Formalising Abilities and Opportunities of Agents

B. van Linder[*] W. van der Hoek J.-J. Ch. Meyer
Utrecht University
Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht
The Netherlands
wiebe@cs.ruu.nl

Abstract

We present a formal system to reason about and specify the behavior of multiple intelligent artificial agents. Essentially, each agent can perform certain actions, and it may possess a variety of information in order to reason about its and other agent's actions. Thus, our KARO-framework tries to deal formally with the notion of *Knowledge,* possessed by the agents, and their possible execution of actions. In particular, each agent may reason about its —or, alternatively, other's— *Abilities* to perform certain actions, the possible *Results* of such an execution and the availability of the *Opportunities* to take a particular action. Formally, we combine dynamic and epistemic logic into one modal system, and add the notion of ability to it. We demonstrate that there are several options to define the ability to perform a sequentially composed action, and we outline several properties under two alternative choices. Also, the agents' views on the correctness and feasibility of their plans are highlighted. Finally, the complications in the completeness proof for both systems indicate that the presence of abilities in the logic makes the use of infinite proof rules useful, if not inevitable.

1 Introduction

The last ten years have witnessed an intense flowering of interest in artificial agents, both on a theoretical and on a practical level. The ACM devoted a special issue of its 'Communications' to intelligent agents [6], and Scientific American ranked intelligent software agents among the key technologies for the 21st century [35]. Also various conferences and workshops were initiated that specifically address agents, their theories, languages, architectures and applications [8, 28, 49, 50]. Consequently, terms like agent-based computing, agent-based software engineering and agent-oriented programming have become widely used in research on AI. Despite its wide use, there is no agreement on what the term 'agent' means. Riecken remarks that 'at best, there appears to be a rich set of emerging views' and that 'the terminology is a bit messy' [42]. Existing definitions range from 'any entity whose state is viewed as consisting of mental objects ' [46] and 'autonomous objects with the capacity to learn, memorize and communicate' [9], to 'systems whose behavior is neither casual nor strictly causal, but teleonomic, goal-oriented toward a certain state of the world' [3]. Other authors, and truly not the least, use the term 'robot' instead of agent [27], or take the common-sense definition of

[*] Currently at ABN-AMRO, Amsterdam.

agents for granted [40]. In practical applications agents are 'personal assistant[s] who [are] collaborating with the user in the same work environment' [34], or 'computer programs that simulate a human relationship, by doing something that another person could otherwise do for you' [45].

The informal description of an (artificial) agent in its most primitive form, which we distill from the definitions given above and which the reader is advised to keep at the back of his/her mind throughout reading this paper, is that of an entity which has the possibility to execute certain *actions*, and is in the possession of certain *information*, which allows it to *reason* about its own and other agents' actions. In general, these agents will also have motives that explain why they act the way they do. The treatment of these motives is however not the subject of this paper (but see [32]). Moreover, although we borrow a lot of terminology and notions from philosophy, the reader should keep in mind that it is our main goal to describe artificial agents, rather than humans.

Currently several applications of agent-technology are in use. Among those listed by Wooldridge & Jennings [48] are air-traffic control systems, spacecraft control, telecommunications network management and particle acceleration control. Furthermore, interface agents are used that for instance take care of email administration, as well as information agents that deal with information management and retrieval. In all probability, these implemented agents will be rather complex. In addition, life-critical implementations like air-traffic control systems and spacecraft control systems need to be highly reliable. To guarantee reliability it is probably necessary to use formal methods in the development of these agent systems, since such a guarantee can never be given by just performing tests on the systems. Besides this general reason for using formal techniques in any branch of AI and computer science, there is another reason when dealing with agents. These agents will in general be equipped with features representing common-sense concepts as knowledge, belief and ability. Since most people do have their own conception of these concepts, it is very important to unambiguously establish what is meant by these concepts when ascribed to some specific implemented agent. Formal specifications allow for such an unambiguous definition.

The formal tool that we propose to model agency is *modal logic* [4, 20, 21]. Using modal logics offers a number of advantages. Firstly, using an intensional logic like modal logic allows one to come up with an intuitively acceptable formalisation of intensional notions with much less effort than it would take to do something similar using fully-fledged first-order logic. Secondly, the reducibility of modal logic to (fragments of) first-order logic ensures that methods and techniques developed for first-order logic are still applicable to modal logic. Lastly, using possible worlds models as originally proposed by Kripke [25], provides for a uniform, clear, intelligible, and intuitively acceptable means to give mathematical meaning to a variety of modal operators. The modal systems that we propose to formalise agents belong to what we call the *KARO-framework*. In this framework, the name of which is inspired by the well-known BDI-architecture [40], special attention is paid to the agents' *knowledge* and *abilities*, and to the *results* of and *opportunities* for their actions. We present two different systems, both belonging to the KARO-framework, that differ in their treatment of abilities for certain actions. To show the expressive power of the framework we formalise various notions that are interesting maybe from a philosophical point of view, but that above all should help to understand and model artificially intelligent agents—we like to stress that our aim is to describe artificial agents like softbots and robots by means of these notions, rather than human agents, which are far more complex and for which one probably needs more complicated descriptions.

The agent attitudes in this paper are limited to knowledge, and abilities, results and opportunities with respect to his/its actions, i.e. the KARO framework. We stress that the purpose of this paper is to give a thorough treatment of this KARO framework, which has been used by us as a basis for a much more extensive description of agents, incorporating such notions as observations ([30]), communication ([29]), default reasoning ([33]), belief revision ([31]), and goals ([32]).

The philosophy adhered to in this endeavour is that the primary attitude of agents is to *act*, by the very meaning of the word 'agent', so that a specification logic for the behaviour of agents should start out from a logic of action (for which we have chosen an extension of dynamic logic in which knowledge and ability is expressible as well). In this enterprise we owe to Bob Moore's work combining a version of dynamic logic and epistemic logic for the first time [37] So here we deviate from the philosophy of other foundational work on agents, in particular that of Rao & Georgeff [40, 39, 41], who take belief, desire, intentions as well as time as primitive notions for agents. (One could argue that our approach is more in line with that of Cohen & Levesque [5]. They, too, take actions as basic building blocks for their theory of agents. However, they consider only *models* of their framework and provide no formal proof system. Furthermore, they are mainly concerned with the formalisation of motivational attitudes such as goals and intentions of agents, and employ actions merely as a basis to obtain this, while here we are interested in actions and aspects of these, such as opportunities and abilities, in their own right.)

Therefore, the main contribution of this paper is to investigate the KARO logic and provide meta-results such as completeness, which, particularly by the addition of abilities, will turn out to be a non-trivial extension of that for basic dynamic logic.

Organisation of the paper The rest of the paper is organised as follows. In Section 2 we look at the philosophical foundations of the KARO-framework. In Section 3 we present the formal definitions constituting the two systems belonging to the KARO-framework. We start by defining the language common to the two systems, where-after a common class of models and two different interpretations for the formulas from the language in the models are presented. In Section 4 various properties of knowledge and action in the KARO-framework are considered. In Section 5 we consider the notion of practical possibility, and formalise part of the reasoning of agents on the correctness and feasibility of their plans. In Section 6 we present two slightly different proof systems that are sound and complete with respect to the notions of validity associated with the two interpretations. Section 7 concludes this paper with a brief summary, an overview of related work, and some suggestions for future research. In the appendix we present the proofs of soundness and completeness in considerable detail.

2 The KARO-framework from a philosophical perspective

As mentioned in the previous section, in its simplest form an agent is an entity that performs actions and possesses information. The informational attitude that we equip our agents with is termed *knowledge*. Our use of the term knowledge agrees with the common one in AI and computer science [14, 36], i.e. knowledge is veridical information with respect to which the agent satisfies conditions of both positive and negative introspection. Veridicality implies that only true formulas are known by agents, positive introspection states that agents know that they know something whenever they know it, and negative introspection states that agents know that they do not know something as soon as they do not know it.

To explain the concept of action, we first have to spend some words on the ontology of states of affairs that we presuppose. By a *state of affairs* we mean the way the world is, or one way it might be, at a moment. The (currently) actual state of affairs is composed of the facts about the world as it actually is at this very moment. But there are, presumably, various other states of affairs which could have applied to the world as this moment instead, or that would apply as the result of a given action. An agent, with limited knowledge of the facts, might consider various merely hypothetical states of affairs consistent with the agent's knowledge *Actions* are now considered to be descriptions of causal processes, which upon execution by an agent may turn one state of affairs into another one. Thus, our intuitive idea of actions corresponds to what Von Wright calls the *generic* view on actions [52]. An *event* consists of the performance of a particular action by a particular agent, and is as such related to Von Wright's *individual* view on actions [52]. We will use the plain term events, although perhaps the term *agent-driven event* would be more appropriate, here. Given the ontology of actions and events as somehow causing transitions between states of affairs, we deem two aspects of these notions to be crucial: when is it possible for an agent to perform an action, and what are the effects of the event consisting of the performance by a particular agent of a particular action in a particular state of affairs? To investigate these questions we focus on three aspects of actions and events that are in our opinion essential, viz. *result*, *opportunity* and *ability*. Slightly simplifying ideas of Von Wright [51], we consider any aspect of the state of affairs brought about by the occurrence of an event in some state of affairs to be among the results of that particular event in that particular state of affairs. In adopting this description of results we abstract from all kinds of aspects of results that would probably have to be dealt with in order to come up with an account that is completely acceptable from a philosophical point of view, such as for instance the question whether all changes in a state of affairs have to be ascribed to the occurrence of some event, thereby excluding the possibility of external factors influencing these changes. However, it is not our aim to provide a complete theory of results incorporating all these aspects, but instead combine results with other notions that are important for agency. From this point of view it seems that our definition of results is adequate to investigate the effects of actions, and, given the complexity already associated with this simple definition, it does not make much sense to pursue even more complex ones.

Along with the notion of the result of events, the notions of ability and opportunity are among the most discussed and investigated in analytical philosophy. Ability plays an important part in various philosophical theories, as for instance the theory of free will and determinism, the theory of refraining and seeing-to-it, and deontic theories. Following Kenny [22], we consider ability to be the complex of physical, mental and moral capacities, internal to an agent, and being a positive explanatory factor in accounting for the agent's performing an action. Opportunity on the other hand is best described as circumstantial possibility, i.e. possibility by virtue of the circumstances. The opportunity to perform some action is external to the agent and is often no more than the absence of circumstances that would prevent or interfere with the performance. Although essentially different, abilities and opportunities are interconnected in that abilities can be exercised only when opportunities for their exercise present themselves, and opportunities can be taken only by those who have the appropriate abilities. From this point of view it is important to remark that abilities are understood to be *reliable* (cf. [2]), i.e. having the ability to perform a certain action suffices to take the opportunity to perform the action every time it presents itself. The combination of ability and opportunity determines whether or not an agent has the (practical) possibility to perform an action.

For our artificial agents, we will in Section 5 study 'correctness' of α for i to bring about ϕ in terms of having the opportunity: for the artificial agents that we have in mind and the actions ('programs') they perform correctness has to do with intrinsic (rather than external) features of the action: its halting and the intended outcome. Such features are still beyond the scope of the agent's abilities, of course.

3 The KARO-framework from a formal perspective

For the reasons already given in Section 1, we propose the use of a propositional multi-modal language to formalise the knowledge and abilities of agents, and the results of and opportunities for their actions. In contrast with most philosophical accounts, but firmly in the tradition of theoretical computer science, this language is an *exogenous* one, i.e. actions are represented explicitly. Although it is certainly possible to come up with accounts of action without representing actions (see for instance [38, 44]), we are convinced that many problems that plague these endogenous formalisations can be avoided in exogenous ones.

The language contains modal operators to represent the knowledge of agents as well as to represent the result and opportunity of events. The ability of agents is formalised by a factually non-modal operator. Following the representation of Hintikka [17] we use the operator $\mathbf{K}_{\_}\_$ to refer to the agents' knowledge: $\mathbf{K}_i\varphi$ denotes the fact that agent i knows φ to hold. To formalise results and opportunities we borrow constructs from dynamic logic: $\langle \mathrm{do}_i(\alpha)\rangle\varphi$ denotes that agent i has the opportunity to perform the action α and that φ will result from this performance. The abilities of agents are formalised through the $\mathbf{A}_{\_}\_$ operator: $\mathbf{A}_i\alpha$ states that agent i has the ability to perform the action α. The class of actions that we consider here is built up from a set of atomic actions using a variety of constructors. These constructors deviate somewhat from the standard actions from dynamic logic [12, 15], but are both well-known from high-level programming languages and somewhat closer to philosophical views on actions than the standard constructors. When defining the models we will ensure that atomic actions are deterministic, i.e. the event consisting of an agent performing an action in some state of affairs has a unique outcome. As we will see later on this ensures that all actions are deterministic.

