GOAL Agents Instantiate Intention Logic

Koen Hindriks[1] and Wiebe van der Hoek[2]

[1] Delft University of Technology, The Netherlands, k.v.hindriks@tudelft.nl
[2] University of Liverpool, UK wiebe@csc.liv.ac.uk

Abstract. It is commonly believed there is a big gap between agent logics and computational agent frameworks. In this paper, we show that this gap is not as big as believed by showing that GOAL agents instantiate Intention Logic of Cohen and Levesque. That is, we show that GOAL agent programs can be formally related to Intention Logic. We do so by proving that the GOAL Verification Logic can be embedded into Intention Logic. It follows that (a fragment of) Intention Logic can be used to prove properties of GOAL agents. The work reported is an important step towards the application of standard tools from modal logic for e.g. model checking agent programs. Our results also prove useful for extending the expressiveness of the GOAL agent language. This is illustrated by incorporating temporally extended goals into GOAL agents.

1 Introduction

As has been observed by many others, there is still a considerable gap between logical theories of rational agents and most computational frameworks for such agents [10, 12]. Though it is generally hard to connect computational frameworks for rational agents to logics for such agents, in this paper we show that it is possible to formally relate the GOAL agent programming language [4, 8] and Intention Logic of Cohen and Levesque [3]. The result proven establishes that GOAL agents instantiate the theory of rational agents as proposed by Intention Logic, although we also argue that the theory needs revision at a number of points.

The motivation behind our work is the observation that there are a number of basic similarities between Intention Logic and the GOAL Verification Logic ("GOAL Logic" for short, see [4]). Most notably, both are based on linear time frames and both incorporate basic notions of a common sense perspective on rational action - beliefs and goals in relation to action. Intention Logic has been proposed as a *theory of the "rational balance" of beliefs, goals, intentions and actions*, inspired by Bratman's theory of intention. It thus proposes a set of rationality principles rational agents should comply with. The GOAL agent programming language is based on and aspires to incorporate similar rationality principles, and has been proposed as a *theory of computation* based on the common-sense notions of belief and goal. Relating both formally thus would be a significant step in bridging the gap between agent theory and engineering.

Establishing a formal connection between GOAL and Intention Logic is useful for a number of reasons. First of all, it connects the GOAL agent programming language to an agent logic in a formally precise sense, contributing to one of the long-standing

challenges of agent research of bridging the gap between agent theory and agent programming [10]. It shows that agent logics such as Intention Logic can be applied and used for the verification of properties of computational agents. Conceptually it is interesting to compare the agent concepts and rationality principles incorporated in Intention Logic with those used by GOAL agents. Related to this we show that establishing a formal connection turns out to be useful for extending GOAL agents with temporally extended goals [1]. On top of this, technically, the mapping of GOAL Logic into a standard modal logic is useful since it makes available the rich set of tools available for such logics. These include, for example, tools for model checking, which can be used to achieve one of the main goals of our work - to establish verification tools that can be practically applied to computational rational agents. Finally, combining the frameworks of two approaches also has an effect in the opposite direction: we will argue that assumptions made for Intention Logic can be broadly categorised threefold: those that constitute a basic logic for intention, those that can be conceived of as natural in special situations, and those that seem to be not necessary, or even, not intuitive.

The paper is organized as follows. In Section 2 we briefly introduce the agent programming language GOAL and its verification logic as proposed in [4]. In Section 3 the propositional fragment of Intention Logic used in this paper is introduced. In Section 4 we show that GOAL Logic can be embedded into Intention Logic. In Section 5 we (re)use the embedding proof to show how to incorporate temporally extended goals into GOAL agents. Finally, in Section 6 we conclude the paper and discuss possible directions for future work.

2 The Agent Programming Language GOAL

GOAL agents derive their choice of action from their *beliefs* and their *goals*. GOAL agents consist of four components: (i) a set of beliefs called a *belief base*, (ii) a set of goals called a *goal base*, (iii) a set of action rules, called the *agent program*, and (iv) a set of *action specifications*. The beliefs and goals are drawn from some logical language. The basic ingredients needed are a knowledge representation language and associated inference relation and update operators. Here we follow [4] and throughout the paper we assume a propositional language $\mathcal{L}_0$ (with typical elements ϕ) defined over a set of Atoms with entailment operator $\models$. The beliefs and goals of a GOAL agent define its *mental state*, which needs to satisfy a number of rationality constraints.

Definition 1. (Mental State)
A mental state *of a GOAL agent is a pair $\langle \Sigma, \Gamma \rangle$ with Σ a belief base and Γ a goal base consisting of sentences drawn from a classical propositional language $\mathcal{L}_0$, i.e. $\Sigma, \Gamma \subseteq \mathcal{L}_0$. A mental state needs to satisfy the following* rationality constraints*:*

- *Belief bases are consistent:* $\Sigma \not\models \mathbf{false}$,
- *Individual goals are consistent:* $\forall \gamma \in \Gamma : \gamma \not\models \mathbf{false}$,
- *Goals are not believed to be achieved:* $\forall \gamma \in \Gamma : \Sigma \not\models \gamma$.

Rational agents are assumed to have consistent beliefs and goals that are not (logically) impossible to achieve which motivates the introduction of the first two rationality constraints. Goals of a GOAL agent are *achievement goals* that the agent wants to

achieve some time in the future. As such, an agent may have multiple achievement goals that taken together are inconsistent but may be achieved in either order over time (cf. [4, 7, 8]). A GOAL agent is assumed to be committed to achieving these goals. A rational agent however will not invest resources in pursuing goals that are already (completely) achieved, which motivates the third rationality constraint.

In order to be able to decide on its next action a GOAL agent inspects its belief and goal bases. To do so, so-called *mental state conditions* are introduced to reason about the agent's beliefs and goals. The language $\mathcal{L}_m$ of mental state conditions extends $\mathcal{L}_0$ with a modal belief $\mathbf{B}$ and goal $\mathbf{G}$ operator, which can be used to express conditions on the mental state of an agent.

