $\gamma \models_h \mathbf{G}\psi$.

We show that $\gamma \models_s \mathbf{D_1}$. We have to show that $\gamma \not\models_s \mathbf{G}\bot$, i.e., that $\neg \exists \gamma' \subseteq \gamma : \gamma' \not\models \bot$ and $\gamma' \models \bot$. This is obviously the case.

The proof for $\gamma \models_s \mathbf{M}$ is analogous to the proof for $\gamma \models_h \mathbf{M}$.

(3.2) We show that $\gamma \models_h \mathbf{D_1}$. We have to show that $\gamma \not\models_h \mathbf{G}\bot$, i.e., that $\neg \exists \phi \in \gamma : \phi \models \bot$. Since each $\phi \in \gamma$ is consistent by assumption, this follows immediately.

(3.3) We show that $\gamma \models_b \mathbf{D_1}$. We have to show that $\gamma \models_b \neg \mathbf{G}\bot$, i.e., that $\gamma \not\models_b \mathbf{G}\bot$, i.e., that $\gamma \not\models \bot$. This follows immediately, since γ is assumed to be consistent.

We show that $\gamma \models_b \mathbf{D_2}$. We have to show that $\gamma \models_b \neg(\mathbf{G}\phi \wedge \mathbf{G}\neg\phi)$, i.e., that $\gamma \not\models_b \mathbf{G}\phi \wedge \mathbf{G}\neg\phi$, i.e., that it is not the case that $\gamma \models_b \mathbf{G}\phi$ and $\gamma \models_b \mathbf{G}\neg\phi$, i.e., that it is not the case that $\gamma \models \phi$ and $\gamma \models \neg\phi$. This is the case, since γ is assumed to be consistent.

We show that $\gamma \models_h \mathbf{D_2}$. We have to show that $\gamma \models_h \neg(\mathbf{G}\phi \wedge \mathbf{G}\neg\phi)$, i.e., that $\gamma \not\models_h \mathbf{G}\phi \wedge \mathbf{G}\neg\phi$, i.e., that it is not the case that $\gamma \models_h \mathbf{G}\phi$ and $\gamma \models_h \mathbf{G}\neg\phi$, i.e., that it is not the case that $\exists \phi' \in \gamma : \phi' \models \phi$ and $\exists \phi' \in \gamma : \phi' \models \neg\phi$. This is the case, since γ is assumed to be consistent.

We show that $\gamma \models_s \mathbf{K}$. We have to show that $\gamma \models_s \mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi)$. This means we have to show that $\gamma \models_s \mathbf{G}(\phi \rightarrow \psi) \Rightarrow (\gamma \models_s \mathbf{G}\phi \Rightarrow \gamma \models_s \mathbf{G}\psi)$. Assume that $\gamma \models_s \mathbf{G}(\phi \rightarrow \psi)$ and $\gamma \models_s \mathbf{G}\phi$. This means that $\exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \phi \rightarrow \psi)$ and $\exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \phi)$. As γ is assumed to be consistent, we can conclude that $\exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \phi \rightarrow \psi$ and $\gamma' \models \phi)$. From this we can conclude that $\exists \gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \psi)$, which is the definition of $\gamma \models_s \mathbf{G}\psi$, yielding the desired result.

We have to show that $\gamma \models_s \mathbf{M}$ and $\gamma \models_s \mathbf{C}$. The proofs are analogous to the proof of $\gamma \models_s \mathbf{K}$. □

There are several important things to note about this proposition. First, we can see that the basic semantics does not satisfy $\mathbf{D_1}$ in the general case. This means that $\mathbf{G}\bot$ is satisfiable, namely, in case $\gamma \models \bot$. We consider this to be undesirable for a semantics of goals, as we want to allow the goal base to be inconsistent without the logic being trivialized. The consistent subset semantics and the semantics of Hindriks et al. *do* satisfy the axiom $\mathbf{D_1}$ (although in case of the semantics of Hindriks et al., we need the additional assumption that individual formulas in the goal base are consistent).

Second, we can see that in the general case, the semantics of Hindriks et al. and the consistent subset semantics do not satisfy the **K** axiom, but only the weaker **M** axiom, while the basic semantics does satisfy **K**. We consider the satisfaction of the **K** axiom in principle desirable. However, it results in the logic being trivialized, i.e., in the satisfiability of $\mathbf{G}\bot$, in case the goal base is inconsistent. We view the satisfaction of $\mathbf{D_1}$ as more important than the satisfaction of **K** for modeling goals, as a failure to satisfy $\mathbf{D_1}$ means that the semantics cannot handle inconsistent goals. This is the main reason that we consider the semantics of Hindriks et al. and the consistent subset semantics to be more suitable for modeling conflicting goals than the basic semantics.

Third, the proposition shows that if the goal base is assumed to be consistent, the consistent subset semantics *does* satisfy **K**, in contrast with the semantics of Hindriks et al. Given that we see the satisfaction of **K** as desirable, this is one reason that we consider the consistent subset semantics to have better properties than the semantics of Hindriks et al. Recall that a reason for our point of view that the satisfaction of **K** is desirable, is the fact that this property is satisfied by motivational attitudes in BDI logics [11, 44, 45]. In fact, the consistent subset semantics satisfies both **K** *and* $\mathbf{D_1}$ and $\mathbf{D_2}$ if the goal base is consistent, which corresponds exactly with these logics. That is, in these logics motivational attitudes are assumed to be consistent, and satisfy the **K** axiom and the **D** axiom.

Fourth, we can conclude that under the semantics of Hindriks et al. and the consistent subset semantics, the axioms $\mathbf{D_1}$ and $\mathbf{D_2}$ are not equivalent. The axiom $\mathbf{D_1}$ is satisfied in the general case (that is, with one additional constraint in the case of the semantics of Hindriks et al.), while $\mathbf{D_2}$ is not. As mentioned, in normal modal logics the axioms $\mathbf{D_1}$ and $\mathbf{D_2}$ are equivalent. This is not surprising, as in normal modal logics we have the **K** axiom, which means we have the **C** axiom. This axiom tells us in particular that if $\mathbf{G}\phi$ and $\mathbf{G}\neg\phi$ hold, we can derive $\mathbf{G}(\phi \wedge \neg\phi)$, i.e., we can derive $\mathbf{G}\bot$. Given that the **C** axiom does not hold for the semantics of Hindriks et al. and the consistent subset semantics in the general case, it is perhaps not surprising that the two axioms are not equivalent. In fact, these two semantics were particularly designed to allow the satisfiability of $\mathbf{G}\phi \wedge \mathbf{G}\neg\phi$, without making $\mathbf{G}\bot$ satisfiable.

Our final observation regarding this proposition is that in case we assume that each formula in the goal base is consistent, the semantics of Hindriks et al. and the consistent subset semantics satisfy the same axioms (with respect to the axioms **K**, $\mathbf{D_1}$, $\mathbf{D_2}$, **C**, and **M** considered here). However, as we will see next, these semantics are *not* equivalent.

3.2.2 Relations between the semantics

The investigation of which axioms are satisfied by our semantics already tells us a little about how they are related. In this section, we investigate the relations between the semantics more directly.

Proposition 3 *Let* $\forall\phi \in \gamma : \phi \not\models \bot$.

$$\gamma \models_h \mathbf{G}\phi \Rightarrow \gamma \models_s \mathbf{G}\phi \tag{3.4}$$

Let $\gamma \not\models \bot$.

$$\gamma \models_b \mathbf{G}\phi \Leftrightarrow \gamma \models_s \mathbf{G}\phi \tag{3.5}$$

Proof (3.4) Assume $\gamma \models_h \mathbf{G}\phi$, i.e., $\exists\phi' \in \gamma : \phi' \models \phi$. We have that $\phi' \not\models \bot$ by assumption. Then we also have that $\exists\gamma' \subseteq \gamma : (\gamma' \not\models \bot$ and $\gamma' \models \phi)$, as we know that there is a $\phi' \in \gamma$

such that $\phi' \models \phi$, and we can take $\gamma' = \{\phi'\}$. This yields the desired result. (3.5) 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 $\gamma \models \phi$.