Definition 3.1 The language $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ is founded on three denumerable, non-empty sets, each of which is disjoint of the others: Π is the set of propositional variables, $\mathrm{A} \subseteq \mathbb{N}$ is the set of (names of) agents, and At is the set of atomic actions. The alphabet contains the well-known connectives $\neg$ and $\wedge$, the epistemic operator $\mathbf{K}_{\_}\_$, the dynamic operator $\langle \mathrm{do}_{\_}(\_)\rangle\_$, the ability operator $\mathbf{A}_{\_}\_$, the action constructors `confirm_` (confirmations), _;_ (sequential composition), `if_then_else_fi` (conditional composition) and `while_do_od` (repetitive composition).

Definition 3.2 The language $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ is the smallest superset of Π such that

- if $\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ and $\psi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ then $\neg\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ and $\varphi \wedge \psi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$
- if $\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$, $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}(\mathrm{At})$ then $\mathbf{K}_i\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$, $\langle \mathrm{do}_i(\alpha)\rangle\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ and $\mathbf{A}_i\alpha \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$

where the class $\mathrm{Ac}(\mathrm{At})$ of actions is the smallest superset of At such that

- if $\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ then $\texttt{confirm}\,\varphi \in \mathrm{Ac}(\mathrm{At})$
- if $\alpha_1 \in \mathrm{Ac}(\mathrm{At}), \alpha_2 \in \mathrm{Ac}(\mathrm{At})$ then $\alpha_1; \alpha_2 \in \mathrm{Ac}(\mathrm{At})$
- if $\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$, $\alpha_1 \in \mathrm{Ac}(\mathrm{At}), \alpha_2 \in \mathrm{Ac}(\mathrm{At})$ then $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi} \in \mathrm{Ac}(\mathrm{At})$
- if $\varphi \in \mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$, $\alpha \in \mathrm{Ac}(\mathrm{At})$ then $\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od} \in \mathrm{Ac}(\mathrm{At})$

The constructs $\vee$, $\rightarrow$, $\leftrightarrow$, $\top$, denoting the canonical tautology and $\bot$, denoting the canonical contradiction, are defined in the usual way. Other constructs are introduced by definitional abbreviation:

$\mathbf{M}_i\varphi$	$=^{\mathrm{def}}$	$\neg\mathbf{K}_i\neg\varphi$
$[\mathrm{do}_i(\alpha)]\varphi$	$=^{\mathrm{def}}$	$\neg\langle\mathrm{do}_i(\alpha)\rangle\neg\varphi$
`skip`	$=^{\mathrm{def}}$	$\texttt{confirm}\,\top$
`fail`	$=^{\mathrm{def}}$	$\texttt{confirm}\,\bot$
α^0	$=^{\mathrm{def}}$	`skip`
α^{n+1}	$=^{\mathrm{def}}$	$\alpha; \alpha^n$

The following letters, possibly marked, are used as typical elements:

- p, q, r for the elements of Π
- i, j for the elements of A
- a, b, c for the elements of At
- φ, ψ, ρ for the elements of $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$
- α, β, γ for the elements of $\mathrm{Ac}(\mathrm{At})$

Whenever the sets $\Pi, \mathrm{A}, \mathrm{At}$ are understood, which we assume to be the case unless explicitly stated otherwise, we write L and Ac rather than $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ and $\mathrm{Ac}(\mathrm{At})$.

The intuitive interpretation of formulas $\mathbf{K}_i\varphi$, $\langle\mathrm{do}_i(\alpha)\rangle\varphi$ and $\mathbf{A}_i\alpha$ is discussed above. The formula $\mathbf{M}_i\varphi$ is the dual of $\mathbf{K}_i\varphi$ and represents the epistemic possibility of φ for agent i, i.e. on the basis of its knowledge, i considers φ to be possible. The formula $[\mathrm{do}_i(\alpha)]\varphi$ is the dual of $\langle\mathrm{do}_i(\alpha)\rangle\varphi$; this formula is noncommittal about the opportunity of agent i to perform the action α but states that if the opportunity to do α is present, then φ would be among the results of $\mathrm{do}_i(\alpha)$. The action constructors presented in Definition 3.2 constitute the class of so-called *strict programs* (cf. [13, 15]). Their intuitive interpretation is as follows:

$\texttt{confirm}\,\varphi$	verify φ
$\alpha_1; \alpha_2$	α_1 followed by α_2
$\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$	α_1 if φ holds and α_2 otherwise
$\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}$	α as long as φ holds

The action `skip` represents the void action, and `fail` denotes the abort action. The action α^n consists of sequentially doing α n times.

3.1 The KARO-framework: semantics

The vast majority of all interpretations proposed for modal languages is based on the use of Kripke-style possible worlds models. The models that we use to interpret formulas from L contain a set S of possible worlds, representing actual and hypothetical states of affairs, a valuation π on the elements of Π, indicating which atomic propositions are true in which possible world, a relation R denoting epistemic accessibility, and two functions $\mathtt{r}_0$ and $\mathtt{c}_0$ dealing with (the result, opportunity and ability for) atomic actions. In the sequel $\mathrm{S}^{\cdot}$ denotes the lift of S, i.e. $\mathrm{S}^{\cdot} =^{\mathrm{def}} \mathrm{S} \cup \{\emptyset\}$, and bool $=^{\mathrm{def}} \{\mathbf{1}, \mathbf{0}\}$ is a set of truth values.

Definition 3.3 A model M for L is a tuple consisting of the following five elements:

- a non-empty set S of possible worlds or states.
- a valuation $\pi : \Pi \times \mathrm{S} \to \mathrm{bool}$ on propositional symbols.
- a function $\mathrm{R} : \mathrm{A} \to \wp(\mathrm{S} \times \mathrm{S})$ indicating the epistemic alternatives of agents. This function is required to be such that $\mathrm{R}(i)$ is an equivalence relation for all $i \in \mathrm{A}$.
- a function $\mathtt{r}_0 : \mathrm{A} \times \mathrm{At} \to \mathrm{S} \to \mathrm{S}^{\cdot}$ indicating the state-transitions caused by the execution of atomic actions.
- a function $\mathtt{c}_0 : \mathrm{A} \times \mathrm{At} \to \mathrm{S} \to \mathrm{bool}$ determining the abilities of agents with regard to atomic actions.

Note that models in principle depend on the language: valuations π depend on Π, there are epistemic alternatives for each agent and the state transitions and abilities are agent- and atomic action-dependent. However, we will in the semantics often omit reference to the sets Π, A and At. The class containing all models for L is denoted by $\mathbf{M}$. The letter M, possibly marked, denotes a typical model, and s, t, u, possibly marked, are used as typical elements of the set of states.

The relation $\mathrm{R}(i)$ indicates which pairs of worlds are indistinguishable for agent i on the basis of its knowledge: if $(s, s') \in \mathrm{R}(i)$ then whenever s is the description of the actual world, s' might as well be for all agent i knows. To ensure that knowledge indeed has the properties sketched in Section 2, it is demanded that $\mathrm{R}(i)$ is an equivalence relation for all i. That this demand ensures that knowledge behaves as desired is stated in Proposition 4.1 and explained in Proposition 4.7. The function $\mathtt{r}_0$ characterises occurrences of atomic events, i.e. events consisting of an agent performing an atomic action: whenever s is some possible world, then $\mathtt{r}_0(i, a)(s)$ represents the state of affairs following execution of the atomic action a in the possible world s by the agent i. Since atomic actions are inherently deterministic, $\mathtt{r}_0(i, a)(s)$ yields at most one state of affairs as the one resulting from the occurrence of the event $\mathrm{do}_i(a)$ in s. If $\mathtt{r}_0(i, a)(s) = \emptyset$, we will sometimes say that execution of a by i in s leads to the (unique) *counterfactual state of affairs*, i.e. a state of affairs which is neither actual nor hypothetical, but counterfactual. One may think of $\mathtt{r}_0(i, a)(s) = \emptyset$ as indicating a serious failure, rather than just a disappointment: from $\emptyset$, no further actions can be taken. The function $\mathtt{c}_0$ acts as a kind of valuation on atomic actions, i.e. $\mathtt{c}_0(i, a)(s)$ indicates whether agent i has the ability to perform the action a in the possible world s.

Formulas from the language L are interpreted on the possible worlds in the models from $\mathbf{M}$. Propositional symbols are directly interpreted using the valuation π: a propositional symbol p is true in a state s iff $\pi(p, s)$ yields the value **1**. Negations and conjunctions are interpreted as in classical logic: a formula $\neg\varphi$ is true in a state s iff φ is not true in s and $\varphi \wedge \psi$ is true in s iff both φ and ψ are true in s. The knowledge formulas $\mathbf{K}_i\varphi$ are interpreted using the epistemic accessibility relation $\mathrm{R}(i)$: agent i knows that φ in s iff φ is true in all the possible worlds that are epistemically equivalent to s, for that agent. The dynamic formulas $\langle \mathrm{do}_i(\alpha)\rangle\varphi$ and the ability formulas $\mathbf{A}_i\alpha$ are interpreted through the extensions $\mathtt{r}$ and $\mathtt{c}$ of the functions $\mathtt{r}_0$ and $\mathtt{c}_0$, respectively. These extensions $\mathtt{r}$ and $\mathtt{c}$ will be defined in Definition 3.4. Informally, a formula $\langle \mathrm{do}_i(\alpha)\rangle\varphi$ is true in some possible world s, if the extension $\mathtt{r}$ of $\mathtt{r}_0$ applied to i, α and s yields some successor state s' in which the formula φ holds. A formula $\mathbf{A}_i\alpha$ is true in a state s if the extension $\mathtt{c}$ of $\mathtt{c}_0$ yields the value **1** when applied to i, α and s. Before defining the extended versions of $\mathtt{r}_0$ and $\mathtt{c}_0$, we first motivate the choices underlying these extensions.

3.1.1 Results and opportunities for composite actions

Recall from the introduction to this section that $\langle \mathrm{do}_i(\alpha)\rangle\varphi$ denotes that agent i has the opportunity to perform action α in such a way that φ will result from this performance. Thus, we can define the opportunity *sec* to do α as $\langle \mathrm{do}_i(\alpha)\rangle\top$. Note that under our assumption about determinism of actions, the formula $\langle \mathrm{do}_i(\alpha)\rangle\varphi$ is in fact stronger than $[\mathrm{do}_i(\alpha)]\varphi$: whereas the diamond-formula expresses that i has the opportunity to do α and φ will be among the results of i's doing α, the box-formula conditions φ being a result of i performing α upon i's opportunity to do α.

The extension $\mathtt{r}$ of the function $\mathtt{r}_0$ as we will present it is originally due to Halpern& Reif [13]. Although Halpern & Reif's logic is meant to reason about computer programs and not about agents performing actions, we argue that their definition is also adequate for our purposes. Using this definition, actions $\mathtt{confirm}\,\varphi$ are interpreted as genuine confirmations: whenever the formula φ is true in a state s, s is its own $\mathrm{do}_i(\mathtt{confirm}\,\varphi)$-successor. If φ does not hold in a possible world s, then the $\mathtt{confirm}\,\varphi$ action fails, and no successor state results. In practice this implies that (all) agents have the opportunity to confirm the truth of a certain formula iff the formula holds. Execution of such an action does not have any effects in the case that the formula that is confirmed holds, and leads to the counterfactual state of affairs if the formula does not hold[1].

Since the action $\alpha_1; \alpha_2$ is intuitively interpreted as 'α_1 followed by α_2', the transition caused by execution of an action $\alpha_1; \alpha_2$ equals the 'sum' of the transition caused by α_1 and the one caused by α_2 in the state brought about by execution of α_1. In the case that execution of α_1 leads to an empty set of states, execution of the action $\alpha_1; \alpha_2$ also leads to an empty set: there is no escape from the counterfactual state of affairs. In practice this implies that an

[1] Originally in dynamic logic [15, 24] these actions were referred to as tests instead of confirmations. As long as one deals with the behaviour of computer programs, the term 'test' is quite acceptable. However, as soon as formalisations of (human) agents are concerned, one should be careful with using this term. The common-sense notion of test is that of an action, execution of which provides some kind of information (in our terminology of [31], a test is a 'knowledge producing action' and 'informative'.). For example dope-tests and eye-tests are performed in order to acquire information on whether some athlete has been taking drugs, or whether someone's eyesight is adequate. The nature of this kind of tests is not captured by the action which just checks for the truth of some proposition, without yielding any information whatsoever. To avoid confusion we have chosen to refer to these latter kinds of actions as confirmations. Thus, in terms of our models, we think of a confirmation as an action that does not change the state, whereas an agent testing for φ might end up in a different (epistemic) state.

agent has the opportunity to perform a sequential composition $\alpha_1; \alpha_2$ iff it has the opportunity to do α_1 (now), and doing α_1 results in the agent having the opportunity to do α_2. The results of performing $\alpha_1; \alpha_2$ equal the results of doing α_2, having done α_1.