Definition 2. (Mental State Conditions: Syntax)
The language $\mathcal{L}_m$ (with typical elements ψ, ψ') of mental state conditions *is defined by:*

$$\begin{array}{l} \phi \in \mathcal{L}_0 \ ::= \textit{ any element in } \mathcal{L}_0 \\ \psi \in \mathcal{L}_m ::= \mathbf{B}\phi \mid \mathbf{G}\phi \mid \neg\psi \mid \psi \wedge \psi \end{array}$$

The set of mental state conditions consists of Boolean combinations of formulae of the form $\mathbf{B}\phi$ and $\mathbf{G}\phi$ with $\phi \in \mathcal{L}_0$. It is not allowed to nest the operators $\mathbf{B}$ and $\mathbf{G}$ in mental state conditions. Also note that simple propositional formulas without occurrences of $\mathbf{B}$ or $\mathbf{G}$ operators are not mental state conditions. These formulae are called *objective* and are used to represent properties of the agent's environment instead. The semantics of mental state conditions is evaluated with respect to mental states.

Definition 3. (Mental State Conditions: Semantics)
The semantics of mental state conditions is defined relative to a mental state $\langle \Sigma, \Gamma \rangle$.

$$\begin{array}{ll} \langle \Sigma, \Gamma \rangle \models \mathbf{B}\phi & \textit{iff } \Sigma \models \phi, \\ \langle \Sigma, \Gamma \rangle \models \mathbf{G}\phi & \textit{iff } \exists \gamma \in \Gamma \textit{ such that } \gamma \models \phi \textit{ and } \Sigma \not\models \phi, \\ \langle \Sigma, \Gamma \rangle \models \neg\psi & \textit{iff } \langle \Sigma, \Gamma \rangle \not\models \psi, \\ \langle \Sigma, \Gamma \rangle \models \psi \wedge \psi' & \textit{iff } \langle \Sigma, \Gamma \rangle \models \psi \textit{ and } \langle \Sigma, \Gamma \rangle \models \psi'. \end{array}$$

The semantics of the goal operator $\mathbf{G}$ defines an agent's achievement goals as those propositions that follow from a single goal in the agent's goal base that is not believed to be the case; in other words, $\mathbf{G}\phi$ expresses that ϕ is an achievement goal in this sense.

GOAL agents select actions using a rule-based action selection mechanism. In the remainder, we assume a set of actions A (with typical element α, α') has been provided. *Action rules* of the form $\mathbf{if}\ \psi\ \mathbf{then}\ \alpha$ are used to specify that action α can be performed, or, is *enabled*, whenever condition ψ holds, where ψ is a mental state condition. This mechanism allows agents to derive their choice of action from their beliefs and goals. The semantics of action selection and execution are formally specified in GOAL by means of an operational semantics; here, however, we abstract from the formal details (see [4]) and we will represent action selection implicitly by means of action occurrences in a set of possible traces. A *trace* simply is a sequence of mental states and actions.

Definition 4. (Trace)
A trace t *is an infinite sequence $m_0, \alpha_0, m_1, \alpha_1, \ldots$ of mental states m_i and actions α_i. We also write t_i^m to denote the ith mental state and t_i^a to denote the ith action.*

Intuitively, a trace corresponding to a possible computation of a GOAL agent needs to start with a mental state that corresponds to the initial state of the GOAL agent. The changes in mental states over time are the result of executing actions (which ideally correspond to changes in the agent's environment). Action rules and preconditions do not need to determine a unique action to be taken by the agent at a time point. The semantics associated with the action selection and execution of a GOAL agent thus does not define a unique computation but corresponds to a set of computations. This motivates defining the meaning of a GOAL agent $\mathcal{A}$ as a set of traces, in line with the fact that we abstract from the semantics of action selection and execution in this paper.

2.1 GOAL Logic

To obtain a verification logic for GOAL agents temporal operators are added on top of mental state conditions to be able to express temporal properties over traces. Additionally an operator $\mathbf{start}$ is introduced to be able to pinpoint the start of a trace.

Definition 5. (Temporal Language: Syntax)
The temporal language $\mathcal{L}_G$ *(with typical elements χ, χ') is defined by:*

$$\chi \in \mathcal{L}_G ::= \mathbf{start} \mid \psi \in \mathcal{L}_m \mid \neg\chi \mid \chi \wedge \chi \mid \chi \ \mathbf{until}\ \chi \mid [\alpha \in A]\chi$$

The semantics of $\mathcal{L}_G$ is defined relative to an agent $\mathcal{A}$, trace $t \in \mathcal{A}$ and time point i.

Definition 6. (Temporal Language: Semantics)
The truth conditions of sentences from $\mathcal{L}_G$ given a GOAL agent $\mathcal{A}$, trace $t \in \mathcal{A}$ and time point i are inductively defined by:

$$\begin{array}{ll}
\mathcal{A}, t, i \models \mathbf{start} & \textit{iff } i = 0, \\
\mathcal{A}, t, i \models \mathbf{B}\phi & \textit{iff } t_i^m \models \mathbf{B}\phi, \\
\mathcal{A}, t, i \models \mathbf{G}\phi & \textit{iff } t_i^m \models \mathbf{G}\phi, \\
\mathcal{A}, t, i \models \neg\varphi & \textit{iff } \mathcal{A}, t, i \not\models \varphi, \\
\mathcal{A}, t, i \models \varphi \wedge \psi & \textit{iff } \mathcal{A}, t, i \models \varphi \textit{ and } \mathcal{A}, t, i \models \psi, \\
\mathcal{A}, t, i \models \varphi \ \mathbf{until}\ \psi & \textit{iff } \exists j \geq i : \mathcal{A}, t, j \models \psi \textit{ and } \forall i \leq k < j : \mathcal{A}, t, k \models \varphi, \\
\mathcal{A}, t, i \models [\alpha]\varphi & \textit{iff } \forall t \in \mathcal{A}(t_i^a = \alpha \Rightarrow \mathcal{A}, t, i+1 \models \varphi).
\end{array}$$

Note that formulas of the form $[\alpha]\varphi$ specify *universal action postconditions*, in particular, we have $\mathcal{A}, t, i \models [\alpha]\varphi$ iff $\mathcal{A}, t', i' \models [\alpha]\varphi$ iff $\mathcal{A} \models [\alpha]\varphi$. This operator allows to define the Hoare system for GOAL which was proven complete in [4] and facilitates reasoning about actions. This operator is crucial in GOAL Logic to be able to compositionally prove properties of all traces induced by a GOAL agent [4].