Comparing the consistent subset semantics with the semantics of Hindriks et al., we see that the set of goals derivable under the semantics of Hindriks et al. is a subset of those derivable under the consistent subset semantics (3.4) (under the assumption that the formulas in the goal base are consistent).[8] The opposite of (3.4) does not hold in general. Take, e.g., a goal base $\{p, q\}$. In that case, $\mathbf{G}(p \wedge q)$ holds under the consistent subset semantics, but does not hold under the semantics of Hindriks et al.

We thus have that the consistent subset semantics allows the derivation of strictly more goals than the semantics of Hindriks et al. if the formulas in the goal base are consistent. However, the properties considered in Proposition 2 do *not* distinguish the two. That is, under the condition that the formulas in the goal base are consistent both semantics satisfy axioms $\mathbf{D_1}$ and $\mathbf{M}$, although we have just argued that the semantics are not equivalent.

Finding a discriminating property is not a trivial task. It seems that a weaker version of the axiom $\mathbf{C}$ might be what we are looking for to characterize the consistent subset semantics, since in this semantics two goals may sometimes be combined into one, but not always. For example, given a goal base $\{p, q\}$, we have that $\mathbf{G}p$ and $\mathbf{G}q$ hold under the consistent subset semantics, and $\mathbf{G}(p \wedge q)$ also holds (but $\mathbf{G}(p \wedge q)$ does not hold under the semantics of Hindriks et al.). Given a goal base $\{p \wedge r, q \wedge \neg r\}$, we have that $\mathbf{G}p$ and $\mathbf{G}q$ hold, but $\mathbf{G}(p \wedge q)$ does *not* hold under the consistent subset semantics, i.e., in this case, p and q may not be combined.

Since p and q may be combined in some but not all cases, it is not very likely that we can use a version of axiom $\mathbf{C}$ in which we put conditions on ϕ and ψ only. We might end up concluding that the strongest property we can come up with, is that axiom $\mathbf{C}$ holds iff there is a consistent subset of γ from which both ϕ and ψ follow. This property is however not very informative, since it essentially repeats the semantic definition. Also, it is a property which is not very general, since it does not depend on general properties of γ or of ϕ and ψ. Further investigations along these lines are left for future research.

A nice property that is worth noting regarding the consistent subset semantics, is expressed in Corollary 1 below. That is, Property (3.6) states that under the assumption of consistency of the goal base, the basic semantics and the consistent subset semantics are equivalent. This is in line with Proposition 2, as in this case the semantics satisfy the same axioms. We thus have that consistent subset semantics fully exploits the logical structure of the goal base in case the goal base is consistent, while this is not the case for the semantics of Hindriks et al. This is expressed by Property (3.7), which says that the set of goals derivable under the semantics of Hindriks et al. is a subset of those derivable under the basic semantics, in case the goal base is consistent. The implication does not hold in the other direction.

Corollary 1 *Let* $\gamma \not\models \bot$.

$$\gamma \models_b \kappa \Leftrightarrow \gamma \models_s \kappa \tag{3.6}$$

$$\gamma \models_h \mathbf{G}\phi \Rightarrow \gamma \models_b \mathbf{G}\phi \tag{3.7}$$

Proof (3.6) Immediate from (3.5) and Definitions 2 and 4. (3.7) Immediate from (3.4) and (3.5). □

[8] Note that the property does not hold for goal bases in general, as it can then be the case that $\mathbf{G}\bot$ holds in case of the semantics of Hindriks et al., while this never holds for the consistent subset semantics. This occurs if there is an inconsistent formula in the goal base.

4 Conditional goals

In Sect. 3, we presented a number of semantics for goals which were based on a goal base consisting of a set of propositional formulas. In this section, we propose another structure for the representation of goals with an accompanying logical language and semantics. This new structure has two main advantages over the goal base of Sect. 3. First, it allows to represent that goals are *conflicting*, even though they are *logically consistent*. Second, it allows the representation of *conditional* goals, that is, goals that are conditional on beliefs and/or other goals.

Being able to represent conditional goals seems intuitively desirable. One is then able to represent that, e.g., an agent wants to have a pizza if it is hungry. This intuition that the representation of conditional goals is important, is also backed by research in philosophical logic [24] in which mental attitudes are argued to be conditional by nature, and it forms the basis for the BOID architecture (see Sect. 5.2 for more details on that work).

The logical language of goals that we use in this section, is the same as the one that we have used in Sect. 3 (Definition 1). The semantics of goal formulas and the structure on which the formulas are evaluated, however, differs from the previous section. To be more specific, the construct that we propose for representing conditional goals are *goal inference rules*. A goal inference rule has the form $\beta, \kappa^+, \kappa^- \Rightarrow \phi$, where ϕ represents the goal that can be inferred using the rule, β is a condition on the agent's beliefs, κ^+ represents the goals on the basis of which ϕ can be inferred, and κ^- represents the goals that are conflicting with ϕ. That is, β and κ^+ allow the representation of conditional goals, and κ^- allows to represent that goals are conflicting. The set of goal inference rules is formally defined below.

Definition 5 (*goal inference rule*) The set of goal inference rules $\mathcal{R}_{\mathsf{GI}}$ is defined as follows: $\{\beta, \kappa^+, \kappa^- \Rightarrow \phi \mid \beta \subseteq \mathcal{L}, \kappa^+ \subseteq \mathcal{L}, \kappa^- \subseteq \mathcal{L}, \phi \in \mathcal{L}\}$.

A goal inference rule $\{source\}^{\beta}, \{target\}^{\kappa^+} \Rightarrow waypoint$,[9] for example, intuitively expresses that if the agent is at some source location and has the goal to be at a target location, it may infer the goal to be at a waypoint in between the source and the target. This rule shows how goals can be conditional on beliefs and other goals. An example of a goal inference rule that represents conflicting goals is $\{hungry\}^{\beta}, \{sushi\}^{\kappa^-} \Rightarrow pizza$, which intuitively expresses that if the agent is hungry, it may derive the goal to have pizza, but the goal to have sushi is in conflict with the goal to have pizza.

The semantics of goals based on goal inference rules is defined through a translation of goal inference rules into so-called default rules of default logic. In Sect. 4.1, we introduce default logic. In Sect. 4.2, we present the semantics of goals, and in Sect. 4.3 we investigate properties of this semantics, and show how it is related to the consistent subset semantics of goals of Definition 4.

4.1 Default Logic

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, 7]. In this paper, we consider a purely propositional variant of default logic without variables.

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

[9] For reasons of presentation, we use superscripts to indicate whether a set of propositional formulas is a condition on beliefs, on goals, or represents conflicting goals, and if a set is empty, it is left out from the rule.

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 $\mathcal{L}$, a *default rule* has the form $\phi\colon \psi_1, \ldots, \psi_n/\chi$, 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* [7] 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 entailment relation for $\mathcal{L}$ and $Th(E_i)$ is the closure under classical logical consequence of E_i.

$$\begin{aligned} E_0 &= W \\ E_{i+1} &= Th(E_i) \cup \{\chi \mid \phi\colon \psi_1, \ldots, \psi_n/\chi \in D,\ E_i \models \phi,\ 1 \leq j \leq n,\ \forall j\colon E \not\models \neg\psi_j\} \end{aligned}$$

A set $E \subseteq \mathcal{L}$ is then an extension of $\langle W, D\rangle$ iff $E = \bigcup_{i=0}^{\infty} E_i$. In the sequel, we will sometimes be somewhat imprecise and say that, e.g., $\{p\}$ is an extension, where we should, strictly speaking, say that $Th(\{p\})$ is an extension.

It is important to note that extensions are always *consistent* sets[10] that are *closed* under the application of default rules. A rule $\phi\colon \psi_1, \ldots, \psi_n/\chi$ 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 for all rules $\phi\colon \psi_1, \ldots, \psi_n/\chi \in D$, that if the rule is applicable to E, then $E \models \chi$. Moreover, an extension may not entail the negation of a justification of a default rule applicable to it (specified by the last requirement of the definition above).

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. An extension may not entail the negation of a justification of any default rule applicable to it. 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.