Given its intuitive meaning, it is obvious that the transition caused by a conditional composition $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ equals the one associated with α_1 in the case that φ holds and the one caused by execution of α_2 in the case that $\neg\varphi$ holds. This implies that an agent has the opportunity to perform an action $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ if (it has the opportunity to confirm that) φ holds and it has the opportunity to do α_1, or (it has the opportunity to confirm that) $\neg\varphi$ holds and the agent has the opportunity to do α_2. The result of performing $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ equals the result of α_1 in the case that φ holds and that of α_2 otherwise.

The definition of the extension $\mathtt{r}$ of $\mathtt{r}_0$ for the repetitive composition is based on the idea that execution of the action $\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}$ comes down to sequentially testing for the truth of φ and executing α until a state is reached in which $\neg\varphi$ holds. For deterministic while-loops $\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}$, at most one of the actions $\beta_k = ((\texttt{confirm}\,\varphi; \alpha)^k; \texttt{confirm}\,\neg\varphi)$, with $k \in \mathbb{N}$, has an execution which does not lead to the counterfactual state of affairs. Now if such an action β_k exists, the resulting state of execution of the while-loop is defined to be the state resulting from execution of β_k, and otherwise execution of the loop is taken to lead to the counterfactual state of affairs.

3.1.2 Abilities for composite actions

Whereas the extension $\mathtt{r}$ of $\mathtt{r}_0$ for composite actions is more or less standard, the extension $\mathtt{c}$ of $\mathtt{c}_0$ as determining the abilities of agents for composite actions, is not. Since we are (among) the first to give a formal, exogenous account of ability, extending the function $\mathtt{c}_0$ to the class of all actions involves a couple of personal choices.

We start with motivating our definitions of ability for confirmations and conditional compositions since neither of these is really controversial: the definition of ability for confirmations is indisputable since it represents a highly personal choice (and there is no accounting for tastes), and that of the ability for the conditional composition is too obvious and natural to be questioned.

We have decided to let an agent have the ability to confirm any formula that is actually true. Since confirmations do not correspond to any actions usually performed by humans, this definition seems to be perfectly acceptable, or at least it is hard to come up with any convincing counterarguments to it. Note that this definition implies that in a situation where some proposition is true, (all) agents have both the opportunity and the ability to confirm this proposition.

Let us continue with defining abilities for conditionally composed actions. For these actions, ability is defined analogously to opportunity: an agent is able to perform the action $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ iff either it is able to confirm the condition φ and perform α_1 afterwards, or it is able to confirm the negation of the condition and perform α_2. In practice this implies that having the ability to perform an action $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ boils down to being able to do α_1 whenever φ holds and being able to do α_2 whenever φ does not hold. In our opinion this is *the* natural way to define the ability for conditionally composed actions, thereby accepting the import of some oddities of conditionals like the following. With our definition, an agent may claim on Tuesday that it has the ability to jump over the moon if it is Wednesday and scratch its nose otherwise'[2]. This may be undesirable, but also note that it is not the same as (the even worse) claim that on Wednesday it is able to jump over the moon.

Whereas the definitions of the ability for confirmations and conditional compositions are easily explained and motivated, this is not the case for those describing the ability for sequential and repetitive compositions, even though the basic ideas underlying these definitions are perfectly clear.

Informally, having the ability to perform a sequentially composed action $\alpha_1;\alpha_2$ is defined as having the ability to do α_1 now, while being able to do α_2 as a result of having done α_1. If the opportunity to perform α_1 exists, i.e. performing α_1 does not result in the counterfactual state of affairs, there is no question concerning the intuitive correctness of this definition, but things are different when this opportunity is absent. It is not clear how the abilities of agents are to be determined in the counterfactual state of affairs. Probably the most acceptable approach would be to declare the question of whether the agent is able to perform an action in the counterfactual state of affairs to be meaningless, which could be formalised by extending the set of truth-values to contain an element representing undefinedness of a proposition. Since this would necessitate a considerable complication of our classical, two-valued approach, we have chosen not to explore this avenue, which leaves us with the task of assigning a classical truth-value to the agents' abilities in the counterfactual state of affairs. In general we see two ways of doing this, the first of which would be to treat all actions equally and come up with a uniform truth value for the abilities of all agents to perform any action in the counterfactual state of affairs. This approach is relatively simply to formalise, and is in fact the one that we will pursue. The second approach would be to treat each action individually, and determine the agents' abilities through other means, such as by assuming an agent to be in the possession of certain default, or typical, abilities. This approach is further discussed in Section 7. Coming back to the first approach, it is obvious that — given that there are exactly two truth-values — two ways exist to treat all actions equally with respect to the agents' abilities in the counterfactual state of affairs. The first of these could be called an *optimistic*, or bold, approach, and states that agents are omnipotent in the counterfactual state of affairs. According to this approach, in situations where an agent does have the ability but not the opportunity to perform an action α_1 it is concluded that the agent has the ability to perform the sequential composition $\alpha_1;\alpha_2$ for arbitrary actions α_2. The second approach is a *pessimistic*, or careful one. In this approach agents are assumed to be nilpotent in counterfactual situations. Thus, in situations in which an agent does have the ability but not the opportunity to perform an action α_1 it is concluded that the agent is unable to perform the sequential composition $\alpha_1;\alpha_2$ for all α_2. Note that in the case that the agent has the opportunity to do α_1, optimistic and pessimistic approaches towards the agent's ability to do $\alpha_1;\alpha_2$ coincide. Although there is a case for both definitions, neither is completely acceptable. Consider the example of a lion in a cage, which is perfectly well capable of eating a zebra, but ideally never has the opportunity to do so. Using the first definition we would have to conclude that the lion is capable of performing the sequential composition 'eat zebra; fly to the moon', which hardly seems intuitive. Using the second definition it follows that the lion is unable to perform the action 'eat zebra; do nothing', which seems equally counterintuitive. Fortunately, the problems associated with these definitions are not really serious. They occur

[2] these oddities, and, in particular, this example, was suggested by a referee of a preliminary version of this paper

only in situations where an agent has the ability but not the opportunity to perform some action. And since it is exactly the combination of opportunity and ability that is important, no unwarranted conclusions can be drawn in these situations. Henceforth, we pursue both the optimistic and the pessimistic approach; in Section 7 we suggest alternative approaches in which the aforementioned counterintuitive situations do not occur.

Defining abilities for while-loops is even more hazardous than for sequential compositions. Intuitively it seems a good point of departure to let an agent be able to perform a while-loop only if it is at any point during execution capable of performing the next step. However, using this intuitive definition one has to be careful not to jump to undesired conclusions in the case of an action for which execution does not terminate. It seems highly counterintuitive to declare an agent, be it artificial or not, to have the reliable ability to perform an action that goes on indefinitely. For no agent is eternal: human agents die, artificial agents break down, and after all even the lifespan of the earth and the universe is bounded. Hence agents should not be able to perform actions that take infinite time. Therefore it seems reasonable to equate the ability to perform a while-loop with the ability to perform some finite-length sequence of confirmations and actions constituting the body of the while-loop, which ends in a confirmation for the negation of the condition of the loop, analogously to the equation used in extending the function $\mathtt{r}_0$ to while-loops. Accepting this equation, it is obvious that the discussion concerning the ability of agents for sequentially composed actions also becomes relevant for the repetitive composition, i.e. also with respect to abilities for while-loops a distinction between optimistic and pessimistic agents can be made. In the case that the while-loop terminates, optimistic and pessimistic approaches coincide, but in the case that execution of the action leads to the counterfactual state of affairs, they differ. Consider the situation of an agent that up to a certain point during the execution of an action $\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$ has been able to perform the confirmation for φ followed by α, and now finds itself in a state where φ holds, it is able to do α but does not have the opportunity for α. An optimistic agent concludes that it would have been able to finish the finite-length sequence constituting the while-loop after the (counterfactual) execution of α, and therefore considers itself to be capable of performing the while-loop. A pessimistic agent considers itself unable to finish the sequence, and thus is unable to perform the while-loop. The demand for finiteness of execution of the while-loop and the pessimistic view on abilities provide for a very interesting combination. For in order for an agent to be able to perform an action $\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$ it has to have the opportunity to perform all the steps in the execution of $\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$, possibly except for the last one. Furthermore, as as result of performing the last but one step in the execution the agent should obtain the ability to perform the last one, which is a confirmation for $\neg\varphi$. Since ability and opportunity coincide for confirmations this implies that the agent has the opportunity to confirm $\neg\varphi$, i.e. the agent has the opportunity to perform the last step in the execution of $\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$. But then the agent has the opportunity to perform all the steps in the execution of the while-loop, and thus has the opportunity to perform the while-loop. Hence in the pessimistic approach the ability to perform a while-loop implies the opportunity!

3.1.3 Formally interpreting knowledge, abilities, results and opportunities

To interpret dynamic and ability formulas from L in a model M for L, the functions $\mathtt{r}_0$ and $\mathtt{c}_0$ from M are extended to deal with composite, i.e. non-atomic actions. To account for the difference between the optimistic and the pessimistic outlook on the agents' abilities, we define two different extensions of $\mathtt{c}_0$, and thereby also two different interpretations. The optimistic and the pessimistic approach coincide in their extension of $\mathtt{r}_0$, but differ in the extension of $\mathtt{c}_0$ for sequentially composed actions, and hence also in their treatment of ability for repetitive compositions. The following definition presents the extensions of $\mathtt{r}_0$ and $\mathtt{c}_0$. Here functions with the superscript **1** correspond to the optimistic view, and those with the superscript **0** to the pessimistic view on the agents' abilities in the counterfactual state of affairs.

Definition 3.4 For $\mathbf{b} \in$ bool we inductively define the binary relation $\models^{\mathbf{b}}$ between a formula from L and a pair M, s consisting of a model M for L and a state s in M for the dynamic and ability formulas as follows:

$$\begin{array}{ll}
\mathrm{M}, s \models^{\mathbf{b}} p & \Leftrightarrow \pi(p, s) = \mathbf{1} \text{ for } p \in \Pi \\
\mathrm{M}, s \models^{\mathbf{b}} \neg\varphi & \Leftrightarrow \text{not}\,(\mathrm{M}, s \models^{\mathbf{b}} \varphi) \\
\mathrm{M}, s \models^{\mathbf{b}} \varphi \wedge \psi & \Leftrightarrow \mathrm{M}, s \models^{\mathbf{b}} \varphi \text{ and } \mathrm{M}, s \models^{\mathbf{b}} \psi \\
\mathrm{M}, s \models^{\mathbf{b}} \mathbf{K}_i\varphi & \Leftrightarrow \forall s' \in \mathrm{S}((s, s') \in \mathrm{R}(i) \Rightarrow \mathrm{M}, s' \models^{\mathbf{b}} \varphi) \\
\mathrm{M}, s \models^{\mathbf{b}} \langle \mathrm{do}_i(\alpha)\rangle\varphi & \Leftrightarrow \exists s' \in \mathrm{S}(s' = \mathtt{r}^{\mathbf{b}}(i, \alpha)(s)\, \&\, \mathrm{M}, s' \models^{\mathbf{b}} \varphi) \\
\mathrm{M}, s \models^{\mathbf{b}} \mathbf{A}_i\alpha & \Leftrightarrow \mathtt{c}^{\mathbf{b}}(i, \alpha)(s) = \mathbf{1}
\end{array}$$

where $\mathtt{r}^{\mathbf{b}}$ and $\mathtt{c}^{\mathbf{b}}$ are defined by:

$$\begin{array}{lcl}
\mathtt{r}^{\mathbf{b}} & : & \mathrm{A} \times \mathrm{Ac} \to \mathrm{S}^{\cdot} \to \mathrm{S}^{\cdot} \\
\mathtt{r}^{\mathbf{b}}(i, a)(s) & = & \mathtt{r}_0(i, a)(s) \\
\mathtt{r}^{\mathbf{b}}(i, \mathtt{confirm}\,\varphi)(s) & = & s \text{ if } \mathrm{M}, s \models^{\mathbf{b}} \varphi \\
 & = & \emptyset \text{ otherwise} \\
\mathtt{r}^{\mathbf{b}}(i, \alpha_1; \alpha_2)(s) & = & \mathtt{r}^{\mathbf{b}}(i, \alpha_2)(\mathtt{r}^{\mathbf{b}}(i, \alpha_1)(s)) \\
\mathtt{r}^{\mathbf{b}}(i, \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})(s) & = & \mathtt{r}^{\mathbf{b}}(i, \alpha_1)(s) \text{ if } \mathrm{M}, s \models^{\mathbf{b}} \varphi \\
 & = & \mathtt{r}^{\mathbf{b}}(i, \alpha_2)(s) \text{ otherwise} \\
\mathtt{r}^{\mathbf{b}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) & = & s' \text{ if } s' = \mathtt{r}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s) \\
 & & \quad \text{for some } k \in \mathbb{N} \\
 & = & \emptyset \text{ otherwise} \\
\mathtt{r}^{\mathbf{b}}(i, \alpha)(\emptyset) & = & \emptyset \\
 & & \\
\mathtt{c}^{\mathbf{b}} & : & \mathrm{A} \times \mathrm{Ac} \to \mathrm{S}^{\cdot} \to \text{bool} \\
\mathtt{c}^{\mathbf{b}}(i, a)(s) & = & \mathtt{c}_0(i, a)(s) \\
\mathtt{c}^{\mathbf{b}}(i, \mathtt{confirm}\,\varphi)(s) & = & \mathbf{1} \text{ iff } \mathrm{M}, s \models^{\mathbf{b}} \varphi \\
\mathtt{c}^{\mathbf{b}}(i, \alpha_1; \alpha_2)(s) & = & \mathbf{1} \text{ iff } \mathtt{c}^{\mathbf{b}}(i, \alpha_1)(s) = \mathbf{1}\, \&\, \mathtt{c}^{\mathbf{b}}(i, \alpha_2)(\mathtt{r}^{\mathbf{b}}(i, \alpha_1)(s)) = \mathbf{1} \\
\mathtt{c}^{\mathbf{b}}(i, \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})(s) & = & \mathbf{1} \text{ iff } \mathtt{c}^{\mathbf{b}}(i, \mathtt{confirm}\,\varphi; \alpha_1)(s) = \mathbf{1} \text{ or} \\
 & & \quad \mathtt{c}^{\mathbf{b}}(i, \mathtt{confirm}\,\neg\varphi; \alpha_2)(s) = \mathbf{1} \\
\mathtt{c}^{\mathbf{b}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) & = & \mathbf{1} \text{ iff } \mathtt{c}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s) = \mathbf{1} \\
 & & \quad \text{for some } k \in \mathbb{N} \\
\mathtt{c}^{\mathbf{b}}(i, \alpha)(\emptyset) & = & \mathbf{b}
\end{array}$$