3 Basic Intention Logic

Our interest in this paper is in the *single-agent, propositional fragment* of Intention Logic without dynamic (composition) operators such as sequential composition. In essence, Intention Logic can be considered a single-agent logic (cf. [12]) and the single agent restriction boils down to excluding multiple agent labels and variables ranging

over such labels from the logical language. The restriction to the propositional fragment implies that we do not introduce quantifiers and variables ranging over events, agents or domains. Temporal operators are also introduced explicitly in the language rather than defining these as rather complex quantifications over events. The fragment of Intention Logic introduced here is referred to henceforth as *Basic Intention Logic*, or sometimes also simply as Intention Logic.

Definition 7. (Basic Intention Logic: Syntax)
The language $\mathcal{L}_{BI}$ is defined by:

$$\begin{aligned}
\alpha &::= \textit{any element from } A \mid \mathsf{IF}\ \varphi\ \mathsf{THEN}\ \alpha\ \mathsf{ELSE}\ \mathsf{NIL},\\
\phi &::= \textit{any element from } \mathsf{Atom},\\
\varphi &::= \phi \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{BEL}\ \varphi \mid \mathsf{GOAL}\ \varphi \mid \mathsf{HAPPENS}\ \alpha \mid\\
&\quad\ \mathsf{DONE}\ \alpha \mid \mathbf{t} \mid \mathsf{BEFORE}\ \varphi\ \varphi \mid E\varphi,\\
\mathbf{t} &::= \textit{any non-negative numeral } (\mathbf{0}, \mathbf{1}, \ldots)
\end{aligned}$$

The main modification made to Intention Logic is the addition of a global modal operator E (cf. [2]). The operator HAPPENS is too weak to reason about *all possible effects* of executing an action which is crucial for verifying properties of the behaviour of an agent program (compare the dynamic operator $[\alpha]\chi$ introduced above and the usual dynamic modality in Dynamic Logic [6]). The standard abbreviations are used for true and disjunction $\vee$. Some additional abbreviations used are:

$$\mathsf{UNTIL}\ \varphi\ \psi \stackrel{df}{=} \neg(\mathsf{BEFORE}\ \psi\ \neg\varphi), \qquad \Diamond\varphi \stackrel{df}{=} (\mathsf{true}\ \mathsf{UNTIL}\ \varphi), \qquad \Box\varphi \stackrel{df}{=} \neg\Diamond\neg\varphi$$
$$\mathsf{KNOW}\ \varphi \stackrel{df}{=} \varphi \wedge \mathsf{BEL}\ \varphi, \qquad \mathsf{KNOWIF}\ \varphi \stackrel{df}{=} \mathsf{KNOW}\ \varphi \vee \mathsf{KNOW}\ \neg\varphi.$$

After introducing the fragment we refer to as *Basic Intention Logic*, the question remains how much of the *theory* of Intention Logic about rational agency survives. As it will turn out, a large part can be (re)formulated by using temporal operators only. This issue will be revisited at the end of this Section.

3.1 A Run-Based Semantics for Intention Logic

Semantically we first introduce a run-based semantics for Intention Logic and then discuss how our semantics relates to that introduced in [3]. Different from [7] we use standard linear orders $\mathbb{L}$ to define models for Intention Logic to ensure our models have the same basic structure as traces of GOAL agents. Here, we will restrict ourselves to $\mathbb{L} = \langle \mathbb{N}, < \rangle$ and $\mathbb{L} = \langle \mathbb{Z}, < \rangle$. We use linear orders to define the concept of a *run*.

Definition 8. (Run-Based Model)
Let an arbitrary set of labels S also called states *be given. A* run based on S and A *is a function $r : \mathbb{L} \to (S \times A)$ that assigns to every time point a state-action pair. Given $n \in \mathbb{L}$, we will write r_n^{st} for the first component of $r(n)$, and r_n^{ac} for the second. The set of runs based on S and A is denoted $\mathcal{R}(S, A)$.*

A run-based model M *(over* Atoms*) is a tuple $M = \langle S, \mathbb{L}, B, G, V \rangle$, where*

- *S is a non-empty set of* states*;*

- $\mathbb{L}$ *is a linear order;*
- $B \subseteq \mathcal{R} \times \mathbb{L} \times \mathcal{R} \times \mathbb{L}$ *is a Euclidean, transitive and serial belief accessibility relation,*
- $G \subseteq \mathcal{R} \times \mathbb{L} \times \mathcal{R} \times \mathbb{L}$ *is a serial goal accessibility relation, and*
- $V : S \to \mathsf{Atoms}$.

The semantics of Basic Intention Logic can now be defined using run-based models.