4.2 Semantics

Given a set of goal inference rules, we want to define the semantics of goal formulas. We have found that an appropriate translation of goal inference rules into default rules can be used for defining a semantics that has intuitive characteristics. We believe it to be an advantage that goal inference rules are translated into default logic rather than defining the semantics "directly", as default logic is well investigated. As remarked in Sect. 2.1, results regarding

[10] Assuming W is consistent.

the relation between default logic and answer set programming [21, 22], can be used as a basis for implementing the ideas presented in this paper in an agent programming language using existing answer set solvers [33, 37]. Moreover, work on prioritized default logic (see, e.g., [6, 8, 16, 34]) can be used as a basis for the representation of priorities among goals.

The semantics is based on the idea of default logic that each extension represents a possible view on the world, while extensions are mutually conflicting, i.e., a default theory represents that the world cannot be in a state that corresponds to multiple extensions of the theory. In the context of goals, we similarly take each extension to represent a non-conflicting set of goals.[11] We define the semantics of goals based on the (extensions of the) default theory that result(s) from translating the goal inference rules into a set of default rules.

4.2.1 Translating goal inference rules into default rules

The general idea of the translation is that κ^+ is translated into the prerequisite of the default rule, κ^- is translated into the justification of the default rule, and the consequent of the goal inference rule is translated into the consequent of the default rule. We will provide the explanation of why such a translation has desired characteristics after giving the formal definition of the translation. The belief condition β is dealt with separately, which will also be explained in the sequel. We define a function f that takes a set of goal inference rules without belief condition and yields a set of propositional default rules.

Definition 6 (*goal inference rules to default rules*) Let DR denote the set of propositional default rules. The function $t\colon \mathcal{R}_{\mathsf{GI}} \rightarrow \wp(\mathsf{DR})$, taking a goal inference rule and yielding a default rule, is then defined as follows, where $\kappa^+, \kappa^- \Rightarrow \chi$ is a goal inference rule without belief condition, $\kappa^+ = \{\phi_1, \ldots, \phi_m\}$ and $\kappa^- = \{\psi_1, \ldots, \psi_n\}$, and $m, n \geq 1$. If $\kappa^+ = \emptyset$, κ^+ is translated to $\top$, and if $\kappa^- = \emptyset$, the sequence $\neg\psi_1, \ldots, \neg\psi_n$ is empty.

$$t(\kappa^+, \kappa^- \Rightarrow \chi) = \{\phi_1 \wedge \ldots \wedge \phi_m : \neg\psi_1, \ldots, \neg\psi_n, \chi/\chi\}$$

The function $f\colon \wp(\mathcal{R}_{\mathsf{GI}}) \rightarrow \wp(\mathsf{DR})$ taking a set of goal inference rules of the form $\kappa^+, \kappa^- \Rightarrow \chi$ and yielding a set of default rules, is defined as follows.

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

We explain this definition using an (abstract) example. Consider the goal inference rules

$$\begin{array}{ll} (g_1) & \{p\}^{\kappa^+}, \{q\}^{\kappa^-} \Rightarrow r \\ (g_2) & \Rightarrow p \\ (g_3) & \{r\}^{\kappa^+} \Rightarrow q \end{array}$$

which, respectively, correspond with the default rules

$$\begin{array}{ll} (d_1) & p : \neg q, r \;/\; r \\ (d_2) & \top : p \;/\; p \\ (d_3) & r : q \;/\; q \end{array}$$

When transforming a goal inference rule $\kappa^+, \kappa^- \Rightarrow \chi$, the condition κ^+ is mapped onto the prerequisite of a default rule, and the formulas in κ^- are *negated* and mapped onto the

[11] A difference with default logic for representing knowledge is that different extensions in the context of goals intuitively represent conflicting sets of goals, but all goals represented by the various extensions can exist simultaneously.

justification of the default rule. Considering goal inference rules g_1 and g_2, the set $\{p, r\}$ is an extension of their corresponding default rules d_1 and d_2. This reflects our intuition about goal inference rules: p can be derived on the basis rule g_2, and if p is a goal we can derive goal r, but not in combination with the conflicting goal q (rule g_1). 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 to our intuition about goal inference rules: given rule g_1, r can only be a goal if q is not, since q conflicts with r. The goals r and q thus cannot be part of the same extension.

The conflicting goals in κ^- are mapped to a sequence of justifications, rather than to one conjunctive justification. The reason is, that we want to allow goal inference rules such as $\{p, \neg p\}^{\kappa^-} \Rightarrow q$, specifying that goal q can be derived but that both p and $\neg p$ are conflicting with q. 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 inference 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 inference rules for which $\kappa^- = \emptyset$ 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 [7].

4.2.2 Semantic definition

We define the semantics of goals on the basis of a structure consisting of a belief base and a set of goal inference rules (the rule base). In order to define the semantics, we transform only those goal inference rules into default rules of which *the belief condition holds*, given the belief base. The goal inference rules can be transformed into default rules by means of the function f of Definition 6, after removing the (true) belief condition. Given an extension of the generated default rules, we define that $\mathbf{G}\phi$ holds iff ϕ follows from one of the extensions of the resulting default theory.

Definition 7 (*semantics of goals*) Let $\mathsf{GI} \subseteq \mathcal{R}_{\mathsf{GI}}$ be a finite set of goal inference rules, and let σ be a belief base. Let GI_σ be defined as follows, where $\models_{\mathcal{L}}$ is the standard entailment relation of propositional logic, lifted to *sets* of sentences on the right-hand side of $\models_{\mathcal{L}}$.

$$\mathsf{GI}_\sigma = \{\kappa^+, \kappa^- \Rightarrow \phi \mid \exists(\beta, \kappa^+, \kappa^- \Rightarrow \phi) \in \mathsf{GI} : \sigma \models_{\mathcal{L}} \beta\}.$$

The default semantics $\models_d$ for goal formulas is then defined as follows on the basis of belief base σ and set of goal inference rules GI.

$$\begin{aligned}
\langle \sigma, \mathsf{GI}\rangle \models_d \mathbf{G}\phi \quad &\Leftrightarrow \exists E : E \text{ is an extension of } \langle \emptyset, f(\mathsf{GI}_\sigma)\rangle \text{ and } E \models \phi \\
\langle \sigma, \mathsf{GI}\rangle \models_d \neg\kappa \quad &\Leftrightarrow \langle \sigma, \mathsf{GI}\rangle \not\models_d \kappa \\
\langle \sigma, \mathsf{GI}\rangle \models_d \kappa \wedge \kappa' \;&\Leftrightarrow \langle \sigma, \mathsf{GI}\rangle \models_d \kappa \text{ and } \langle \sigma, \mathsf{GI}\rangle \models_d \kappa'
\end{aligned}$$

Note that the set of facts of the default theory used in the definition above, i.e., of $\langle \emptyset, f(\mathsf{GI}_\sigma)\rangle$, is empty in our case. This means that we do not have an "indisputable" set of goals. The framework could be easily extended to include such a set of indisputable goals by adding these as facts to the default theory. However, the semantics as defined above has a close relation with the consistent subset semantics, as will be shown in Sect. 4.3. In the sequel, we will omit the set of facts and speak of extensions of a set of default rules.

We have defined the semantics of goals using a credulous interpretation. If one would define a skeptical semantics, one would specify that $\mathbf{G}\phi$ holds, iff ϕ follows from all extensions. This would mean that only the non-conflicting parts of the extensions are used for deriving goals. This is not what we want, as an agent may *have* conflicting goals, but it should take care that it handles these appropriately during execution.

The idea that goals have a credulous semantics is also supported by [42]. That paper proposes "an argument-based semantics for combined epistemic and practical reasoning, taking seriously the idea that in certain contexts epistemic reasoning is sceptical while practical reasoning is credulous". Practical reasoning here means reasoning about which of multiple, possibly conflicting, objectives to pursue, which is argued to be generally a credulous form of reasoning.