The formula φ is $\models^{\mathbf{b}}$-satisfiable in the model M iff M, $s \models^{\mathbf{b}} \varphi$ for some s in M; φ is $\models^{\mathbf{b}}$-valid in M, denoted by $\mathrm{M} \models^{\mathbf{b}} \varphi$, iff M, $s \models^{\mathbf{b}} \varphi$ for all s in M. The formula φ is $\models^{\mathbf{b}}$-satisfiable in $\mathcal{M}$ iff φ is $\models^{\mathbf{b}}$-satisfiable in some $\mathrm{M} \in \mathcal{M}$; φ is $\models^{\mathbf{b}}$-valid in $\mathcal{M}$, denoted by $\models^{\mathbf{b}} \varphi$, iff φ is $\models^{\mathbf{b}}$-valid in all $\mathrm{M} \in \mathcal{M}$. Whenever $\models^{\mathbf{b}}$ is clear from the context, we drop it as a prefix and simply speak of a formula φ being satisfiable or valid in a (class of) model(s). For a given model M, we define $[s]_{\mathrm{R}(i)} =^{\mathrm{def}} \{s' \in \mathrm{S} \mid (s, s') \in \mathrm{R}(i)\}$ and $[\![\varphi]\!]_{\mathrm{M}} =^{\mathrm{def}} \{s \in \mathrm{S} \mid \mathrm{M}, s \models^{\mathbf{b}} \varphi\}$. Whenever the model M is clear from the context, the latter notion is usually simplified to $[\![\varphi]\!]$.

4 Properties of knowledge and actions in the KARO-framework

In this section we look at the properties that knowledge and actions have in the KARO-framework. We furthermore consider additional properties, and show how some of these additional properties can be brought about by imposing constraints on the interpretation of atomic actions. We start with the properties of knowledge. When demanding the agents' epistemic accessibility relations to be equivalence relations, the modal operator **K** indeed formalises the notion of knowledge discussed in Section 2.

Proposition 4.1 For all $i \in \mathrm{A}$ and $\varphi, \psi \in \mathrm{L}$ we have:

1. $\models^{\mathbf{b}} \mathbf{K}_i(\varphi \to \psi) \to (\mathbf{K}_i\varphi \to \mathbf{K}_i\psi)$ K
2. $\models^{\mathbf{b}} \varphi \Rightarrow \models \mathbf{K}_i\varphi$ N
3. $\models^{\mathbf{b}} \mathbf{K}_i\varphi \to \varphi$ T
4. $\models^{\mathbf{b}} \mathbf{K}_i\varphi \to \mathbf{K}_i\mathbf{K}_i\varphi$ 4
5. $\models^{\mathbf{b}} \neg\mathbf{K}_i\varphi \to \mathbf{K}_i\neg\mathbf{K}_i\varphi$ 5

The first two items of Proposition 4.1 formalise that $\mathbf{K}_i$ is a normal modal operator: $\mathbf{K}_i$ satisfies both the K-axiom and the necessitation rule N (the names of these and other modal axioms are according to the Chellas classification [4]). Furthermore, $\mathbf{K}_i$ satisfies the axioms of veridicality (the T-axiom), positive introspection (axiom 4) and negative introspection (axiom 5).

Although $\models^{\mathbf{1}}$ differs from $\models^{\mathbf{0}}$, the compositional behaviour of actions with respect to opportunities and results is identical in the two interpretations.

Proposition 4.2 For $\mathbf{b} \in \mathrm{bool}$, $i \in \mathrm{A}$, $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$ and $\varphi, \psi \in \mathrm{L}$ we have:

1. $\models^{\mathbf{b}} \langle \mathrm{do}_i(\mathtt{confirm}\,\varphi)\rangle\psi \leftrightarrow (\varphi \wedge \psi)$
2. $\models^{\mathbf{b}} \langle \mathrm{do}_i(\alpha_1;\alpha_2)\rangle\psi \leftrightarrow \langle \mathrm{do}_i(\alpha_1)\rangle\langle \mathrm{do}_i(\alpha_2)\rangle\psi$
3. $\models^{\mathbf{b}} \langle \mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})\rangle\psi \leftrightarrow ((\varphi \wedge \langle \mathrm{do}_i(\alpha_1)\rangle\psi) \vee (\neg\varphi \wedge \langle \mathrm{do}_i(\alpha_2)\rangle\psi))$
4. $\models^{\mathbf{b}} \langle \mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})\rangle\psi \leftrightarrow ((\neg\varphi \wedge \psi) \vee (\varphi \wedge \langle \mathrm{do}_i(\alpha)\rangle\langle \mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})\rangle\psi))$
5. $\models^{\mathbf{b}} [\mathrm{do}_i(\alpha)](\varphi \to \psi) \to ([\mathrm{do}_i(\alpha)]\varphi \to [\mathrm{do}_i(\alpha)]\psi)$
6. $\models^{\mathbf{b}} \psi \Rightarrow \models^{\mathbf{b}} [\mathrm{do}_i(\alpha)]\psi$

Proposition 4.2 is in fact nothing but a formalisation of the intuitive ideas on results and opportunities for composite actions as expressed above. The first item states that agents have the opportunity to confirm exactly the formulas that are true, and that no state-transition takes place as the result of such a confirmation. The second item deals with the separation of the sequential composition into its elements: an agent has the opportunity to do $\alpha_1;\alpha_2$ with result ψ iff it has the opportunity to do α_1 (now) and doing so will result in having the opportunity to do α_2 with result ψ. The third item states that a conditionally composed action equals its 'then'-part in the case that the condition holds, and its 'else'-part if the condition does not hold. The fourth item formalises a sort of fixed-point equation for execution of while-loops: if an agent has the opportunity to perform a while-loop then it keeps this opportunity under execution of the body of the loop as long as the condition holds. The result of performing a while-loop is also fixed under executions of the body of the loop in states where φ holds, and is determined by the propositions that are true in the first state where $\neg\varphi$ holds. Note that a validity like this one does not suffice to axiomatise the repetitive composition: although it captures the idea of while-loops representing fixed-points, it fails to force termination, i.e. this formula on its own does not guarantee that agents do not have the opportunity to bring an infinitely non-terminating while-loop to its end. In the proof systems that we present in Section 6 this problem is solved by including suitable proof rules guiding the repetitive composition. The last two items state the normality of $[\mathrm{do}_i(\alpha)]$.

As soon as the abilities of agents come into play, the differences between $\models^{\mathbf{1}}$ and $\models^{\mathbf{0}}$ become visible, in particular for sequential and repetitive compositions.

Proposition 4.3 For $\mathbf{b} \in \mathrm{bool}$, $i \in \mathrm{A}$, $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$ and $\varphi \in \mathrm{L}$ we have:

1. $\models^{\mathbf{b}} \mathbf{A}_i\mathtt{confirm}\,\varphi \leftrightarrow \varphi$
2. $\models^{\mathbf{1}} \mathbf{A}_i\alpha_1;\alpha_2 \leftrightarrow \mathbf{A}_i\alpha_1 \wedge [\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2$
3. $\models^{\mathbf{0}} \mathbf{A}_i\alpha_1;\alpha_2 \leftrightarrow \mathbf{A}_i\alpha_1 \wedge \langle\mathrm{do}_i(\alpha_1)\rangle\mathbf{A}_i\alpha_2$
4. $\models^{\mathbf{b}} \mathbf{A}_i\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi} \leftrightarrow ((\varphi \wedge \mathbf{A}_i\alpha_1) \vee (\neg\varphi \wedge \mathbf{A}_i\alpha_2))$
5. $\models^{\mathbf{1}} \mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \leftrightarrow (\neg\varphi \vee (\varphi \wedge \mathbf{A}_i\alpha \wedge [\mathrm{do}_i(\alpha)]\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}))$
6. $\models^{\mathbf{0}} \mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \leftrightarrow (\neg\varphi \vee (\varphi \wedge \mathbf{A}_i\alpha \wedge \langle\mathrm{do}_i(\alpha)\rangle\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}))$

The first and the fourth items of Proposition 4.3 deal with the actions for which abilities are defined in a straightforward manner: agents are able to confirm exactly the true formulas, and having the ability to perform a conditional composition comes down to having the 'right' ability, dependent on the truth or falsity of the condition. The differences between the optimistic and the pessimistic outlook on abilities in the counterfactual state of affairs are clearly visible in the other items of Proposition 4.3. Optimistic agents are assumed to be omnipotent in counterfactual situations, and therefore it suffices for the agent to be able to do α_2 as a conditional result of doing α_1. A pessimistic agent needs certainty, and therefore demands to have the opportunity to do α_1 before concluding anything on its abilities following execution of α_1. This behaviour of optimistic and pessimistic agents is formalised in the second and the third item, respectively. The fifth and sixth item formalise an analogous behaviour for repetitive compositions: optimistic agents are satisfied with conditional results (item 5) whereas pessimistic agents demand certainty (item 6).

The compositional behaviour of sequential and repetitive compositions differs for the two interpretations only in situations where an agent lacks opportunities. If all appropriate opportunities are present, there is no difference for the two interpretations, a property which is formalised in the following corollary.

Corollary 4.4 For $i \in \mathrm{A}$, $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$ and $\varphi \in \mathrm{L}$ we have:

- $\models^{\mathbf{1}} \langle\mathrm{do}_i(\alpha_1)\rangle\top \rightarrow (\mathbf{A}_i\alpha_1;\alpha_2 \leftrightarrow \mathbf{A}_i\alpha_1 \wedge \langle\mathrm{do}_i(\alpha_1)\rangle\mathbf{A}_i\alpha_2)$
- $\models^{\mathbf{1}} \langle\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})\rangle\top \rightarrow$
  $(\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \leftrightarrow (\neg\varphi \vee (\varphi \wedge \mathbf{A}_i\alpha \wedge \langle\mathrm{do}_i(\alpha)\rangle\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})))$

4.1 Frames and correspondences

To investigate properties of knowledge and actions, it will often prove useful to refer to *schemas*, which are sets of formulas, usually of a particular form. Using schemas one may abstract from particular agents and particular formulas, thereby having the possibility to formulate certain qualities of knowledge and action in a very general way. For instance, the axiom 4, $\mathbf{K}_i\varphi \to \mathbf{K}_i\mathbf{K}_i\varphi$, considered as a schema *in* φ expresses positive introspection of the agent i, and, as a schema in φ and i, it denotes positive introspection of all of the agents. If context allows it, we will remain implicit about what exactly are the varying elements in a schema.

Where schemas are used to express general properties of knowledge and actions on the syntactic level, *frames* can be used to do so on the semantic level. Informally speaking, a frame can be seen as a model without a valuation. By leaving out the valuation one may abstract from particular properties of knowledge and actions that are due to the valuation rather than inherently due to the nature of knowledge and/or action itself. Truth in a frame is defined in terms of truth in all models that can be constructed by adding a valuation to the frame.