Definition 9. (Run-Based Semantics for Basic Intention Logic)
Let $M = \langle S, \mathbb{L}, B, G, V\rangle$ be a run-based model, $r \in \mathcal{R}$, and $n \in \mathbb{L}$. Then the satisfaction relation $\models$ relative to M is defined by:

$M, r, n \models p$	*iff* $p \in V(r^{st}_n)$
$M, r, n \models \neg\varphi$	*iff* $M, r, n \not\models \varphi$
$M, r, n \models \varphi \wedge \varphi'$	*iff* $M, r, n \models \varphi$ *and* $M, r, n \models \varphi'$
$M, r, n \models \mathbf{t}$	*iff* $\mathbf{t}$ *denotes* n,
$M, r, n \models \mathsf{DONE}\ \alpha$	*iff* $\exists j \in \mathbb{L}\ M, r, j[\![\alpha]\!]n$
$M, r, n \models \mathsf{HAPPENS}\ \alpha$	*iff* $\exists j \in \mathbb{L}\ M, r, n[\![\alpha]\!]j$
$M, r, n \models \mathsf{BEL}\ \varphi$	*iff* $\forall r', n'(B(r, n, r', n') \Rightarrow M, r', n' \models \varphi)$
$M, r, n \models \mathsf{GOAL}\ \varphi$	*iff* $\forall r' n'(G(r, n, r', n') \Rightarrow M, r', n' \models \varphi)$
$M, r, n \models \mathsf{BEFORE}\ \varphi\ \psi$	*iff* $\forall j \geq n(M, r, j \models \psi \Rightarrow \exists i \leq j(M, r, i \models \varphi))$
$M, r, n \models E\varphi$	*iff* $\exists r', n'\ \ M, r', n' \models \varphi$

where $M, r, n[\![\alpha]\!]n'$, to interpret $\mathsf{DONE}\ \alpha$ *and* $\mathsf{HAPPENS}\ \alpha$*, is defined as follows:*

1. $M, r, n[\![\alpha]\!]n'$ *iff* $r^{ac}_n = \alpha$ *and* $n' = n + 1$.
2. $M, r, n[\![\mathsf{IF}\ \varphi\ \mathsf{THEN}\ \alpha\ \mathsf{ELSE\ NIL}]\!]n'$ *iff* $M, r, n \models \varphi \Rightarrow M, r, n[\![\alpha]\!]n'$.

Note in particular the definition of semantics of the global modality E, which is an extension of Intention Logic: this operator allows inspection of arbitrary states within a model, which is useful to translate the dynamic operator $[\alpha]\chi$ of GOAL Logic into Intention Logic.

3.2 CL Models for Intention Logic

How do our Run-Based Models (RBM, from now) compare to the Cohen & Levesque Models, as presented in [3] (CLM, henceforth)?

Observation 1. *The following relates* RBM *with* CLM*:*

1. CLM *models are a special case of* RBM *models in the following sense: In* CLM *models,*
   (a) $\mathbb{L}$ *is taken to be* $\mathbb{Z}$
   (b) agents know the correct time: *If* $B(r, n, r', n')$ *then* $n = n'$
   (c) agents "want" the current time: *If* $G(r, n, r', n')$ *then* $n = n'$
   (d) G *and* B *are related through* realism*:* $G \subseteq B$
   (e) a run is of type $\mathbb{L} \to A$

(f) runs are determined by their action part, i.e.,

$$\forall r, r'(\forall n : r_n^{ac} = r_n'^{ac} \Rightarrow \forall n\ V(r_n^{st}) = V(r_n'^{st})$$

(g) agents remember the last atomic action they have done: if $B(r, n, r', n')$ *then* $n = n'$ *and* $r_n^{ac} = r_n'^{ac}$.

(h) assume the property of No persistence / deferral forever*, see below.*

2. *However,* RBM *models are also a specialisation of* CLM *models:*
   (a) CLM *allows for quantification over a domain of objects and events;*
   (b) CLM *models have a richer notion of composed actions, and accordingly an extended definition of* $M, r, n[\![\alpha]\!]n'$.
   (c) CLM *models are defined for multiple agents.*

Some of the differences mentioned above are merely a matter of choice or design. For instance, it is straightforward to extend the notion of Run-Based Model in such a way that they encompass item 2(b,c) of Observation 1. As regarding item 1, there are some deeper issues involved. As to 1a, it seems natural for computational systems to assume that computations have a start somewhere. Syntactically, item 1a amounts to the requirement that there is always some atomic action α for which $\mathsf{DONE}\ \alpha$ holds.[3] To assume that agents know the correct time (1b) makes sense in many scenario's, and, given that an agent knows the time, it does not make sense to have a "goal" that the time were different. Where realism of CLM ensures $\mathsf{BEL}\ \varphi \rightarrow \mathsf{GOAL}\ \varphi$, the weak realism of RBM amounts to $\mathsf{BEL}\ \varphi \rightarrow \neg\mathsf{GOAL}\ \neg\varphi$. We don't think realism is a very realistic(!) assumption, and we even think that Cohen and Levesque had *weak realism* in mind when they presented their semantics ([3][p. 227]):

> *. . . 'the worlds that are consistent with what the agent has chosen are not ruled out by his beliefs. Without this constraint, the agent could choose world involving (for example) future events that he believes will never happen.'*

Hence, we will assume that R and G satisfy *weak realism*: for every $r \in \mathcal{R}$, and $n \in \mathbb{L}$, there is a $r' \in \mathcal{R}$ and $n' \in \mathbb{L}$ such that $(r, n, r', n') \in G \cap B$.

Let us now consider item 1f, which is related to item 1e which restricts runs to $\mathbb{L} \rightarrow A$. Suppose we take runs as basic entities, like in CLM. This does not do justice to the intensional notion of the logic, as can be seen as follows. Suppose that we have only one atom $p \in \mathsf{Atoms}$, and two basic actions α, α'. Let $\widehat{\mathsf{BEL}}\ \varphi$ be $\neg\mathsf{BEL}\ \neg\varphi$: the agent considers φ doxastically possible. Let ψ be $\mathbf{0} \wedge \Box(p \wedge \mathsf{DONE}\ \alpha)$. Now consider $\widehat{\mathsf{BEL}}\ (\psi \wedge \mathsf{GOAL}\ p) \wedge \widehat{\mathsf{BEL}}\ (\psi \wedge \neg\mathsf{GOAL}\ p)$. This is not satisfiable in CLM, since ψ determines a unique run, and what the goals and beliefs of an agent are is determined by the run. More natuaral examples present themselves in the multi-agent case, where we would have for instance that $\widehat{\mathsf{BEL}}_1(\psi \wedge \mathsf{BEL}_2 p) \wedge \widehat{\mathsf{BEL}}_1(\psi \wedge \neg\mathsf{BEL}_2 p)$ is unsatisfiable[4].