We illustrate the difference between a credulous and skeptical semantics using a simple example. The two goal inference rules $\{sushi\}^{\kappa^-} \Rightarrow pizza$ and $\{pizza\}^{\kappa^-} \Rightarrow sushi$ express that *sushi* and *pizza* are conflicting. These goal inference rules have $\top : \neg sushi, pizza/pizza$ and $\top : \neg pizza, sushi/sushi$ as corresponding default rules, with the extensions $\{sushi\}$ and $\{pizza\}$, which reflects that *sushi* and *pizza* are conflicting. We would like to have that $\mathbf{G}(sushi) \wedge \mathbf{G}(pizza)$ holds, given these goal inference rules, since, even though the goals *sushi* and *pizza* are conflicting, they *are* both goals. This is achieved by taking the credulous interpretation, i.e., the credulous interpretation allows the agent to have conflicting goals. In the skeptical interpretation, the agent would not have any goals in this example.

Note that the fact that the agent has both *sushi* and *pizza* as goals does not mean that it will also pursue both of these goals. In this paper, we are concerned with the representation of goals that the agent would in principle like to pursue as opposed to those that the agent is currently pursuing (see Sect. 2.2). The former ones may be conflicting, while the latter ones should be conflict-free. Once the agent starts the pursuit of goals, it will have to take into account the conflicts between goals and make sure that it does not pursue conflicting goals simultaneously, e.g., that it chooses between pursuing *sushi* or pursuing *pizza*. The latter is, however, not addressed in this paper.

Another related important point to note is that the formula $\mathbf{G}(sushi \wedge pizza)$ does *not* hold in the example. The semantics of goal formulas thus prevents the conjunction of two conflicting goals. This is important, since the fact that these goals are conflicting would otherwise not be reflected in the semantics. This is analogous to the consistent subset semantics, in which the semantics does not allow to derive, e.g., $\mathbf{G}(sushi \wedge \neg sushi)$ from the conflicting goals $\mathbf{G}(sushi)$ and $\mathbf{G}(\neg sushi)$. In the consistent subset semantics, we aimed for two mutually inconsistent goals not leading to the derivation of one inconsistent goal. We believe it to be intuitive that the same kind of behavior, i.e., the behavior that mutually conflicting goals ϕ and ψ do not lead to the derivation of one conjunctive goal $\phi \wedge \psi$, is exhibited by the default semantics.

Another aspect worth mentioning about this semantics is the following. Since goals may be conditional on the agent's beliefs, the goals may change as the beliefs change. For example, assume that the agent has a goal inference rule $\{sunny\}^{\beta} \Rightarrow sunscreen$ saying that if it is sunny, the agent may derive the goal of putting on sunscreen. Now assume the agent first believes it is sunny, and then the sun disappears (and the agent's beliefs are updated accordingly). In this case, the agent first can derive the goal to put on sunscreen, and after the sun has gone, it will no longer be able to derive this goal. That is, one could say that the agent has dropped the goal of putting on sunscreen.

In some approaches to goals, goals are required to have a certain persistency in that they are not dropped until they are believed to be achieved (see Sect. 2.2). As illustrated by the example above, goals do not have this kind of persistency in our approach since goals are

conditional on beliefs. That is, goals may be dropped if the beliefs change, even though they have not yet been reached.

We believe it is desirable that the goals change in this way as the beliefs of the agent change, when considering the goals that the agent would in principle like to pursue, as we do in this paper (see Sect. 2.2). If the agent has not yet started to put on sunscreen, we find it intuitive that the goal of putting on sunscreen is dropped if it is no longer sunny. For the goals that the agent is currently pursuing, it makes more sense to require that these have more persistency, to get a certain stability in the agent's behavior. In our example, this would mean that if the agent has started to put on sunscreen, it should not stop because the sun has disappeared for a moment.

4.2.3 Examples

We present two simple examples, in order to provide a better idea of the kinds of situations which can be modeled using goal inference rules. The first example is about an agent which has to carry cargo from a source location to a target location.

Example 2 (*carrying cargo*) Consider that one wants to express that if the agent is at the location of the source, it should have the goal to have cargo, and if it is at the target, it should have the goal not to have cargo. Further, if the agent believes he is at the source and he has cargo, he should have the goal to be at the target. Finally, if he believes he is at the source, and has the goal to be at the target, he should have the goal to be at some waypoint[12] in-between the source and the target. This can be useful if the agent only has plans to get from the source to the waypoint, and from the waypoint to the target. This could be modeled using the following goal inference rules.

$$\begin{aligned} \{source, \neg haveCargo\}^{\beta} &\Rightarrow haveCargo \\ \{target, haveCargo\}^{\beta} &\Rightarrow \neg haveCargo \\ \{source, haveCargo\}^{\beta} &\Rightarrow target \\ \{source\}^{\beta}, \{target\}^{\kappa^{+}} &\Rightarrow waypoint \end{aligned}$$

If we assume the agent believes he is at the source and has cargo, then $\mathbf{G}(target)$ and $\mathbf{G}(waypoint)$ will hold. Also, the formula $\mathbf{G}(target \wedge waypoint)$ holds, as the two goals are not conflicting.

This example illustrates that goals might be conditional on beliefs, and also on other goals. The fourth rule is an example of the specification of the derivation of landmarks, as it was called in [55]. A landmark is a goal which the agent has to achieve, on its way to achieving another goal. In this example, the waypoint can be viewed as a landmark, which the agent has to achieve in order to achieve the goal of being at the target.

The next example, which we have already mentioned above, is about an agent wanting either sushi or pizza, if he is hungry.

Example 3 (*sushi or pizza*) Consider that one wants to express that if an agent is hungry, he may have pizza or sushi, but should not simultaneously pursue these goals, i.e., the goals pizza and sushi are conflicting. Moreover, if an agent has sushi, he wants to drink tea with it,

[12] According to *The American Heritage: Dictionary of the English Language*, a waypoint is a point between major points on a route, as along a track.

and if he has pizza, he wants to drink soda. This could be modeled using the following goal inference rules.

$$\begin{aligned} \{hungry\}^{\beta}, \{sushi\}^{\kappa^-} &\Rightarrow pizza \\ \{hungry\}^{\beta}, \{pizza\}^{\kappa^-} &\Rightarrow sushi \\ \{sushi\}^{\kappa^+} &\Rightarrow tea \\ \{pizza\}^{\kappa^+} &\Rightarrow soda \end{aligned}$$

Assuming the agent indeed believes he is hungry, the default rules corresponding with these goal inference rules are $\top : \neg sushi, pizza/pizza$, $\top : \neg pizza, sushi/sushi$, $sushi : tea/tea$, and $pizza : soda/soda$. There are two extensions of these default rules, i.e., $\{pizza,soda\}$ and $\{sushi, tea\}$. This means we have $\mathbf{G}(sushi \wedge tea)$ and $\mathbf{G}(pizza \wedge soda)$, but we do not have $\mathbf{G}(sushi \wedge pizza)$, $\mathbf{G}(tea \wedge soda)$, $\mathbf{G}(pizza \wedge tea)$, nor $\mathbf{G}(sushi \wedge soda)$.

This example illustrates how to express that the goals $\{sushi\}$ and $\{pizza\}$ are conflicting, and how to use this in combination with the derivation of other goals on the basis of these goals. The conflict between $\{sushi\}$ and $\{pizza\}$ is reflected by the fact that they are not part of the same extension, and by the fact that the conjunctive goal $\mathbf{G}(sushi \wedge pizza)$ cannot be derived. Using goal inference rules, we can thus express that certain goals are conflicting, even though they are logically consistent. Since the derivation of *tea* and *soda* depends on the derivation of $\{sushi\}$ and $\{pizza\}$, the conflict between the latter two is "transferred" to the former two.

Note that in order to express that two goals are mutually conflicting, one needs to incorporate a condition to express this conflict in *both* rules. Replacing the second rule by the rule $\{hungry\}^{\beta} \Rightarrow sushi$ does not yield the same behavior, for the following reason. The new rule would have $\top : sushi/sushi$ as its corresponding default rule. When deriving extensions, we can apply this second rule, which yields the extension $\{sushi\}$. The first default rule is not applicable anymore, due to its justification. Alternatively, we can apply the first rule, yielding the set $\{pizza\}$. The second rule is applicable, as *sushi* is consistent with this set. The resulting set $\{pizza, sushi\}$, however, is *not* an extension, as the justification of the first rule is violated. We thus have only the extension $\{sushi\}$ in this case, and consequently $\mathbf{G}(sushi)$ holds and $\mathbf{G}(pizza)$ does not hold.