Definition 4.5 A frame F for a model $\mathrm{M} \in \mathbf{M}$ is a tuple consisting of the elements of M except for the valuation π. The class of all frames for models from $\mathbf{M}$ is denoted by $\mathbf{F}$. If F is some frame then (F, π) denotes the model generated by the elements of F and the valuation π. For F a frame, s one of its states and $\varphi \in \mathrm{L}$ we define

- $\mathrm{F}, s \models \varphi \Leftrightarrow (\mathrm{F}, \pi), s \models \varphi$ for all valuations π
- $\mathrm{F} \models \varphi \Leftrightarrow \mathrm{F}, s \models \varphi$ for all states s of F.
- $\mathbf{F} \models \varphi \Leftrightarrow \mathrm{F} \models \varphi$ for all $\mathrm{F} \in \mathbf{F}$

Since schemas are used to express general properties of knowledge and action syntactically, and frames can be used to do this semantically, the question arises as to how these notions relate. In particular, it is both interesting and important to try to single out first-order constraints on frames that exactly correspond to certain properties of knowledge and/or action, expressed in the form of schemas. The area of research called *correspondence theory* deals with finding relations — *correspondences* — between schemas and (first-order expressible) constraints on frames. A modal schema is said to correspond to a first-order constraint on frames if the schema is satisfied in exactly those frames that obey the constraint. A good introduction into correspondence theory is given in [1].

Definition 4.6 If φ is a schema and F is some frame then $\mathrm{F} \models \varphi$ iff $\mathrm{F} \models \chi$ for all formulas χ that are an instantiation of φ. If P is a formula in the first-order language subsuming the functions $\mathrm{R}, \mathtt{r}_0, \mathtt{c}_0$ and equality, then $\mathrm{F} \models^{\mathrm{fo}} P$ iff F satisfies P. The schema φ corresponds to the first-order formula P, notation $\varphi \sim P$ iff $\forall \mathrm{F}$, $(\mathrm{F} \models \varphi \Leftrightarrow \mathrm{F} \models^{\mathrm{fo}} P)$.

As already hinted at above, the properties that we require knowledge to obey correspond to constraints on the epistemic accessibility relations $\mathrm{R}(i)$. In Definition 3.3 we required these relations to be equivalence relations, and this demand indeed corresponds to knowledge being veridical and satisfying the properties of positive and negative introspection. The proof of the following proposition is standard and well-known from the literature [21, 36].

Proposition 4.7 The following correspondences hold.

1. T $\sim \forall s((s,s) \in \mathrm{R}(i))$, i.e. $\mathrm{R}(i)$ is reflexive
2. 4 $\sim \forall s, s', s''((s,s') \in \mathrm{R}(i) \,\&\, (s',s'') \in \mathrm{R}(i) \Rightarrow (s,s'') \in \mathrm{R}(i))$, i.e. $\mathrm{R}(i)$ is transitive
3. 5 $\sim \forall s, s', s''((s,s') \in \mathrm{R}(i) \,\&\, (s,s'') \in \mathrm{R}(i) \Rightarrow (s',s'') \in \mathrm{R}(i))$, i.e. $\mathrm{R}(i)$ is Euclidean

4.2 Additional properties of actions

The language L is sufficiently expressive to formalise various properties of knowledge, actions and their interplay, that are interesting both from a philosophical point of view as from the point of view of AI. The first of the properties that we consider here is *accordance*. Informally speaking, accordant actions are known to behave according to plan, i.e. for an accordant action it will be the case that things that an agent expects — on the basis of its knowledge — to hold in the future state of affairs that will result from it executing the action, are indeed known to be true by the agent when that future state of affairs has been brought about. Accordance of actions may be an important property in the context of agents planning to achieve certain goals. For if the agent knows (now) that performing some accordant action will bring about some goal, then it will be satisfied after it has executed the action: the agent knows that the goal is brought about. From a formal point of view, i-accordance of an action α corresponds to the schema $\mathbf{K}_i[\mathrm{do}_i(\alpha)]\varphi \to [\mathrm{do}_i(\alpha)]\mathbf{K}_i\varphi$.

The notion of *determinism* was already touched upon in the explanation of Definition 3.3 where it was stated that atomic actions are inherently deterministic. As we will see later on, viz. in Proposition 4.11, the determinism of atomic actions implies that of all actions. The notion of i-determinism of an action α is formalised through the schema $\langle \mathrm{do}_i(\alpha)\rangle\varphi \to [\mathrm{do}_i(\alpha)]\varphi$.

Whenever an action is *idempotent*, consecutively executing the action twice — or in general an arbitrary number of times — will have exactly the same results as performing the action just once. In a sense, the state of affairs reached after the first performance of the action can be seen as a kind of fixed-point of execution of the action. The simplest idempotent action in our framework is the void action `skip`: performing it once, twice or an arbitrary number of times will not affect the state of affairs in any way whatsoever. More interesting idempotent actions were determined in our paper on actions that change the agent's epistemic state ([31]); there, we claimed that such actions (we distinguished *retracting*, *expanding* and *revising*) have idempotency as a characterising property. Formally, i-idempotence of an action α corresponds to the schema $[\mathrm{do}_i(\alpha;\alpha)]\varphi \leftrightarrow [\mathrm{do}_i(\alpha)]\varphi$, or equivalently $\langle \mathrm{do}_i(\alpha;\alpha)\rangle\varphi \leftrightarrow \langle \mathrm{do}_i(\alpha)\rangle\varphi$.

Agents always have the opportunity to perform *realisable* actions, regardless of the circumstances, i.e. there never is an external factor that may prevent the performance of such an action. Typical realisable actions are, again, those in which the agent changes its information; an agent always has the opportunity to change its mind. The property of *A-realisability* relates ability and opportunity. For actions that are A-realisable, ability implies opportunity, i.e. whenever an agent is able to perform the action it automatically has the opportunity to perform it. Realisable actions are trivially A-realisable, and so are actions that no agent is ever capable of performing, but it seems hard to think of non-trivial examples of regular, mundane actions that an agent is able to execute and therefore automatically has the opportunity to do so. Adopting the A-realisability schema as an axiom schema would not be desirable for a general-purpose account of actions, opportunities and abilities, but might be appropriate for some specialized investigations. In any case it is far more reasonable to assume that ability implies opportunity than the reverse, given the fact that 'abilities are states that are acquired with effort [whereas] opportunities are there for the taking until they pass' ([22], p. 133). Realisability of an action α for agent i is formalised through the schema $\langle \mathrm{do}_i(\alpha)\rangle\top$ in $i \in \mathrm{A}$; A-realisability of α for agent i corresponds to $\mathbf{A}_i\alpha \to \langle \mathrm{do}_i(\alpha)\rangle\top$.

The following definition summarises the properties discussed above in a formal way.

Definition 4.8 Let $\alpha \in \mathrm{Ac}$ be some action, i an agent and let F be a frame. The right-hand side of the following definitions is to be understood as a schema in φ.

- α is i-accordant in F iff $\mathrm{F} \models \mathbf{K}_i[\mathrm{do}_i(\alpha)]\varphi \to [\mathrm{do}_i(\alpha)]\mathbf{K}_i\varphi$
- α is i-deterministic in F iff $\mathrm{F} \models \langle \mathrm{do}_i(\alpha)\rangle\varphi \to [\mathrm{do}_i(\alpha)]\varphi$
- α is i-idempotent in F iff $\mathrm{F} \models [\mathrm{do}_i(\alpha;\alpha)]\varphi \leftrightarrow [\mathrm{do}_i(\alpha)]\varphi$
- α is realisable for i in F iff $\mathrm{F} \models \langle \mathrm{do}_i(\alpha)\rangle\top$
- α is A-realisable for i in F iff $\mathrm{F} \models \mathbf{A}_i\alpha \to \langle \mathrm{do}_i(\alpha)\rangle\top$

We often omit explicit reference to the agent i in the above properties. Then, for instance, naming α accordant may either mean that is is i-accordant for all agents i, or that mentioning the particular agent is clear from context, or not important. If Prop is any of the properties defined above, we say that α has the property Prop in $\mathbf{F}$ iff α has the property Prop in every $\mathrm{F} \in \mathbf{F}$.

Here we show how these properties can be brought about to hold for all actions by imposing constraints on the functions R, $\mathtt{r}_0$ and $\mathtt{c}_0$. On the level of atomic actions, these properties correspond to first-order expressible constraints on R, $\mathtt{r}_0$ and $\mathtt{c}_0$. In Proposition 4.10 we present the correspondences for the properties of accordance, determinism, idempotence, realisability and A-realisability, respectively. Since we have defined two possible interpretations, viz. $\models^{\mathbf{1}}$ and $\models^{\mathbf{0}}$, for schemas from L in frames from $\mathbf{F}$ we have to be precise on the meaning of these correspondences.

Definition 4.9 For $\mathbf{b} \in \mathrm{bool}$ we define the schema φ to correspond to the first-order formula P given the interpretation $\models^{\mathbf{b}}$ iff $\forall \mathrm{F}(\mathrm{F} \models^{\mathbf{b}} \varphi \Leftrightarrow \mathrm{F} \models^{\mathrm{fo}} P)$. In such a case, we write $\varphi \sim^{\mathbf{b}} P$.

Proposition 4.10 For atomic actions $a \in \mathrm{At}$, the following correspondences hold in the class $\mathbf{F}$ of frames for $\mathbf{M}$ both for $\mathbf{b} = \mathbf{1}$ and $\mathbf{b} = \mathbf{0}$. The left-hand side of these correspondences is to be understood as a schema in φ.

1. $\mathbf{K}_i[\mathrm{do}_i(a)]\varphi \to [\mathrm{do}_i(a)]\mathbf{K}_i\varphi \sim^{\mathbf{b}}$
   $\forall s_0 \in \mathrm{S}\forall s_1 \in \mathrm{S}(\exists s_2 \in \mathrm{S}(s_2 = \mathtt{r}_0(i,a)(s_0)\,\&\,(s_2, s_1) \in \mathrm{R}(i)) \Rightarrow$
   $\exists s_3 \in \mathrm{S}((s_0, s_3) \in \mathrm{R}(i)\,\&\,s_1 = \mathtt{r}_0(i,a)(s_3)))$
2. $\langle \mathrm{do}_i(a)\rangle\varphi \to [\mathrm{do}_i(a)]\varphi \sim^{\mathbf{b}}$
   $\forall s \in \mathrm{S}\forall s' \in \mathrm{S}\forall s'' \in \mathrm{S}(\mathtt{r}_0(i,a)(s) = s'\,\&\,\mathtt{r}_0(i,a)(s) = s'' \Rightarrow s' = s'')$
3. $[\mathrm{do}_i(a;a)]\varphi \leftrightarrow [\mathrm{do}_i(a)]\varphi \sim^{\mathbf{b}} \forall s \in \mathrm{S}(\mathtt{r}_0(i,a)(\mathtt{r}_0(i,a)(s)) = \mathtt{r}_0(i,a)(s))$
4. $\langle \mathrm{do}_i(a)\rangle\top \sim^{\mathbf{b}} \forall s \in \mathrm{S}(\mathtt{r}_0(i,a)(s) \neq \emptyset)$

5. $\mathbf{A}_i a \to \langle \mathrm{do}_i(a)\rangle\top \sim^{\mathbf{b}} \forall s \in \mathrm{S}(\mathtt{c}_0(i,a)(s) = \mathbf{1} \Rightarrow \mathtt{r}_0(i,a)(s) \neq \emptyset)$

Since the functions $\mathtt{r}_0$ and $\mathtt{c}_0$ are defined for atomic actions only, and the functions $\mathtt{r}^{\mathbf{b}}$ and $\mathtt{c}^{\mathbf{b}}$ — which are the extensions of $\mathtt{r}_0$ and $\mathtt{c}_0$ for arbitrary actions — are constructed out of $\mathtt{r}_0$ and $\mathtt{c}_0$ and have no existence on their own, it is not possible to prove correspondences like those of Proposition 4.10 for non-atomic actions. There simply is no semantic entity to correspond the syntactic schemas with. This implies that it is in general not possible to ensure that arbitrary actions satisfy a certain property. However, it turns out that some of the properties considered above straightforwardly extend from the atomic level to the level of arbitrary actions, regardless of the interpretation that is used. This is in particular the case for the properties of A-realisability and determinism.

Proposition 4.11 The following lifting results hold for all F and $i \in \mathrm{A}$, in the case of $\models^{\mathbf{1}}$ as well as that of $\models^{\mathbf{0}}$:

- $\forall a \in \mathrm{At}(a$ is A-realisable for i in F$) \Rightarrow \forall \alpha \in \mathrm{Ac}(\alpha$ is A-realisable for i in F$)$
- $\forall a \in \mathrm{At}(a$ is i-deterministic in F$) \Rightarrow \forall \alpha \in \mathrm{Ac}(\alpha$ is i-deterministic in F$)$

Since the range of the function $\mathtt{r}_0$ is the set $\mathrm{S}^{\cdot}$, it follows directly that atomic actions are deterministic in $\mathbf{F}$: for if $a \in \mathrm{At}$, $i \in \mathrm{A}$ and s a state in some model, then $\mathtt{r}_0(i,a)(s)$ is either the empty set, or a single state from S, and hence the frame condition for determinism as given in Proposition 4.10 is satisfied. Using the lifting result obtained in Proposition 4.11 one then concludes that all actions are deterministic in $\mathbf{F}$.