[3] Since we do not have quantification over events in the propositional version of IL, we assume that all transitions are labeled with an atomic action.

[4] for readers familiar with modal epistemic logic, this is exactly the reason why states are not identified with valuations: there would not be enough valuations (in case of one atom) to satisfy $\neg K_1 \neg(p \wedge K_2 p) \wedge \neg K_1 \neg(p \wedge \neg K_2 p)$

Given that CLM models identify runs and paths, and a run in CLM is of type $\mathbb{L} \to A$, already brings a problem to the fore that is more basic than on the intensional level. In CLM, a valuation Φ checks whether $\Phi(p, \sigma, n)$ holds, where p is an atomic proposition, σ is a 'event-run': $\mathbb{Z} \to A$ and $n \in \mathbb{Z}$. But this implies that the truth of atomic propositions (and hence, of objective formulas) is compeletely determined once we know which actions are taken along σ. In other words, it is not possible to have two event-runs that agree on all the actions, but still objective formulas along them differ. Suppose the event α represents the throwing of a dice, ($\sigma(n) = \alpha$, for all n) and that $p_i (i \leq 6)$ represents the outcome. Now, let Φ determine how the propositions p_i are distributed over σ, say $\Phi(p_i, \sigma, n)$ iff $i = n \bmod 6$. Now, the type of Φ dictates that there *cannot* be another event run σ' in which a dice is continuously thrown ***but the outcomes are different!*** In particular, this implies that if our agent knows that α always happened and will always happen, he will also know all the outcomes (there is no alternative run with the same actions and different outcomes). Summarising (let G^{-1} refer to the past):

$$\mathsf{KNOW}\ (\Box\ \mathsf{HAPPENS}\ \alpha \wedge \mathsf{G}^{-1}\mathsf{DONE}\ \alpha) \to \mathsf{KNOWIF}\ \varphi \tag{1}$$

Property 1g of Observation 1 is implicitly imposed by [3] since they require

$$[3, \textbf{Assumption 3.20}]\ \models (\mathsf{DONE}\ \alpha) \leftrightarrow (\mathsf{BEL}\ (\mathsf{DONE}\ \alpha))$$

Let us now look at 1h. This is the semantic counterpart of another assumption made in [3], motivated by the fact that an agent should not endlessly pursue the same goal:

$$[3, \textbf{Assumption 3.25}]\ \models \Diamond\neg(\mathsf{GOAL}\ (\neg\varphi \wedge \Diamond\varphi))$$

Writing $\widehat{\mathsf{GOAL}}\ \varphi$ for $\neg\mathsf{GOAL}\ \neg\varphi$, it is not hard to see that this is equivalent to:

$$\models \Diamond\ \widehat{\mathsf{GOAL}}\ (\varphi \to \Box\varphi) \tag{2}$$

However, (2) *as a scheme* corresponds, in the sense of modal logic [2] to a semantic property that is incompatible with the models we are currently looking at. Note that for $\varphi \to \Box\varphi$ to be true in a world x corresponds to the fact that $\forall y(x \leq y \to y = x)$ (there is at most one instance that is later than x, and this is x itself. Then for $\Diamond\ \widehat{\mathsf{GOAL}}\ (\varphi \to \Box\varphi)$ to be true in all worlds z corresponds to

$$\forall z \exists u \exists x(z \leq u\ \&\ (Gux\ \&\ \forall y(x \leq y \to y = x))) \tag{3}$$

In words: for every time point, there is a future time point with a GOAL-accessible point, such that the latter point only has itself as a future successor. This property is incompatible with our models (and indeed, with CLM models), since (1) time is supposed to go on forever, and (2) we have 'nominals' that are true at only one time point: in the x state above, some time expression $\mathbf{x}$ must be only in x itself, and not its successors.

[3, **Assumption 3.25**] expresses that 'there is a future point such that in some goal-accessible world, no goal is true anymore', while the intuition [3] seem to want to capture is 'for every goal there will be a time point in the future that it is dropped'. The latter seems hard to be conceived of as a structural property on models, and indeed, we think it should be a property of the protocol, or behaviour, of the agent.

Summarising, for our semantics, we assume time has a starting point, and that agents know and want the time. The other restrictions 1d - 1h are either properties that give undesired properties (1d, 1e, 1f, 1h), or can be added on top of a basic class of models (1g).

Definition 10 (Run-Based IL-Models). *The class of Run-Based Basic Intention Logic Models,* RBBILM *for short, is the class of Run-Based Models* $M = \langle S, \mathbb{L}, B, G, V \rangle$ *such that:*

1. $\mathbb{L} = \mathbb{Z}$
2. *agents know the correct time*
3. *agents want the correct time*
4. B *and* G *are connected through weak realism.*

Validity in the class RBBILM *is denoted* $\models_{BI}$.

4 Connecting GOAL and Intention Logic

In this Section we show how to formally relate GOAL and Intention Logic. First, we define a translation function from GOAL into Intention Logic. Except for the goal operator and the dynamic modality of GOAL Logic this is straightforward. The main result we want to prove is that properties proven to hold in one logic are preserved under translation from that logic to the other. We do so by showing that satisfaction of a formula is preserved under translation.