That is, if one does not express that *pizza* is in conflict with *sushi* in the second rule, it means that *sushi* can always be derived as a goal, preventing the derivation of the goal *pizza* using the first rule since *sushi* should occur in all extensions. We expect that in most cases in which one goal conflicts with another, the other goal also conflicts with the first, i.e., that if goals conflict they are mutually conflicting. In a practical setting, one could support the programmer by automatically defining goals as mutually conflicting, or giving a warning if the programmer has nog programmed this himself.

4.3 Properties

In this section, we investigate properties of the default semantics, and we show how it is related to the consistent subset semantics. The first kind of property we are interested in, is how the antecedent of goal inference rules is related to its consequent, i.e., the question is whether we can say anything about whether the consequent of a rule is a goal, given certain assumptions about the antecedent.

For a goal inference rule $\kappa^+ \Rightarrow \phi \in \mathsf{GI}$ it does not hold in general that if the formulas in κ^+ can be derived to be goals (on the basis of a belief base and GI), that ϕ can then also be derived to be a goal. This only holds if the formulas in κ^+ have not been defined as conflicting using other rules of GI. If the formulas in κ^+ are conflicting, they cannot be used to derive

the goal ϕ, since these would then not be part of the same extension, and the default rule corresponding to this goal inference rule would not be applicable. We formally define when goals are non-conflicting and, using this, formulate the property that ϕ can be derived as a goal on the basis of a set of goal inference rules GI if the formulas in κ^+ are non-conflicting with respect to GI.

Proposition 4 *Goals $\phi_1, \ldots, \phi_n \in \mathcal{L}$ are non-conflicting with respect to a belief base σ and set of goal inference rules GI, iff $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}(\phi_1 \wedge \ldots \wedge \phi_n)$. Let $\{\phi_1, \ldots, \phi_n\}^{\kappa^+} \Rightarrow \chi \in \mathsf{GI}$ be a goal inference rule. Then the following holds: if $\phi_1, \ldots, \phi_n$ are non-conflicting with respect to a belief base σ and set of goal inference rules GI, and $\langle \sigma, \mathsf{GI} \rangle \models_d \neg\mathbf{G}\neg\chi$, then we have $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\chi$.*

Proof Let E be an extension of $\langle \emptyset, f(\mathsf{GI}_\sigma) \rangle$. Assume $\phi_1, \ldots, \phi_n$ are non-conflicting with respect to a belief base σ and set of goal inference rules GI. This means that $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}(\phi_1 \wedge \ldots \wedge \phi_n)$. This means that $\exists E : E \models \phi_1 \wedge \ldots \wedge \phi_n$ (Definition 7). Assume $\langle \sigma, \mathsf{GI} \rangle \models_d \neg\mathbf{G}\neg\chi$, which means that $\neg\exists E : E \models \neg\chi$. The default rule corresponding with the goal inference rule $\{\phi_1, \ldots, \phi_n\}^{\kappa^+} \Rightarrow \chi$ is $\phi_1 \wedge \ldots \wedge \phi_n : \chi / \chi$. This default rule is applicable to the extension E for which $E \models \phi_1 \wedge \ldots \wedge \phi_n$, as we know that $E \not\models \neg\chi$. As extensions are closed under the application of default rules, we have that $E \models \chi$, which means that $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\chi$, yielding the desired result. □

In Proposition 4, we have shown how the antecedent κ^+ is related to the consequent of its goal inference rule. We now show how κ^- is related to the consequent of its goal inference rule. A goal inference rule $\kappa^- \Rightarrow \phi$ expresses that ϕ can in principle always be inferred as a goal, but it conflicts with the goals of κ^-. This means that ϕ can always be inferred as a goal, with the exception of the case where the formulas of κ^- are part of all extensions. If that is the case, the default rule corresponding with this goal inference rule can never be applied to derive ϕ. This property is formulated formally as follows, where we use the notion of "definite goal" to indicate that a goal can be derived on the basis of every extension.

Proposition 5 *A definite goal ϕ, represented as $\mathbf{G}_d\phi$, is defined as follows.*

$$\langle \sigma, \mathsf{GI} \rangle \models_{dd} \mathbf{G}_d\phi \Leftrightarrow \forall E : E \text{ is an extension of } \langle \emptyset, f(\mathsf{GI}_\sigma) \rangle \text{ and } E \models \phi$$

Let $\{\psi_1, \ldots, \psi_m\}^{\kappa^-} \Rightarrow \chi \in \mathsf{GI}$ be a goal inference rule. Then the following holds: if $\langle \sigma, \mathsf{GI} \rangle \models_{dd} \neg\mathbf{G}_d(\psi_1 \wedge \ldots \wedge \psi_m)$ and $\langle \sigma, \mathsf{GI} \rangle \models_d \neg\mathbf{G}\neg\chi$ hold, then we have $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\chi$.

Proof Let E be an extension of $\langle \emptyset, f(\mathsf{GI}_\sigma) \rangle$. Assume $\langle \sigma, \mathsf{GI} \rangle \models_{dd} \neg\mathbf{G}_d(\psi_1 \wedge \ldots \wedge \psi_m)$ and $\langle \sigma, \mathsf{GI} \rangle \models_d \neg\mathbf{G}\neg\chi$ hold. This means that $\exists E : E \not\models \psi_1 \wedge \ldots \wedge \psi_m$ and $\neg\exists E : E \models \neg\chi$. The default rule corresponding with the goal inference rule $\{\psi_1, \ldots, \psi_m\}^{\kappa^-} \Rightarrow \chi$ is $\top : \neg\psi_1, \ldots, \neg\psi_m / \chi$. This rule is applicable to the extension E for which $E \not\models \psi_1 \wedge \ldots \wedge \psi_m$, as we know that $E \not\models \neg\chi$. As extensions are closed under the application of default rules, we have that $E \models \chi$, which means that $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\chi$, yielding the desired result. □

The next proposition establishes which axioms of modal logic are satisfied by the default semantics. For ease of reference, we repeat the axioms below.

$$\begin{aligned}
\mathbf{K} &: \mathbf{G}(\phi \rightarrow \psi) \rightarrow (\mathbf{G}\phi \rightarrow \mathbf{G}\psi) \\
\mathbf{D_1} &: \neg\mathbf{G}\bot \\
\mathbf{D_2} &: \neg(\mathbf{G}\phi \wedge \mathbf{G}\neg\phi) \\
\mathbf{M} &: \mathbf{G}(\phi \wedge \psi) \rightarrow (\mathbf{G}\phi \wedge \mathbf{G}\psi) \\
\mathbf{C} &: (\mathbf{G}\phi \wedge \mathbf{G}\psi) \rightarrow \mathbf{G}(\phi \wedge \psi)
\end{aligned}$$

The proposition shows that in the general case, the default semantics satisfies the same axioms (from the ones stated above) as the consistent subset semantics (Proposition 2). Also, if there is only one extension, it satisfies the same axioms as the consistent subset semantics in case the goal base was consistent and therefore also of the basic semantics.

Proposition 6 *Let σ be an arbitrary belief base and let GI be an arbitrary set of goal inference rules. Then the following holds.*

$$\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{D_1}, \mathbf{M}$$

Let $f(\mathsf{GI}_\sigma)$ have only one extension. Then the following holds.

$$\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{D_2}, \mathbf{C}, \mathbf{K}$$

Proof As we do not have a set of facts in the default theory resulting from the goal inference rules, extensions are always consistent. Therefore, we have $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{D_1}$. If $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}(\phi \wedge \psi)$, it means there is an extension of $f(\mathsf{GI}_\sigma)$ from which $\phi \wedge \psi$ follows. This means there is an extension from which ϕ and ψ follow, which means that $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\phi \wedge \mathbf{G}\psi$, yielding $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{M}$.