Corollary 4.12 All actions $\alpha \in \mathrm{Ac}$ are deterministic in $\mathbf{F}$, both for $\models^{\mathbf{1}}$ and $\models^{\mathbf{0}}$.

Thus two of the properties formalised in Definition 4.8 can be ensured to hold for arbitrary actions by imposing suitable constraints on the frames for $\mathbf{M}$. For the other three properties, viz. accordance, idempotence and realisability, constraining the function $\mathtt{r}_0$ for atomic actions does not suffice, since this does not conservatively extend to the class of all actions. That realisability may not be lifted is easily seen by considering the action `fail`. Independent of the realisability of atomic actions, `fail` will never be realisable: the formula $\neg\langle \mathrm{do}_i(\mathtt{fail})\rangle\top$ is valid, both for $\models^{\mathbf{1}}$ and for $\models^{\mathbf{0}}$. The following examples show why accordance and idempotence are in general not to be lifted.

Example 4.13 Consider the language $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ with $\Pi = \{p, q\}$, $i \in \mathrm{A}$ and At arbitrary. Let $\mathrm{F} \in \mathbf{F}$ be a frame such that the set S of states in F contains at least two elements, say s and t, on which the relation $\mathrm{R}(i)$ is defined to be universal, and the first-order property corresponding with accordance of atomic actions is met. Let π be a valuation such that $\pi(p,s) = \pi(q,s) = \mathbf{1}, \pi(p,t) = \pi(q,t) = \mathbf{0}$. Then we have that $(\mathrm{F},\pi), s \models^{\mathbf{b}} \mathbf{K}_i[\mathrm{do}_i(\mathtt{confirm}\,p)]q$, and furthermore that $(\mathrm{F},\pi), s \not\models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{confirm}\,p)]\mathbf{K}_i q$. Hence $\mathrm{F} \not\models^{\mathbf{b}} \mathbf{K}_i[\mathrm{do}_i(\mathtt{confirm}\,p)]q \to [\mathrm{do}_i(\mathtt{confirm}\,p)]\mathbf{K}_i q$, which provides a counterexample to the lifting of accordance.

Example 4.14 Consider the language $\mathrm{L}(\Pi, \mathrm{A}, \mathrm{At})$ with $\Pi = \{p\}$, $i \in \mathrm{A}$ $\langle \mathrm{S}, \mathrm{R}, \mathtt{r}_0, \mathtt{c}_0\rangle$, where

- $\mathrm{S} = \{s_1, s_2, s_3, s_4\}$
- $\mathrm{R}(i)$ is an arbitrary equivalence relation on S

- $\mathtt{r}_0(i, a_1)(s_1) = s_1 \quad \mathtt{r}_0(i, a_1)(s_2) = s_3 \quad \mathtt{r}_0(i, a_1)(s_3) = s_3 \quad \mathtt{r}_0(i, a_1)(s_4) = \emptyset$
  $\mathtt{r}_0(i, a_2)(s_1) = s_2 \quad \mathtt{r}_0(i, a_2)(s_2) = s_2 \quad \mathtt{r}_0(i, a_2)(s_3) = s_4 \quad \mathtt{r}_0(i, a_2)(s_4) = s_4$
- $\mathtt{c}_0 : \mathrm{A} \times \mathrm{S} \to$ bool is arbitrary

It is easily checked that both a_1 and a_2 are idempotent in F. However, it is not the case that all actions that can be built on At are idempotent in F. For it holds for arbitrary $\mathbf{b} \in$ bool that $\mathrm{F} \not\models^{\mathbf{b}} [\mathrm{do}_i((a_1; a_2); (a_1; a_2))]p \leftrightarrow [\mathrm{do}_i(a_1; a_2)]p$. To see this take $\mathrm{M} = (\mathrm{F}, \pi)$ where $\pi(p, s_2) \neq \pi(p, s_4)$. In this model it holds that $\mathrm{M}, s_1 \models^{\mathbf{b}} [\mathrm{do}_i((a_1; a_2); (a_1; a_2))]p \leftrightarrow [\mathrm{do}_i(a_1; a_2)]\neg p$. Hence $\mathrm{M} \not\models^{\mathbf{b}} [\mathrm{do}_i((a_1; a_2); (a_1; a_2))]p \leftrightarrow [\mathrm{do}_i(a_1; a_2)]p$, and therefore also $\mathrm{F} \not\models^{\mathbf{b}} [\mathrm{do}_i((a_1; a_2); (a_1; a_2))]p \leftrightarrow [\mathrm{do}_i(a_1; a_2)]p$. Thus neither for $\models^{\mathbf{1}}$ nor for $\models^{\mathbf{0}}$ is $a_1; a_2$ idempotent in F.

Although we showed in Example 4.13 that accordance is not to be lifted from atomic actions to general ones, we can prove a restricted form of lifting for accordance. That is, if we leave confirmations out of consideration, we can prove that accordance is lifted.

Proposition 4.15 Let Ac^- be the confirmation-free fragment of Ac, i.e. the fragment built from atomic actions through sequential, conditional or repetitive composition. Then we have for all $\mathrm{F} \in \mathbf{F}$ and for all $\mathbf{b} \in$ bool:

- $\forall a \in \mathrm{At}(a$ is accordant for $\models^{\mathbf{b}}$ in F$) \Rightarrow \forall \alpha \in \mathrm{Ac}^-(\alpha$ is accordant for $\models^{\mathbf{b}}$ in F$)$

The properties of idempotence and (A-)realisability are in general undesirable ones. If all actions were idempotent, it would be impossible to walk the roads by taking one step at a time. Realisability would render the notion of opportunity meaningless and A-realisability would tie ability and opportunity in a way that we feel is unacceptable. Therefore we consider neither the lifting result for A-realisability to be very important, nor the absence of such a result for idempotence and realisability. And even though the property of accordance is, or may be, important, it is not one that typically holds in the lively world of human agents. Therefore we consider this property to be an exceptional one, that holds for selected actions only. Hence also for accordance the absence of a lifting result is not taken too seriously.

5 Correctness and feasibility of actions: practical possibility

Within the KARO-framework, several notions concerning agency may be formalised that are interesting not only from a philosophical point of view, but also when analysing agents in planning systems. The most important one of these notions formalises the knowledge that agents have about their practical possibilities. We consider the notion of practical possibility as relating an agent, an action, and a proposition: agents may have the practical possibility to bring about (truth of) the proposition by performing the action. We think of practical possibility as consisting of two parts, viz. correctness and feasibility. Correctness implies that no external factors will prevent the agent from performing the action and thereby making the proposition true. As such, correctness is defined in terms of opportunity and result: an action is correct for some agent to bring about some proposition iff the agent has the opportunity to perform the action in such a way that its performance results in the proposition being true. Feasibility captures the internal aspect of practical possibility. It states that it is within the agent's capacities to perform the action, and as such is nothing but a reformulation of ability. Together, correctness and feasibility constitute practical possibility.

Definition 5.1 For $\alpha \in \mathrm{Ac}$, $i \in \mathrm{A}$ and $\varphi \in \mathrm{L}$ we define:

- $\mathbf{Correct}_i(\alpha, \varphi) =^{\mathrm{def}} \langle \mathrm{do}_i(\alpha) \rangle \varphi$
- $\mathbf{Feasible}_i \alpha =^{\mathrm{def}} \mathbf{A}_i \alpha$
- $\mathbf{PracPoss}_i(\alpha, \varphi) =^{\mathrm{def}} \mathbf{Correct}_i(\alpha, \varphi) \wedge \mathbf{Feasible}_i \alpha$

The counterintuitive situations that occurred with respect to the ability of agents as described previously do not take root for practical possibility. That is, a lion that has the ability but not the opportunity to eat a zebra will neither have the practical possibility to eat a zebra first and thereafter fly to the moon nor have the practical possibility to eat a zebra and rest on its laurels afterwards. Thus even though the notion of ability suffers from problems like these, the more important notion of practical possibility does not. The importance of practical possibility manifests itself particularly when ascribing — from the outside — certain qualities to an agent. It seems that for the agent itself practical possibilities are relevant in so far as the agent has knowledge of these possibilities. For one may not expect an agent to act on its practical possibilities if the agent does not know of this possibilities. To formalise this kind of knowledge, we introduce the Can-predicate and the Cannot-predicate. The first of these predicates concerns the knowledge of agents about their practical possibilities, the latter predicate does the same for their practical impossibilities.

Definition 5.2 For $\alpha \in \mathrm{Ac}$, $i \in \mathrm{A}$ and $\varphi \in \mathrm{L}$ we define:

- $\mathbf{Can}_i(\alpha, \varphi) =^{\mathrm{def}} \mathbf{K}_i \mathbf{PracPoss}_i(\alpha, \varphi)$
- $\mathbf{Cannot}_i(\alpha, \varphi) =^{\mathrm{def}} \mathbf{K}_i \neg \mathbf{PracPoss}_i(\alpha, \varphi)$

The Can-predicate and the Cannot-predicate integrate knowledge, ability, opportunity and result, and seem to formalise one of the most important notions of agency. In fact it is probably not too bold to say that knowledge like that formalised through the Can-predicate, although perhaps in a weaker form by taking aspects of uncertainty into account, underlies all acts performed by rational agents. For rational agents act only if they have some information on both the possibility to perform the act, and its possible outcome; at least in this paper we restrict ourselves to such actions, leaving mere *experiments* out of our scope. It therefore seems worthwhile to take a closer look at both the Can-predicate and the Cannot-predicate. The following proposition focuses on the behaviour of the *means*-part of the predicates, which is the α in $\mathbf{Can}_i(\alpha, \varphi)$ and $\mathbf{Cannot}_i(\alpha, \varphi)$.

Proposition 5.3 For all $\mathbf{b} \in \mathrm{bool}$, $i \in \mathrm{A}$, $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$ and $\varphi, \psi \in \mathrm{L}$ we have:

1. $\models^{\mathbf{b}} \mathbf{Can}_i(\mathtt{confirm}\,\varphi, \psi) \leftrightarrow \mathbf{K}_i(\varphi \wedge \psi)$
2. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\mathtt{confirm}\,\varphi, \psi) \leftrightarrow \mathbf{K}_i(\neg\varphi \vee \neg\psi)$
3. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha_1; \alpha_2, \varphi) \leftrightarrow \mathbf{Can}_i(\alpha_1, \mathbf{PracPoss}_i(\alpha_2, \varphi))$
4. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha_1; \alpha_2, \varphi) \rightarrow \langle \mathrm{do}_i(\alpha_1) \rangle \mathbf{Can}_i(\alpha_2, \varphi)$ for α_1 accordant in $\mathbf{F}$
5. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\alpha_1; \alpha_2, \varphi) \leftrightarrow \mathbf{Cannot}_i(\alpha_1, \mathbf{PracPoss}_i(\alpha_2, \varphi))$
6. $\models^{\mathbf{b}} \mathbf{Can}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi}, \psi) \wedge \mathbf{K}_i\varphi \leftrightarrow \mathbf{Can}_i(\alpha_1, \psi) \wedge \mathbf{K}_i\varphi$

7. $\models^{\mathbf{b}} \mathbf{Can}_i(\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}, \psi) \wedge \mathbf{K}_i\neg\varphi \leftrightarrow \mathbf{Can}_i(\alpha_2, \psi) \wedge \mathbf{K}_i\neg\varphi$

8. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}, \psi) \wedge \mathbf{K}_i\varphi \leftrightarrow \mathbf{Cannot}_i(\alpha_1, \psi) \wedge \mathbf{K}_i\varphi$

9. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}, \psi) \wedge \mathbf{K}_i\neg\varphi \leftrightarrow \mathbf{Cannot}_i(\alpha_2, \psi) \wedge \mathbf{K}_i\neg\varphi$

10. $\models^{\mathbf{b}} \mathbf{Can}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \wedge \mathbf{K}_i\varphi \leftrightarrow \mathbf{Can}_i(\alpha, \mathbf{PracPoss}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi)) \wedge \mathbf{K}_i\varphi$

11. $\models^{\mathbf{b}} \mathbf{Can}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \wedge \mathbf{K}_i\varphi \rightarrow \langle \mathrm{do}_i(\alpha)\rangle \mathbf{Can}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi)$
    for α accordant in $\mathbf{F}$

12. $\models^{\mathbf{b}} \mathbf{Can}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \rightarrow \mathbf{K}_i(\varphi \vee \psi)$

13. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \wedge \mathbf{K}_i\neg\varphi \leftrightarrow \mathbf{K}_i(\neg\varphi \wedge \neg\psi)$

14. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \wedge \mathbf{K}_i\varphi \leftrightarrow \mathbf{Cannot}_i(\alpha; \texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}, \psi) \wedge \mathbf{K}_i\varphi$

Proposition 5.3 supports the claim about appropriateness of the Can-predicate and Cannot-predicate as formalising knowledge of practical possibilities of actions performed by rational agents. In particular items 6 through 9 and item 14 are genuine indications of the rationality of the agents that we formalised. Consider for example item 7. This item states that whenever an agent knows both that it has the practical possibility to bring about ψ by performing $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ and that the negation of the condition of $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ holds, it also knows that performing the else-part of the conditional composition provides the practical possibility to achieve ψ. Conversely, if agent i knows that it has the practical possibility to bring about ψ by performing α_2 while at the same time knowing that the proposition φ is false, then the agent knows that performing a conditional composition $\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi}$ would also bring about ψ, regardless of α_1. For since it knows that $\neg\varphi$ holds, it knows that this compositional composition comes down to the else-part α_2. Items 4 and 11 explicitly use the accordance of actions. For it is exactly this property of accordance that causes the agent's knowledge of its practical possibilities to persist under execution of the first part of the sequential composition in item 4 and the body of the while-loop in item 11.