Definition 11. (Translating $\mathcal{L}_G$ into $\mathcal{L}_{BI}$)
The translation function τ *mapping GOAL Logic formulae and action rules onto Intention Logic formulae is defined by:*

$$
\begin{array}{ll}
\tau(\mathbf{start}) & = \mathbf{0}, \\
\tau(\mathbf{B}\phi) & = \mathsf{BEL}\ \phi, \\
\tau(\mathbf{G}\phi) & = \mathsf{GOAL}\ \Diamond\phi \wedge \neg\mathsf{BEL}\ \phi, \\
\tau(\neg\chi) & = \neg\tau(\chi), \\
\tau(\chi_1 \wedge \chi_2) & = \tau(\chi_1) \wedge \tau(\chi_2), \\
\tau(\chi_1\ \mathbf{until}\ \chi_2) & = \tau(\chi_1)\mathsf{UNTIL}\ \tau(\chi_2), \\
\tau([\alpha]\chi) & = U(\mathsf{DONE}\ \alpha \rightarrow \tau(\chi)), \\
\tau(\mathbf{if}\ \psi\ \mathbf{then}\ \alpha) & = \mathsf{IF}\ \tau(\psi)\ \mathsf{THEN}\ \alpha\ \mathsf{ELSE\ NIL}.
\end{array}
$$

The most interesting case in the definition of the translation function τ is the translation of $\mathbf{G}\phi$. An achievement goal in GOAL requires that the agent does not believe ϕ to be the case, whereas [3] require the agent to believe that ϕ is not the case. Whereas the goal operator $\mathbf{G}$ does not satsify axiom D (cf. [4]; see also [7] for a discussion), the achievement goal operator of [3] does, implying that an agent cannot have inconsistent achievement goals.

The proof showing that satisfaction is preserved under translation is based on model constructions. Lemma 1 shows how to derive a GOAL Logic model (a trace) from an RBBILM model that preserves satisfaction of formulae from GOAL Logic, whereas

Lemma 2 shows how to construct an RBBILM model from a GOAL trace. Theorem 2 states our main result that satisfaction is preserved under translation, which shows that Basic Intention Logic can be used to prove properties of GOAL agents.

Lemma 1. *Let $M = \langle S, \mathbb{L}, B, G, V\rangle$ be an* RBBILM *model. Then there is a GOAL agent $\mathcal{A}$ and a function f from runs to traces such that the set of traces in $\mathcal{A}$ is $\{f(r) \mid r \in \mathcal{R}(S, A)\}$ and for all $\varphi \in \mathcal{L}_G$:*

$$M, r, n \models_{BI} \tau(\varphi) \textit{ iff } \mathcal{A}, f(r), n \models_G \varphi$$

Proof. We need to construct a GOAL trace $f(r) = t = m_0, \alpha_0, m_1, \alpha_1, \ldots$ for every run $r \in \mathcal{R}$, where each mental state m_i is of the form $\langle \Sigma_i, \Gamma_i\rangle$. The components can be derived from r as follows:

- $\Sigma_i = \{\phi \in \mathcal{L}_0 \mid M, r, i \models_{BI} \mathsf{BEL}\ \phi\}$,
- $\Gamma_i = \{\phi \in \mathcal{L}_0 \mid M, r, i \models_{BI} \mathsf{GOAL}\ \Diamond\phi \wedge \neg\mathsf{BEL}\ \phi\}$, and
- $\alpha_i = r_i^{ac}$.

Since the relation B in RBBILM models is serial, each Σ_i is consistent. For a similar reason every $\gamma \in \Gamma_i$ is consistent. Moreover, by constuction of Γ_i, we have $\forall\gamma \in \Gamma_i, \Sigma \not\models \gamma$. We now show the equivalence of $\tau(\varphi)$ in M, r, n with that of $f(r), n$ in $\mathcal{A}$ by induction on φ as follows.

If φ is $\mathbf{start}$, we have $M, r, n \models_{BI} \mathbf{0}$ iff $n = 0$ iff $\mathcal{A}, f(r), 0 \models_G \mathbf{start}$. For the intensional operators, the equivalence follows immediately from the definition of mental states in the trace $f(r)$. Finally, let $\varphi = [\alpha]\chi$. Then $M, r, n \models_{BI} U((\mathsf{DONE}\ \alpha) \rightarrow \tau(\chi))$ iff for every run r' and n', we have $M, r', n' \models_{BI} \mathsf{DONE}\ \alpha \rightarrow \tau(\chi)$. Now let t' be an arbitrary trace in $\mathcal{A}$, and suppose $t_i'^{a} = \alpha$. Obviously, this trace must be the image of a run r' for which $r_i'^{ac} = \alpha_i$. But then, $M, r', i+1 \models_{BI} \mathsf{DONE}\ \alpha$ and, hence $M, r', i+1 \models_{BI} \tau(\chi)$. By induction, $\mathcal{A}, t', i+1 \models_G \chi$. This demonstrates $\mathcal{A}, t, n \models_G [\alpha]\chi$. The other direction is similar. □

Lemma 2. *Let $\mathcal{A}$ be an agent, that is, a set of traces. Then we can construct an* RBBILM *model $M = \langle S, \mathbb{L}, B, G, V\rangle$ such that there is a function $g : \mathcal{A} \rightarrow \mathcal{R}$ satisfying, for every $\varphi \in \mathcal{L}_G$ and every $n \in \mathbb{N}$:*

$$\mathcal{A}, t, n \models_G \varphi \textit{ iff } M, g(t), n \models_{BI} \tau(\varphi)$$