If there is only one extension, it means that the existential quantification in the definition of the default semantics always refers to the same extension. This means that if $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\phi \wedge \mathbf{G}\psi$, we have that ϕ and ψ follow from the same single extension. This means that we also have $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}(\phi \wedge \psi)$, yielding $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{C}$, and for similar reasons we also have $\mathbf{K}$. Further, if there is only one extension, we have that it cannot be the case that both ϕ and $\neg\phi$ follow from this extension, as the extension is consistent. Therefore, we have $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{D_2}$. □

We continue to investigate how the default semantics is related to the consistent subset semantics. As we will see, the two can be related if we transform the goal base into goal inference rules in the appropriate way, i.e., by transforming each formula ϕ in the goal base into a goal inference rule $\top \Rightarrow \phi$, as formally defined below.

Definition 8 *(goal base to goal inference rules)* The function $g\colon \wp(\mathcal{L}) \rightarrow \wp(\mathcal{R}_{\mathsf{GI}})$, taking a goal base and yielding a set of goal inference rules, is defined as follows: $g(\gamma) = \{\top \Rightarrow \phi \mid \phi \in \gamma\}$.

Note that the default rules corresponding with these goal inference rules of the form $\top \Rightarrow \phi$, have the form $\top : \phi/\phi$. Default rules of this form are often called Poole-type defaults, or supernormal defaults (see, e.g., [5, 41]).

We now have the following theorem, which specifies that if the goal base is transformed into goal inference rules in this way, the consistent subset semantics and the default semantics are equivalent. The consistent subset semantics can thus be considered as a special case of the default semantics, i.e., the case where only goal inference rules are used that correspond to Poole-type defaults.

Theorem 1 *Let $\mathsf{GI} = g(\gamma)$ and let σ be an arbitrary belief base. Then the following holds.*

$$\gamma \models_s \kappa \Leftrightarrow \langle \sigma, \mathsf{GI} \rangle \models_d \kappa$$

In the proof of this theorem, we use the following lemma, in which the notion of a maximal consistent subset is used. 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 $\{\top : \phi/\phi \mid \phi \in \gamma\}$* [7].

Proof Assume there is a consistent subset γ' of γ such that $\gamma' \models \phi$. If γ' is a maximal consistent subset, we are done. If γ' is not a maximal consistent subset, we add formulas $\phi' \in \gamma$ with $\gamma' \cup \{\phi'\} \not\models \bot$ to γ' until γ' becomes maximally consistent. A maximal consistent subset is also a consistent subset, yielding the proof in the other direction. The second part of the lemma is proven in [7]. □

Proof of Theorem 1 We prove the result by induction on the structure of κ. Let κ be of the form $\mathbf{G}\phi$. We then have $\gamma \models_s \mathbf{G}\phi$ iff there is a consistent subset γ' of γ such that $\gamma' \models \phi$ (Definition 4). By Lemma 1, this is equivalent to there being a maximal consistent subset γ' of γ such that $\gamma' \models \phi$.

We have that $\langle \sigma, \mathsf{GI} \rangle \models_d \mathbf{G}\phi$ iff $\exists E$: E is an extension of $\langle \emptyset, f(\mathsf{GI}_\sigma) \rangle$ and $E \models \phi$ (Definition 7). As $\mathsf{GI} = g(\gamma)$, goal adoption rules are not conditional on the beliefs, and therefore $\mathsf{GI}_\sigma = \mathsf{GI}$.

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 $E \models \phi$.[13] By Definition 8, we have that $g(\gamma) = \{\top \Rightarrow \phi \mid \phi \in \gamma\}$ 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 for the case where κ is of the form $\mathbf{G}\phi$.

If κ is of the form $\kappa_1 \wedge \kappa_2$, we have that $\gamma \models_s \kappa_1 \wedge \kappa_2$ iff $\gamma \models_s \kappa_1$ and $\gamma \models_s \kappa_2$, and analogously for the default semantics by definition. By induction, we have that $\gamma \models_s \kappa_1$ iff $\langle \sigma, \mathsf{GI} \rangle \models_d \kappa_1$ and analogously for κ_2, yielding the desired result for the case where κ is of the form $\kappa_1 \wedge \kappa_2$. The case for negation is analogous. □

5 Related work in normative systems

Normative systems are systems in the behavior of which norms play a role and which need normative concepts in order to be described or specified [35]. There are different kinds of norms, such as permissions, prohibitions and obligations. In particular, obligations seem to be related to goals, as they express that a certain state of affairs should be achieved, or that certain actions need to be executed. In this section, we discuss work in the context of normative systems that is based on default logic or other defeasible logics.

5.1 Van Fraassen and Horty

In this section, we discuss the relation of our work with work on deontic logic by Horty [29–31], and with the work of Van Fraassen [53], which Horty addresses in his work. Our work as presented in this paper has been developed independently from the work of Horty and Van Fraassen. It however turns out that some of it is closely related to their work.

Deontic logics are logics for describing normative reasoning. Since its inception in the work of Von Wright [60], deontic logic has been developed primarily as a species of modal logic. In [29], Horty however argues that these modal deontic logics do not allow *normative conflicts*. He argues that normative conflicts occur often in everyday life, and that it is thus important that deontic logics are designed that can be used to represent and reason with these conflicts.

[13] A similar proposition was used, although not proven, by Reiter [47].

A situation gives rise to a normative conflict, if two conflicting propositions can both be said to be obligatory in that situation. Horty considers propositions to be conflicting if they are logically inconsistent. A situation of normative conflict thus occurs if both $\bigcirc\phi$ and $\bigcirc\neg\phi$ hold for some proposition ϕ, where $\bigcirc$ stands for "*o*bligation" or "*o*bliged to". As discussed in Sect. 3.2, in a normal modal logic K, this would imply $\bigcirc(\phi \wedge \neg\phi)$ and therefore $\bigcirc\bot$, thereby trivializing the logic. In standard deontic logic, besides the axiom **K**, also the axiom **D**, i.e., $\neg(\bigcirc\phi \wedge \bigcirc\neg\phi)$, is adopted. By adopting this axiom, standard deontic logic thus rules out normative conflicts [29].

In [29] and the follow-up papers [30, 31], Horty discusses an approach to reasoning in the presence of normative conflicts which was first proposed by Van Fraassen [53]. The latter paper contains two suggestions, where the second is a refinement of the first. Departing from modal logic and its possible world semantics, Van Fraassen defines obligations on the basis of a set of so-called background imperatives. These background imperatives are essentially propositional formulas, and are supposed to represent the (possibly conflicting) obligations as arising from various sources.

Van Fraassen's initial suggestion is to define the obligations that can be derived from a set of background imperatives γ, as follows:

$$\gamma \models_{F1} \bigcirc\phi \Leftrightarrow_{def} \exists\phi' \in \gamma : \phi' \models \phi.^{14}$$

Comparing this definition to Hindriks' definition for the semantics of goals (Definition 3), we can see that it is completely analogous. That is, with the exception that Hindriks requires each individual goal in γ to be consistent. Van Fraassen's definition will allow the derivation of any obligation if there is an inconsistent obligation in the set of background imperatives, while Hindriks prevents this by requiring that each goal in the goal base is consistent. The motivations provided by both authors for their definitions are also very similar: ϕ is a goal or obligation if it is a necessary condition for fulfilling a goal or obligation in the goal base or set of background imperatives, respectively.

As noted by Horty [30], this initial suggestion runs into difficulties, however, when it comes to logical interconnections among imperatives. The example provided by Van Fraassen and Horty to illustrate these difficulties, is the following. Suppose that $\gamma = \{p \vee q, \neg p\}$ is the set of background imperatives. Intuitively, one would want to conclude from this that $\bigcirc q$. This however does not follow under Van Fraassen's initial definition, as there is no single imperative from which q follows. To remedy this problem, Van Fraassen provides another and somewhat involved model theoretic definition, which we will refer to using $\models_{F2}$. We do not repeat that definition here, since this would require the introduction of a number of auxiliary notions, and the definition itself is not important for the current discussion.