In the following proposition we characterise the relation between the Can-predicate and the Cannot-predicate. Furthermore some properties are presented that concern the *end*-part of these predicates, i.e. the φ in $\mathbf{Can}_i(\alpha, \varphi)$ and $\mathbf{Cannot}_i(\alpha, \varphi)$.

Proposition 5.4 For all $\mathbf{b} \in$ bool, $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$ and $\varphi, \psi \in \mathrm{L}$ we have:

1. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi) \rightarrow \neg\mathbf{Can}_i(\alpha, \neg\varphi)$

2. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi) \rightarrow \neg\mathbf{Cannot}_i(\alpha, \varphi)$

3. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi) \rightarrow \mathbf{Cannot}_i(\alpha, \neg\varphi)$

4. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi \wedge \psi) \leftrightarrow \mathbf{Can}_i(\alpha, \varphi) \wedge \mathbf{Can}_i(\alpha, \psi)$

5. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\alpha, \varphi) \vee \mathbf{Cannot}_i(\alpha, \psi) \rightarrow \mathbf{Cannot}_i(\alpha, \varphi \wedge \psi)$

6. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi) \vee \mathbf{Can}_i(\alpha, \psi) \rightarrow \mathbf{Can}_i(\alpha, \varphi \vee \psi)$

7. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\alpha, \varphi \vee \psi) \leftrightarrow \mathbf{Cannot}_i(\alpha, \varphi) \wedge \mathbf{Cannot}_i(\alpha, \psi)$

8. $\models^{\mathbf{b}} \mathbf{Can}_i(\alpha, \varphi) \wedge \mathbf{K}_i[\mathrm{do}_i(\alpha)](\varphi \to \psi) \to \mathbf{Can}_i(\alpha, \psi)$

9. $\models^{\mathbf{b}} \mathbf{Cannot}_i(\alpha, \varphi) \wedge \mathbf{K}_i[\mathrm{do}_i(\alpha)](\psi \to \varphi) \to \mathbf{Cannot}_i(\alpha, \psi)$

Even more than Proposition 5.3 does Proposition 5.4 make out a case for the rationality of agents. Take for example item 3, which states that whenever an agent knows that it has the practical possibility to achieve φ by performing α it also knows that α does not provide for a means to achieve $\neg\varphi$. Items 4 through 7 deal with the decomposition of the end-part of the Can-predicate and the Cannot-predicate, which behaves as desired. Note that the reverse implication of item 5 is not valid: it is quite possible that even though an agent knows that α is not correct to bring about $(p \wedge \neg p)$ it might still be that it knows that α is correct for either p or $\neg p$. An analogous line of reasoning shows the invalidity of the reverse implication of item 6. Items 8 and 9 formalise that agents can extend their knowledge about their practical (im-)possibilities by combining it with their knowledge of the (conditional) results of actions.

6 Proof theory

Here we present a proof theory for the semantic framework defined in the previous section. In general the purpose of a proof theory is to provide a syntactic counterpart of the semantic notion of validity for a given interpretation and a given class of models. The idea is to define a predicate denoting deducibility, which holds for a given formula iff the formula is valid. This predicate is to be defined purely syntactically, i.e. it should depend only on the syntactic structure of formulas, without making any reference to semantic notions such as truth, validity, satisfiability etc. We present two such predicates, viz. $\vdash^{\mathbf{1}}$ and $\vdash^{\mathbf{0}}$, which characterise the notions of validity associated with $\models^{\mathbf{1}}$ and $\models^{\mathbf{0}}$, respectively. The definition of these predicates is based on a set of axioms and proof rules, which together constitute a proof system. The proof systems that we define deviate somewhat from the ones that are common in (modal) logics, the most notable difference being the use of infinitary proof rules. Given the relative rarity of this kind of rules, we feel that some explanation is justified.

6.1 Infinitary proof rules

The proof rules that are commonly employed in proof systems, are inference schemes of the form $\mathrm{P}_1, \ldots, \mathrm{P}_m$ / C, where the premises $\mathrm{P}_1, \ldots, \mathrm{P}_m$ and the conclusion C are elements of the language under consideration. Informally, a rule like this denotes that one may deduce C as soon as $\mathrm{P}_1, \ldots, \mathrm{P}_m$ have been deduced. An infinitary[3] proof rule is a rule containing an infinite number of premises. Although not very common, infinitary proof rules have been used in a number of proof systems: Hilbert used an infinitary proof rule in axiomatising number theory [16], Schütte uses infinitary proof rules in a number of systems [43], and both Kröger [26] and Goldblatt [10, 11] use infinitary proof rules in logics of action.

In finitary proof systems proofs can be carried out completely within the formal system. A proof is usually taken to be a finite sequence of formulas that are either axioms of the proof system or conclusions of proof rules applied to formulas that appear earlier in the sequence. Since finitary proof rules can be applied as soon as all of their finitely many premises have been deduced, there is no need to step outside of the formal system. In order to apply an infinitary

[3] We decided to follow the terminology of Goldblatt [10, 11] and refer to these rules as being infinitary. Other authors call these rules infinite [26, 43].

rule, a meta-logical investigation on the deducibility of the (infinitely many) premises needs to be carried out, which makes it in general impossible to carry out proofs completely within the proof system. As such, proofs are no longer 'schematically' constructed, and theorems are not recursively enumerable. However, there are also advantages associated with the use of infinitary proof rules. One such advantage is that for some systems *strong completeness* can be achieved using infinitary proof rules, whereas this is not possible using finitary proof rules (cf. [10, 43]). The notion of strong completeness implies that fewer sets of formulas are consistent, and in particular that sets of formulas that are seen to be inconsistent can also be proved to be so. After the presentation of the proof systems, we will return to the property of strong completeness in the presence of infinitary rules. Besides the possibility to achieve strong completeness when using infinitary proof rules, there are two other arguments that influenced our decision to use this kind of rule. The first of these is its intuitive acceptability. In particular when dealing with notions with an infinitary character, like for instance while-loops, infinitary proof rules provide a much better formalisation of human intuition on the nature of these notions than do finitary proof rules. The second, perhaps less convincing but certainly more compelling, argument is given by the fact that our attempts to come up with finitary axiomatisations remained unavailing.

6.2 Logics of capabilities

Before presenting the actual axiomatisations, we first make some notions precise that were already informally discussed above. An axiom is a schema in L. A proof rule is a schema of the form $\varphi_1, \varphi_2, \ldots / \psi$ where $\varphi_1, \varphi_2, \ldots, \psi$ are schemas in L. A proof system is a pair consisting of a set of axioms and a set of proof rules. As mentioned above, the presence of infinitary proof rules forces us to adopt a more abstract approach to the notions of deducibility and theorem than the one commonly employed in finitary proof systems. Usually, a formula φ is defined to be a theorem of some proof system if there exist a finite-length sequence of formulas of which φ is the last element and such that each formula in the sequence is either an instance of an axiom or the conclusion of a proof rule applied to earlier members of the sequence. An alternative formulation, which is equally usable in finitary and in infinitary proof systems, is to define φ to be a theorem of a proof system iff it belongs to the smallest subset of L containing all (instances of all) axioms and closed under the proof rules. This latter notion of deducibility is actually the one that we will employ here. We define a logic for a given proof system to be a subset of L containing all instances of the axioms of the proof system and closed under its proof rules. A formula is a theorem for a given proof system iff it is an element of the smallest logic for the proof system. These notions are formalised in Definitions 6.4 through 6.6.

To axiomatise the behaviour of while-loops we propose two infinitary rules. Both these rules are based on the idea to equate a repetitive composition $\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}$ with the infinite set $\{(\texttt{confirm}\,\varphi;\alpha)^k;\texttt{confirm}\,\neg\varphi \mid k \in \mathbb{N}\}$. The two proof rules take as their premises an infinite set of formulas built around this infinite set and have as their conclusion a formula built around $\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}$. To make this idea of 'building formulas around actions' explicit, we introduce the concept of *admissible forms*. The notion of admissible forms as given in Definition 6.1 is an extension of that used by Goldblatt in his language of program schemata [10]. In his investigation of infinitary proof rules, Kröger found that, in order to prove completeness, he needed rules in which the context of the while-loop and of the set $\{(\texttt{confirm}\,\varphi;\alpha)^k;\texttt{confirm}\,\neg\varphi \mid k \in \mathbb{N}\}$ is taken into account [26]. The concept of admissible forms provides an abstract generalisation of this idea of taking contexts into account.

Definition 6.1 The set of admissible forms for L, denoted by Afm(L), is defined by the following BNF.

$$\phi ::= \# \mid [\mathrm{do}_i(\alpha)]\phi \mid \mathbf{K}_i\phi \mid \psi \to \phi$$

where $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$ and $\psi \in \mathrm{L}$. We use ϕ as a typical element of Afm(L).

Usually 'admissible form' is abbreviated to 'afm'. By definition, each afm has a unique occurrence of the special symbol #. By instantiating this symbol with a formula from L, afms are turned into genuine formulas. If ϕ is an afm and $\psi \in \mathrm{L}$ is some formula we denote by $\phi(\psi)$ the formula that is obtained by replacing (the unique occurrence of) # in ϕ by ψ.

The following definition introduces two abbreviations that will be used in formulating the infinitary rules.

Definition 6.2 For all $\psi, \varphi \in \mathrm{L}$, $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$ and $l \in \mathbb{N}$ we define:

- $\psi_l(i, \varphi, \alpha) =^{\mathrm{def}} [\mathrm{do}_i((\mathtt{confirm}\,\varphi; \alpha)^l; \mathtt{confirm}\,\neg\varphi)]\psi$
- $\varphi_l(i, \alpha) =^{\mathrm{def}} \mathbf{A}_i((\mathtt{confirm}\,\varphi; \alpha)^l; \mathtt{confirm}\,\neg\varphi)$

The formulas introduced in Definition 6.2 are used to define the premises of the infinitary rules. The rule formalising the behaviour of while-loops with respect to results and opportunities has as premises all sentences in the infinite set $\{\phi(\psi_l(i, \varphi, \alpha)) \mid l \in \mathbb{N}\}$ for some $\phi \in \mathrm{Afm(L)}\}$. The conclusion of this rule is the formula $\phi([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$. Leaving the context provided by ϕ out of consideration, this rules intuitively states that if it is deducible that ψ holds after executing the actions $(\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi$, for every $k \in \mathbb{N}$, then it is also deducible that ψ holds after executing $\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$. The rule used in formalising the ability of agents for while-loops has as its premises the set $\phi(\neg(\varphi_l(i, \alpha)))$ for $l \in \mathbb{N}, \phi \in \mathrm{Afm(L)}$, and a conclusion $\phi(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})$. This rule states that whenever it is deducible that an agent i is not capable of performing any of the actions $(\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi$, where $k \in \mathbb{N}$, then it is also deducible that the agent is incapable of performing the while-loop itself. Or read in its contrapositive form, that an agent is able to perform a while-loop only if it is able to perform some finite-length sequence of confirmations and actions constituting the while-loop. As such, this rule is easily seen to be the proof-theoretic counterpart of the negated version of the (semantic) definition of $\mathtt{c}^{\mathbf{b}}$ for while-loops. For read in its negative form this semantic definition states that $\mathtt{c}^{\mathbf{b}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) = \mathbf{0}$ iff $\mathtt{c}^{\mathbf{b}}(i, \varphi_l(i, \alpha))(s) = \mathbf{0}$ for all $l \in \mathbb{N}$.

The axioms that are used to build the two proof systems are formulated using the necessity operator for actions, i.e. $[\mathrm{do}\_(\_)]\_$, rather than its dual $\langle \mathrm{do}\_(\_)\rangle\_$. The reason for this is essentially one of convenience: in proving completeness of the axiomatisations it turns out to be useful to deal with two necessity operators, viz. $\mathbf{K}\_$ and $[\mathrm{do}\_(\_)]\_$, to allow proofs by analogy. Since $[\mathrm{do}\_(\_)]\_$ and $\langle \mathrm{do}\_(\_)\rangle\_$ are inter-definable this does not create any essential differences.