Proof. Let $Constraints = \{[\alpha]\chi \mid \mathcal{A} \models_G [\alpha]\chi\}$. Let ε be an action symbol not occurring in $\mathcal{L}_G$. Call a run r *minimal* if for all n, $V(r_n^{st}) = \emptyset$ and $r_n^{ac} = \varepsilon$. Call a run r *peak-once* if it is like a minimal run, except that for at most one $k \in \mathbb{N}$, we can have $V(r_k^{st}) \neq \emptyset$. Given a trace t, we have to find its associated run $g(t)$. Let $t = m_0, \alpha_0, m_1, \alpha_1, \ldots$. For the run $g(t)$, we put $g(t)_i^{ac} = \alpha_i$. Let the mental state at t_i be $\langle \Sigma_i, \Gamma_i\rangle$. For every valuation π for which $\pi \models \Sigma_i$, add a state $\langle t, i, \pi\rangle$. Put $V(\langle t, i, \pi\rangle) = \pi$ and, for every such state $\langle t, i, \pi\rangle$, add a peak-once run r' such that $r_i'^{st} = \langle t, i, \pi\rangle$. Put $B(g(t), i)(r', i)$ for each such run. This procedure guarantees that

$$\text{For all } \phi \;\; M, g(t), i \models_{BI} \mathsf{BEL}\ \phi \text{ iff } \phi \in \Sigma_i \tag{4}$$

For the goals Γ_i, we distinguish two cases. First, suppose $\Gamma_i = \emptyset$. Then, for the goal-associated runs in $g(t), i$ we take exactly the belief-associated runs as described above. Apart from weak realism, this guarantees

$$\Gamma_i = \emptyset \Leftrightarrow \text{ for no } \phi : M, g(t), i \models_{BI} \mathsf{GOAL}\ \Diamond\phi \wedge \neg \mathsf{BEL}\ \phi \tag{5}$$

Now, if $\Gamma_i \neq \emptyset$, let $\gamma_1, \gamma_2, \ldots$ be an infinite enumeration of all elementes of Γ_i: if Γ_i only has a finite number h of elements, we put $\gamma_{h+j} = \gamma_j$. Since each γ_j is consistent, it comes with a set of propositional valuations Π_j. Let k be the biggest cardinality of those sets Π_j, which could be an element of $\mathbb{N}$ or else ∞. Now we associate k goal-accessible runs r with $g(t), i$ such that for every $m, m' > i$, $V(r^{st}_m), V(r^{st}_{m'})$ are valuations from Π_m, whenever $\gamma_m = \gamma_{m'}$ then $V(r^{st}_m) = V(r^{st}_{m'})$, and, conversely, every valuation in Π_m occurs in at least one goal-accessible run. Since the language $\mathcal{L}_G$ cannot talk about the past, it does not matter how such a run looks like at $j \leq i$, although, in order to obtain weak realism, we take care that there is at least one of the goal-runs r_g just created for which $r_g(i) = r_b(i)$, where r_b is one of the belief-accessible runs. We finally specify $r^{ac}_n = \varepsilon$ for all n, for all such runs r. Since we know that $\langle \Sigma_i, \Gamma_i \rangle \models \mathbf{G}\phi$ implies that $\phi \notin \Sigma_i$, this procedure guarantees that

$$\text{For all } \phi\ \phi \in \Gamma_i \text{ iff } M, g(t), i \models_{BI} \mathsf{GOAL}\ \Diamond\phi \wedge \neg \mathsf{BEL}\ \phi \tag{6}$$

Now, the model M is built by taking all runs $g(t)$ from $t \in \mathcal{A}$, and adding the associated goal and belief runs (the states that we need are defined when we defined the runs).

The proof of the overall claim again follows using induction on φ, where the intensional operators follow directly from (4), (5) and (6). The only interesting remaining case are $Constraints$. So let us consider $\varphi = [\alpha]\chi$, and the property proven for χ. Suppose furthermore $\mathcal{A}, t, n \models_G [\alpha]\chi$. This means that for all t' and m that $t'^a_m = \alpha \Rightarrow \mathcal{A}, t', m+1 \models_G \chi$. In M, the only runs r for which there is an i such that $M, r, i+1 \models_{BI} \mathsf{DONE}\ \alpha$ holds, are runs for which there is a trace t such that $r = g(t)$ and in t, α_i equals α (since the constructed goal and belief runs only refer to action ε). But using induction, we have $M, g(t), m+1 \models_{BI} \tau(\chi)$, which completes the proof. □

Theorem 2. *GOAL semantics $\models_G$ and semantics of Run-Based Basic Intention Logic $\models_{BI}$ are equivalent for the $\mathcal{L}_G$ and $\tau(\mathcal{L}_G)$.*

Proof. Imediate from Lemma 1 and 2. □

5 Extending GOAL agents with Temporally Extended Goals

The mapping of goals in the GOAL language onto Intention Logic as in Definition 11 shows that these are naturally interpreted as *achievement goals*, as originally intended [4, 8]. The future-directed interpretation of such goals is left implicit in GOAL whereas it is made explicit in the definition of such goals in Intention Logic. By making the temporal component explicit it is straightforward to define other goal types in Intention Logic. For example, *maintenance goals* can be defined as $\mathsf{GOAL}\ (\Box\phi)$. The idea to introduce a primitive "goal" operator GOAL (or $Choice$ as [7] call it) in Intention Logic

that allows defining various goal types can be introduced in GOAL as well to increase expressivity [9]. In this Section we show how we can apply the result of the previous Section to extend GOAL with *temporally extended goals* [1] while still maintaining the connection between GOAL Logic and Intention Logic.