What *is* important, is that Horty provides an equivalent definition by translating the set of background imperatives of Van Fraassen into default rules [29]. To be more specific, each formula ϕ in the set of background imperatives γ is translated into a default rule $\top : \phi/\phi$. Horty then shows the following, where D_γ is the resulting set of default rules:

$$\gamma \models_{F2} \bigcirc\phi \Leftrightarrow \exists E : E \text{ is an extension of } \langle\emptyset, D_\gamma\rangle \text{ and } E \models \phi.$$

By Theorem 1, we then have that $\models_{F2}$ is equivalent to our consistent subset semantics. This analysis thus suggests that goals and obligations are indeed similar. It also shows, using the results of Horty, that the second proposal of Van Fraassen is a special case of our default semantics of goals. Moreover, our analysis of the relation between Hindriks' semantics and

[14] We rephrase the definition as given in [30] for reasons of comparison, which in turn rephrases the definition given in [53].

the consistent subset semantics also shows how the proposals of Van Fraassen are related, as we have that his first proposal is equivalent to that of Hindriks et al., and his second proposal is equivalent to the consistent subset semantics.

5.2 BOID and related approaches

The idea of using default logic to define the semantics of goal inference rules was inspired by the BOID framework [9, 13], which uses default logic for representing beliefs, obligations, intentions and desires. This framework was in turn inspired by Thomason [52], who uses default logic to develop a formalism to integrate reasoning about desires with planning, and Horty [30], and it is related to work in the area of defeasible logic by Governatori and Rotolo [23]. In this section, we discuss the main differences between our work and these approaches. We will refer to the first variant of BOID [9] as BOID'02 and to the second [13] as BOID'04.

The main difference between our work and the other approaches mentioned above, is that we introduce a *logical language of goals of which we define the semantics*. In the work of Thomason and BOID'02, sets of formulas (extensions) are generated on the basis of (normal) default rules. The formulas (or part of them) are intended to represent the goals (or desires) of the agent. However, these approaches do not come with a logical language of goals. Such a language provides the means for unambiguously expressing that a formula *is* a goal, and also that a formula is *not* a goal. Our logics of goals have allowed us to compare properties of our goal operator with properties of operators from modal logics.

In BOID'04, sets of modal formulas with **B**, **O**, **I**, and **D** operators are generated on the basis of rules comparable to normal default rules. In contrast with normal default rules, however, these rules contain modal formulas rather than propositional (or first order) formulas. These modal formulas are used to express whether the agent has a belief, obligation, intention, or desire. The authors assume some modal logic consequence relation for establishing whether a rule can be applied. BOID'04 thus does not define the semantics of a logical language of goals like we do, but the semantics of their operators is stated to be that of standard modal operators. Governatori and Rotolo [23] build on work on defeasible logic by Nute [39] (see also [12] for a follow-up paper). Their logic allows to derive tagged literals, where the tags express whether the literal is a belief, obligation, etc. They thus do not define the semantics of a logical language of goals.

Another difference is that we exploit the *full power of default rules*, including the justification of the default rules, while in Thomason and BOID only normal default rules are considered. The work of Governatori and Rotolo is an extension of defeasible logic. Defeasible logic has a *skeptical* semantics, while we have argued that a credulous semantics is more appropriate in the context of conflicting goals. Other differences between our work and BOID'04 are that we use standard propositional default logic to interpret our rules, while BOID'04 defines its own procedures for generating extensions, in which a modal logic consequence relation is assumed. Further, in BOID'04, the extensions which are computed are explicitly stored in the agent configurations, including the extension which is chosen to form the agent's intention base.

6 Conclusion

We have explored semantics of a logical language of goals that support the representation of conflicting goals, in the context of logic-based cognitive agent programming languages. We have investigated semantics based on unconditional and conditional goals, and have examined their properties and interrelations. The former are based on propositional logic, and the latter is based on default logic.

Regarding the unconditional semantics, we have considered a basic semantics, the semantics of Hindriks et al., and our own consistent subset semantics. We have shown that the logic of goals is trivialized under the basic semantics if the goal base is inconsistent. Also, we have shown that the consistent subset semantics allows the derivation of more goals than the semantics of Hindriks et al., and that it is equivalent with the basic semantics in case of a consistent goal base. The semantics of Hindriks et al. and the consistent subset semantics turn out to be equivalent to proposals by Van Fraassen in the area of deontic logic.

The main advantages of the default semantics are that it allows the representation of conditional goals, and it allows to express that goals are conflicting, even though they are logically consistent. We have shown that the consistent subset semantics is a special case of the default semantics. The main difference between our work and BOID-like approaches is that, in contrast to those approaches, we introduce a logical language of goals of which we define the semantics in various ways. This facilitates the investigation of properties of the semantics and their comparison, and provides for added expressivity.

We see two main directions of future work. First, we aim to investigate how we can implement our semantics of goals using answer set programming (see Sect. 2.1). Second, we plan to investigate in more detail how our approach can be embedded into existing agent programming languages. The main issue to address will be the relation between the goals that the agent would in principle like to pursue as considered in this paper (see Sect. 2.2), and goals the agent is actively pursuing.

Concluding, we maintain that a systematic analysis of semantics of 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. Also, we would like to thank Jan Broersen, Joris Hulstijn, and Leon van der Torre for discussions on the BOID framework. Finally, we would like to thank the anonymous referees for many valuable comments and suggestions.

References

1. Antoniou, G. (1997). *Nonmonotonic reasoning. Artificial intelligence*. Cambridge, MA: The MIT Press.
2. Besnard, P., & Hunter, A. (1995). Quasi-classical logic: Non-trivializable classical reasoning from inconsistent information. In *Symbolic and quantitative approaches to reasoning and uncertainty* (pp. 44–51). Berlin: Springer.
3. Bordini, R. H., Dastani, M., Dix, J., & El Fallah Seghrouchni, A. (2005). *Multi-agent programming: Languages, platforms and applications*. Berlin: Springer.
4. Braubach, L., Pokahr, A., Moldt, D., & Lamersdorf, W. (2005). Goal representation for BDI agent systems. In *Programming multiagent systems, second international workshop (ProMAS'04)*, volume 3346 of LNAI (pp. 44–65). Berlin: Springer.