To this end, we now allow *pure* temporal formulae φ in the belief and goal base of GOAL agents. As the idea is to define achievement and other goals now in GOAL in the same way as in Intention Logic, the semantics of the goal operator $\mathbf{G}$ in GOAL is modified analogously, and is now simply defined as:

$$\langle \Sigma, \Gamma \rangle \models \mathbf{B}\phi \text{ iff } \Sigma \models_{LTL} \phi,$$
$$\langle \Sigma, \Gamma \rangle \models \mathbf{G}\phi \text{ iff } \Gamma \models_{LTL} \phi.$$

As we now allow temporal formulae ϕ without occurrences of other modal operators in both the belief and goal base, the entailment relation of linear temporal logic is used [5]. It is clear that with these operators we can reintroduce the notion of an achievement goal by definition as $\mathbf{G}\Diamond\phi \wedge \neg\mathbf{B}\phi$. Moreover we no longer require as in Definition 1 that individual goals in a goal base Γ are consistent (this is now taken care of by the temporal operators) but instead require that Γ itself is consistent. A further simplification as a result of this modified setup is that the rationality constraint of Definition 1 that goals are not believed to be the case is no longer needed as this now follows by definition.[5]

It turns out that to show that the connection with Intention Logic is maintained requires only minor modifications of the proofs provided in Section 4 and actually simplifies matters somewhat. The proof of Lemma 1 only requires a modification of the derivation of mental states from a run r, as follows:

- $\Sigma_i = \{\phi \in \mathcal{L}_{LTL} \mid M, r, i \models_{BI} \mathsf{BEL}\ \phi\}$,
- $\Gamma_i = \{\phi \in \mathcal{L}_{LTL} \mid M, r, i \models_{BI} \mathsf{GOAL}\ \phi\}$.

As for Lemma 2, since in the new setup (see definitions above) belief and goal bases have the same logical properties there is no need anymore to distinguish them in the proof. It thus suffices to show how to construct a run r such that we have (4^*) $M, g(t), i \models_{BI} \mathsf{BEL}\ \phi$ iff $\Sigma_i \models_{LTL} \phi$ (cf. Lemma 2). As before, for a given trace t we have to find an associated run $g(t)$. Call a run r *silent* if it consists of ϵ-steps only, i.e. $r^{ac}_n = \epsilon$ for all n. Then put $B(g(t), i, r, i)$ for each silent run such that $M, r, i \models \Sigma_i$. This procedure guarantees (4^*). The same procedure can be used to prove (5^*) $M, g(t), i \models_{BI} \mathsf{GOAL}\ \phi$ iff $\Gamma_i \models_{LTL} \phi$, and we are done. Finally, by changing the translation mapping of Definition 11 for $\mathbf{G}\phi$ to $\mathsf{GOAL}\ \phi$ we obtain:

Theorem 3. *The GOAL semantics $\models_G$ and semantics of Run-Based Basic Intention Logic $\models_{BI}$ are equivalent for the languages $\mathcal{L}_G^{LTL}$ and $\tau(\mathcal{L}_G^{LTL})$ that include temporally extended goals and beliefs.*

[5] There remains however the problem of how and when to remove goals from the goal base of an agent. In [9] a progression operator has been introduced as a solution to this problem (see also [1]). In the setup of this Section the main difference between the belief and goal base is this automatic mechanism of removing goals from the goal base, which represents the default *commitment strategy* of an agent (cf. [4, 8, 11]).

6 Conclusion

We showed that GOAL agents instantiate Intention Logic and can be formally related by means of translating GOAL Logic into Intention Logic. Two important results follow: (i) GOAL Logic is equivalent to a propositional fragment of Intention Logic, and (ii) this fragment - a standard normal tense logic - can be used to *verify* GOAL agents using a Hoare logic for actions performed by GOAL agents (using additional derivation rules for verification introduced in [4]). The result proved useful for incorporating temporally extended goals into GOAL while maintaining the connection with Intention Logic.

We argued that Intention Logic at a number of points needs revision. In particular, we argued that the principle of No Persistence Forever that requires an agent to drop every one of its goals sometime is too strong. Moreover, the notion of achievement goals used in GOAL is slightly different from that of [3] and more in line with that proposed in [7].

Future work will involve applying our results in model checking of GOAL agents. Conceptually we are interested in including preferences into the language while maintaining a logical connection with a standard modal logic, which involves extensions to the programming language GOAL [9] as well as to Basic Intention Logic. The additional expressivity introduced by incorporating temporally extended goals and temporal formulae into the belief base of GOAL agents also raises many new questions about goal persistence and the operationalization of, for example, maintenance goals [9].

References

1. F. Bacchus and F. Kabanza. Planning for temporally extended goals. *Annals of Mathematics and Artificial Intelligence*, 22:5–27, 1998.
2. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.
3. P.R. Cohen and H.J. Levesque. Intention Is Choice with Commitment. *Artificial Intelligence*, 42:213–261, 1990.
4. F. de Boer, K. Hindriks, W. van der Hoek, and J-J.Ch. Meyer. A Verification Framework for Agent Programming with Declarative Goals. *Journal of Applied Logic*, 5(2):277–302, 2007.
5. E.A. Emerson. Temporal and Modal Mogic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B. North-Holland Publishing Company, Amsterdam, 1990.
6. D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.
7. A. Herzig and D. Longin. C&l intention revisited. In *Proc. of the 9th Int. Conference Principles of Knowledge Representation and Reasoning (KR'04)*, pages 527–535, 2004.
8. K.V. Hindriks, F.S. de Boer, W. van der Hoek, and J-J.Ch. Meyer. Agent Programming with Declarative Goals. In *Proc. of the 7th Int. Workshop on Intelligent Agents VII (ATAL'00)*, pages 228–243, 2000.
9. K.V. Hindriks and B. van Riemsdijk. Using temporal logic to integrate goals and qualitative preferences into agent programming. In *Proceedings of the International Workshop on Declarative Agent Languages and Theories*, 2008. accepted.
10. J-J.Ch. Meyer. Our quest for the holy grail of agent verification. In *Automated Reasoning with Analytic Tableaux and Related Methods*, pages 2–9, 2007.
11. A.S. Rao and M.P. Georgeff. Intentions and Rational Commitment. Technical Report 8, Australian Artificial Intelligence Institute, 1993.
12. W. van der Hoek and M. Wooldridge. Towards a Logic of Rational Agency. *Logic Journal of the IGPL*, 11(2):133–157, 2003.