5. Brewka, G. (1991). *Nonmonotonic reasoning: logical foundations of commonsense.* Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.
6. Brewka, G. (1994). Adding priorities and specificity to default logic. In *Logics in artificial intelligence (JELIA'94)*, volume 838 of LNCS (pp. 247–260). Berlin: Springer-Verlag.
7. Brewka, G., Dix, J., & Konolige, K. (1997). *Nonmonotonic reasoning: An overview*. Stanford: CSLI Publications.
8. Brewka, G., & Eiter, T. (2000). Prioritizing default logic. In *Intellectics and computational logic* (pp. 27–45). Devender: Kluwer.
9. Broersen, J., Dastani, M., Hulstijn, J., & van der Torre, L. (2002). Goal generation in the BOID architecture. *Cognitive Science Quarterly, 2*(3–4), 428–447.
10. Chellas, B. F. (1980). *Modal logic: An introduction*. Cambridge: Cambridge University Press.
11. Cohen, P. R., & Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence, 42*, 213–261.
12. Dastani, M., Governatori, G., Rotolo, A., & van der Torre, L. (2005). Programming cognitive agents in defeasible logic. In *Proceedings of Logic for Programming, Artificial Intelligence, and Reasoning (LPAR'05)*, volume 3835 of LNAI (pp. 621–637). Berlin: Springer-Verlag.
13. Dastani, M., & van der Torre, L. (2004). Programming BOID-Plan agents: Deliberating about conflicts among defeasible mental attitudes and plans. In *Proceedings of the 3rd Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04)* (pp. 706–713). New York, USA.
14. Dastani, M., van Riemsdijk, M. B., Dignum, F., & Meyer, J.-J. Ch. (2004). A programming language for cognitive agents: Goal directed 3APL. In *Programming multiagent systems, first international workshop (ProMAS'03)*, volume 3067 of LNAI (pp. 111–130). Berlin: Springer.
15. de Boer, F., Hindriks, K., van der Hoek, W., & Meyer, J.-J. (2007). A Verification framework for agent programming with declarative goals. *Journal of Applied Logic, 5*, 277–302.
16. Delgrande, J. P., & Schaub, T. (1997). Compiling reasoning with and about preferences into default logic. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI'97)* (pp. 168–175).
17. Duff, S., Harland, J., & Thangarajah, J. (2006). On proactivity and maintenance goals. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)* (pp. 1033–1040), Hakodate.
18. Emerson, E. A., & Srinivasan, J. (1989). Branching time temporal logic. In *Linear time, branching time and partial order in logics and models for concurrency, school/workshop* (pp. 123–172). London, UK: Springer-Verlag.
19. Fagin, R., & Halpern, J. (1988). Belief, awareness and limited reasoning. *Artificial Intelligence, 34*, 39–76.
20. Gabbay, D., & Hunter, A. (1991). Making inconsistency respectable: A logical framework for inconsistency in reasoning. In P. Jorrand & J. Kelemen (Eds.), *Proceedings of Fundamentals of Artificial Intelligence Research (FAIR'91)* (pp. 19–32). Berlin: Springer-Verlag.
21. Gelfond, M., & Lifschitz, V. (1990). Logic programs with classical negation. In *Logic programming* (pp. 579–597). Cambridge: MIT Press.
22. Gelfond, M., & Lifschitz, V. (1991). Classical negation in logic programs and disjunctive databases. *New Generation Computing, 9*(3/4), 365–386.
23. Governatori, G., & Rotolo, A. (2004). Defeasible logic: Agency, intention and obligation. In A. Lomuscio & D. Nute (Eds.), *Deontic logic in computer science (DEON'04)*, volume 3065 of LNAI, (pp. 114–128). Berlin: Springer.
24. Hansson, B. (1969). An analysis of some deontic logics. In *Nous 3*, 373–398.
25. Harrenstein, B. P. (2004). *Logic in conflict: Logical explorations in strategic equilibrium*. PhD thesis.
26. Hindriks, K., & Meyer, J.-J. Ch. (2006). Agent logics as program logics: Grounding KARO. In *Proceedings of the 29th German Conference on Artificial Intelligence (KI'06)*.
27. Hindriks, K. V., de Boer, F. S., van der Hoek, W., & Meyer, J.-J. Ch. (1999). Agent programming in 3APL. *International Journal of Autonomous Agents Multi-Agent Systems, 2*(4), 357–401.
28. Hindriks, K. V., de Boer, F. S., van der Hoek, W., & Meyer, J.-J. Ch. (2001). 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. Berlin: Springer.
29. Horty, J. F. (1993). Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence (Special Issue on Deontic Logic in Computer Science), 9*, 69–91.
30. Horty, J. F. (1994). Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic, 23*(1), 35–65.
31. Horty, J. F. (1997). Nonmonotonic foundations for deontic logic. In D. Nute (Ed.), *Defeasible deontic logic* (pp. 17–44). Dordrecht: Kluwer Academic Publishers.

32. Hübner, J. F., Bordini, R. H., & Wooldridge, M. (2006). Declarative goal patterns for AgentSpeak. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*.
33. Leone, N., Pfeifer, G., Faber, W., Eiter, T., Gottlob, G., Perri, S., et al. (2006). The DLV system for knowledge representation and reasoning. *ACM Transactions on Computational Logic, 7*(3), 499–562.
34. Marek, V., & Truszczynski, M. (1993). *Nonmonotonic logic: Context-dependent reasoning*. Berlin: Springer.
35. Meyer, J.-J. C., & Wieringa, R. J. (Eds.) (1993). *Deontic logic in computer science: Normative system specification*. Chichester, UK: Wiley and Sons Ltd.
36. Meyer, J.-J. Ch., & van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.
37. Niemelä, I., & Simons, P. (1997). Smodels—An implementation of the stable model and well-founded semantics for normal logic programs. In *Proceedings of the 4th International Conference on Logic Programming and Nonmonotonic Reasoning*, volume 1265 of Lecture Notes on Artificial Intelligence (pp. 420–429). Berlin: Springer Verlag.
38. Nigam, V., & Leite, J. (2006). A dynamic logic programming based system for agents with declarative goals. In M. Baldoni & U. Endriss, (Eds.), *Declarative agent languages and technologies IV (DALT'06)*, volume 4327 of LNAI (pp. 174–190). Berlin: Springer-Verlag.
39. Nute, D. (1994). Defeasible logic. In *Handbook of logic in artificial intelligence and logic programming* (Vol. 3, pp. 353–395). New York: Oxford University Press.
40. Pokahr, A., Braubach, L., & Lamersdorf, W. (2005). A goal deliberation strategy for BDI agent systems. In *MATES 2005*, volume 3550 of LNAI (pp. 82–93). Berline: Springer-Verlag.
41. Poole, D. (1988). A logical framework for default reasoning. *Artificial Intelligence, 36*, 27–47.
42. Prakken, H. (2006). Combining sceptical epistemic reasoning with credulous practical reasoning. In *Proceedings of the 1st International Conference on Computational Models of Argument* (pp. 311–322).
43. Rao, A. S. (1996). AgentSpeak(L): BDI agents speak out in a logical computable language. In W. van der Velde & J. Perram (Eds.), *Agents breaking away* (pp. 42–55), LNAI 1038. Berlin: Springer-Verlag.
44. Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, & E. Sandewall (Eds.), *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)* (pp. 473–484). San Francisco: Morgan Kaufmann.
45. Rao, A. S., & Georgeff, M. P. (1998). Decision procedures for BDI logics. *Journal of Logic and Computation, 8*(3), 293.
46. Reiter, R. (1980). A logic for default-reasoning. *Artificial Intelligence, 13*, 81–132.
47. Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence, 32*, 57–95.
48. Sardina, S., & Shapiro, S. (2003). Rational action in agent programs with prioritized goals. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)* (pp. 417–424), Melbourne.
49. Thangarajah, J., Padgham, L., & Winikoff, M. (2003). Detecting and avoiding interference between goals in intelligent agents. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI 2003)*.
50. Thangarajah, J., Padgham, L., & Winikoff, M. (2003). Detecting and exploiting positive goal interaction in intelligent agents. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)* (pp. 401–408), Melbourne.
51. Thangarajah, J., Winikoff, M., Padgham, L., & Fischer, K. (2002). Avoiding resource conflicts in intelligent agents. In F. van Harmelen (Ed.), *Proceedings of the 15th European Conference on Artificial Intelligence 2002 (ECAI 2002)*, Lyon, France.
52. Thomason, R. H. (2000). Desires and defaults: A framework for planning with inferred goals. In A. G. Cohn, F. Giunchiglia, & B. Selman (Eds.), *KR2000: Principles of knowledge representation and reasoning* (pp. 702–713). San Francisco: Morgan Kaufmann.
53. van Fraassen, B. C. (1973). Values and the heart's command. *Journal of Philosophy, 70*(1), 5–19.
54. van Riemsdijk, M. B. (2006). *Cognitive agent programming: A semantic approach*. PhD thesis.
55. van Riemsdijk, M. B., Dastani, M., Dignum, F., & Meyer, J.-J. Ch. (2005). Dynamics of declarative goals in agent programming. In J. A. Leite, A. Omicini, P. Torroni, & P. Yolum (Eds.), *Declarative agent languages and technologies II: Second international workshop (DALT'04)*, volume 3476 of LNAI pp. 1–18.
56. van Riemsdijk, M. B., Dastani, M., & Meyer, J.-J. Ch. (2005). Semantics of declarative goals in agent programming. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)* (pp. 133–140), Utrecht.

57. van Riemsdijk, M. B., Dastani, M., Meyer, J.-J. Ch., & de Boer, F. S. (2006). Goal-oriented modularity in agent programming. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)* (pp. 1271–1278), Hakodate.
58. van Riemsdijk, M. B., Dastani, M., & Winikoff, M. (2008). Goals in agent systems: A unifying framework. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08)* (pp. 713–720), Estoril.
59. van Riemsdijk, M. B., van der Hoek, W., & Meyer, J.-J. Ch. (2003). Agent programming in Dribble: From beliefs to goals using plans. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)* (pp. 393–400), Melbourne.
60. von Wright, G. H. (1951). Deontic logic. *Mind, 60*, 1–15.
61. Winikoff, M., Padgham, L., Harland, J., & Thangarajah, J. (2002). Declarative and procedural goals in intelligent agent systems. In *Proceedings of the 8th International Conference on Principles of Knowledge Representation and Reasoning (KR2002)*, Toulouse.