Definition 6.3 The following axioms and proof rules are used to constitute the two proof systems that we consider here. Both the axioms as well as the premises and conclusions of the proof rules are to be taken as schemas in $i \in \mathrm{A}, \varphi, \psi \in \mathrm{L}$ and $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$. The ϕ occurring in the two infinitary rules ΩI and ΩIA is taken to be a meta-variable ranging over Afm(L).

A1. All propositional tautologies and their epistemic and dynamic instances

A2. $\mathbf{K}_i(\varphi \to \psi) \to (\mathbf{K}_i\varphi \to \mathbf{K}_i\psi)$

A3. $\mathbf{K}_i\varphi \to \varphi$

A4. $\mathbf{K}_i\varphi \to \mathbf{K}_i\mathbf{K}_i\varphi$

A5. $\neg\mathbf{K}_i\varphi \to \mathbf{K}_i\neg\mathbf{K}_i\varphi$

A6. $[\mathrm{do}_i(\alpha)](\varphi \to \psi) \to ([\mathrm{do}_i(\alpha)]\varphi \to [\mathrm{do}_i(\alpha)]\psi)$

A7. $[\mathrm{do}_i(\mathtt{confirm}\,\varphi)]\psi \leftrightarrow (\neg\varphi \vee \psi)$

A8. $[\mathrm{do}_i(\alpha_1;\alpha_2)]\varphi \leftrightarrow [\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\varphi$

A9. $[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi \leftrightarrow$
$([\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha_1)]\psi \wedge [\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi;\alpha_2)]\psi)$

A10. $[\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \leftrightarrow ([\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi)]\psi \wedge$
$[\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha)][\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$

A11. $[\mathrm{do}_i(\alpha)]\varphi \vee [\mathrm{do}_i(\alpha)]\neg\varphi$

A12. $\mathbf{A}_i\mathtt{confirm}\,\varphi \leftrightarrow \varphi$

A13$_\mathbf{1}$. $\mathbf{A}_i(\alpha_1;\alpha_2) \leftrightarrow \mathbf{A}_i\alpha_1 \wedge [\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2$

A13$_\mathbf{0}$. $\mathbf{A}_i(\alpha_1;\alpha_2) \leftrightarrow \mathbf{A}_i\alpha_1 \wedge \langle\mathrm{do}_i(\alpha_1)\rangle\mathbf{A}_i\alpha_2$

A14. $\mathbf{A}_i\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi} \leftrightarrow$
$(\mathbf{A}_i\mathtt{confirm}\,\varphi;\alpha_1 \vee \mathbf{A}_i\mathtt{confirm}\,\neg\varphi;\alpha_2)$

A15$_\mathbf{1}$. $\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \leftrightarrow (\mathbf{A}_i(\mathtt{confirm}\,\neg\varphi) \vee$
$(\mathbf{A}_i\mathtt{confirm}\,\varphi;\alpha \wedge [\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha)]\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}))$

A15$_\mathbf{0}$. $\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \leftrightarrow (\mathbf{A}_i(\mathtt{confirm}\,\neg\varphi) \vee$
$(\mathbf{A}_i\mathtt{confirm}\,\varphi;\alpha \wedge \langle\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha)\rangle\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}))$

R1. $\phi(\psi_l(i,\varphi,\alpha))$ all $l \in \mathbb{N}$ / $\phi([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$ ΩI

R2. $\phi(\neg(\varphi_l(i,\alpha)))$ all $l \in \mathbb{N}$ / $\phi(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})$ ΩIA

R3. $\varphi,\ \varphi \to \psi\,/\,\psi$ MP

R4. $\varphi\,/\,\mathbf{K}_i\varphi$ KN

R5. $\varphi\,/\,[\mathrm{do}_i(\alpha)]\varphi$ AN

Most of the axioms are fairly obvious, in particular given the discussion on the validities presented in Section 4. Rule R1, the Omega Iteration rule, is adopted from the axiomatisations given by Goldblatt [10, 11]. Both ΩI and rule R2, which is the Omega Iteration rule for Ability, were already discussed above. Rule R3 is the rule of Modus Ponens, well known from, and used in, both classical and modal logics. R4 and R5 are both instances of the rule of necessitation, which is known to hold for necessity operators. These rules state that whenever some formula is deducible, it is also deducible that an arbitrary agent knows the formula, and that all events have this formula among their conditional results, respectively. Axioms A2 and A6, and the rules R4 and R5 indicate that both knowledge and conditional results are formalised through normal modal operators.

The axioms and proof rules given above are used to define two different proof systems. One of these proof systems embodies the optimistic view on abilities in the counterfactual state of affairs, the other employs a pessimistic view.

Definition 6.4 The proof system $\Sigma_\mathbf{1}$ contains the axioms A1 through A12, A13$_\mathbf{1}$, A14, A15$_\mathbf{1}$ and the proof rules R1 through R5. The proof system $\Sigma_\mathbf{0}$ contains the axioms A1 through A12, A13$_\mathbf{0}$, A14, A15$_\mathbf{0}$ and the proof rules R1 through R5.

As mentioned above, a logic for a given proof system is a set encompassing the proof system.

Definition 6.5 A $\mathbf{b}$-logic is a set Λ that contains all the instances of the axioms of $\Sigma_{\mathbf{b}}$ and is closed under the proof rules of $\Sigma_{\mathbf{b}}$. The intersection of all $\mathbf{b}$-logics, which is itself a $\mathbf{b}$-logic, viz. the smallest one, is denoted by $\mathrm{LCap}_{\mathbf{b}}$. Whenever the underlying proof system is either irrelevant or clear from the context, we refer to a $\mathbf{b}$-logic simply as a logic.

Deducibility in a given proof system is now defined as being an element of the smallest logic for the proof system.

Definition 6.6 For Λ some logic, the unary predicate $\vdash^{\Lambda} \subseteq \mathrm{L}$ is defined by: $\vdash^{\Lambda} \varphi \Leftrightarrow \varphi \in \Lambda$. As an abbreviation we occasionally write $\vdash^{\mathbf{b}} \varphi$ for $\vdash^{\mathrm{LCap}_{\mathbf{b}}}$. Whenever $\vdash^{\Lambda} \varphi$ holds we say that φ is deducible in Λ or alternatively that φ is a theorem of Λ.

The proof systems $\Sigma_{\mathbf{1}}$ and $\Sigma_{\mathbf{0}}$ provide sound and complete axiomatisations of validity for $\models^{\mathbf{1}}$ and $\models^{\mathbf{0}}$ respectively. This is summarized in the following theorem, of which the proof is provided in the appendix.

Theorem 6.7 For $\mathbf{b} \in \mathrm{bool}$ and all $\varphi \in \mathrm{L}$ we have: $\vdash^{\mathbf{b}} \varphi \Leftrightarrow \models^{\mathbf{b}} \varphi$.

Besides the notion of deducibility *per se*, it is also interesting to look at deducibility from a set of premises. In modal logics one may distinguish two notions of deducibility from premises. In the first of these, the premises are considered to be additional axioms, on which also rules of necessitation may be applied. The second notion of deducibility allows necessitation only on the axioms of the proof system, and not on the premises. This latter notion of deducibility is perhaps the more natural one, and is in fact the one that we will concentrate on.

To account for deducibility from premises with respect to the alternative notion of deducibility as being an element of some set of formulas, we introduce the notion of a theory of a logic. Corresponding to the idea that the rules of necessitation are not to be applied on premises, we do not demand that a theory be closed under these rules. A formula is now defined to be deducible from some set of premises iff it is contained in every theory that encompasses the set of premises.

Definition 6.8 For Λ some logic, we define a Λ-theory to be any subset Θ of L that contains Λ and is closed under the rules ΩI, ΩIA, and MP.

Definition 6.9 Let Λ be some logic and $\Phi \cup \{\varphi\} \subseteq \mathrm{L}$. The binary relation $\vdash^{\Lambda} \subseteq \wp(\mathrm{L}) \times \mathrm{L}$ is defined by:

$$\Phi \vdash^{\Lambda} \varphi \Leftrightarrow \varphi \in \bigcap \{,\ \subseteq \mathrm{L} \mid \Phi \subseteq ,\ \text{ and } ,\ \text{ is a } \Lambda\text{-theory}\}$$

Whenever $\Phi \vdash^{\Lambda} \varphi$ we say that φ is deducible from Φ in Λ. A set $\Phi \subseteq \mathrm{L}$ is called Λ-inconsistent iff $\Phi \vdash^{\Lambda} \bot$, and Λ-consistent iff it is not Λ-inconsistent.

Given the ‘overloading’ of the symbol $\vdash^{\Lambda}$ as representing both deducibility *per se* and deducibility from premises, it is highly desirable that the two uses of this symbol coincide in the case that the set of premises is empty: deducibility from an empty set of premises should not differ from deducibility *per se*.

Proposition 6.10 For Λ some logic and $\varphi \in \mathrm{L}$ we have: $\vdash^{\Lambda} \varphi \Leftrightarrow \emptyset \vdash^{\Lambda} \varphi$.

As already mentioned before, using infinitary rules to describe the behaviour of while-loops allows one to achieve strong completeness, the notion which states that every consistent set of formulas is simultaneously satisfiable. Achieving strong completeness is in general not possible when just finitary rules are used. To see this consider the set $\Omega = \{[\mathrm{do}_i(a^k)]p \mid k \in \mathbb{N}\} \cup \{\langle \mathrm{do}_i(\mathtt{while}\, p\, \mathtt{do}\, a\, \mathtt{od})\rangle\top\}$. It is obvious that Ω is not satisfiable. For whenever $\mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(a^k)]p$ for all $k \in \mathbb{N}$ then execution of $\mathtt{while}\, p\, \mathtt{do}\, a\, \mathtt{od}$ does not terminate, and hence $\mathrm{M}, s \not\models \langle \mathrm{do}_i(\mathtt{while}\, p\, \mathtt{do}\, a\, \mathtt{od})\rangle\top$. However, when using just finitary rules to describe while-loops (like for instance the well-known Hoare rule [18]), the set Ω will be consistent. For when restricting oneself to finitary rules, consistency of an infinite set of formulas corresponds to consistency of each of its finite subsets. And in every axiomatisation that is to be sound, all finite subsets of Ω should be consistent, and therefore Ω itself is consistent. In the infinitary proof systems $\Sigma_{\mathbf{1}}$ and $\Sigma_{\mathbf{0}}$, $\perp$ is deducible from Ω, i.e. Ω is inconsistent. More generally, the property of strong completeness holds for both $\Sigma_{\mathbf{1}}$ and $\Sigma_{\mathbf{0}}$.

Proposition 6.11 The proof systems $\Sigma_{\mathbf{1}}$ and $\Sigma_{\mathbf{0}}$ are strongly complete, i.e. every set $\Phi \subseteq \mathrm{L}$ that is $\mathrm{LCap}_{\mathbf{b}}$-consistent is $\models^{\mathbf{b}}$-satisfiable.

Just as deducibility *per se* is the proof theoretic counterpart of the semantic notion of validity, there is also a semantic counterpart to the notion of deducibility from premises.

Proposition 6.12 For $\mathbf{b} \in \mathrm{bool}$, $\Phi \subseteq \mathrm{L}$ and $\varphi \in \mathrm{L}$ we have:

- $\Phi \vdash^{\mathbf{b}} \varphi \Leftrightarrow \Phi \models^{\mathbf{b}} \varphi$

where $\Phi \models^{\mathbf{b}} \varphi$ iff $\mathrm{M}, s \models^{\mathbf{b}} \Phi$ implies $\mathrm{M}, s \models^{\mathbf{b}} \varphi$ for all $\mathrm{M} \in \mathbf{M}$ with state s.

In the light of the strong completeness property, Proposition 6.12 is not very surprising. In fact, the left-to-right implication is a direct consequence of the strong completeness property. The right-to-left implication follows from the observation that the set of formulas that is satisfied in some world forms a theory.

7 Summary and conclusions

In this paper we introduced the KARO-framework, a formal framework based on a combination of various modal logics that can be used to formalise agents. After a somewhat philosophical exposition on knowledge, actions and events, we presented two formal systems, both belonging to the KARO-framework, that share a common language and a common class of models but that differ in the interpretation of dynamic and ability formulas. The language common to the two systems is a propositional, multi-modal, exogenous language, containing modalities representing knowledge, opportunity and result, and an operator formalising ability. The models that are used to interpret formulas from the language L are Kripke-style possible worlds models. These models interpret knowledge by means of an accessibility relation on worlds; opportunity, result and ability are interpreted using designated functions. We explained our intuition on the composite behaviour of results, opportunities and abilities, and presented two formal interpretations that comply with this intuition. These interpretations differ in their treatment of abilities of agents for sequentially composed actions. We considered various properties of knowledge and action in the KARO-framework. In defining some of these properties we used the notions of schemas, frames and correspondences. Using the various modalities present in the framework, we proposed a formalisation of the knowledge of agents about their practical possibilities, a notion which captures an important aspect of agency, particularly in the context of planning agents. We presented two proof systems that syntactically characterise the notion of validity in the two interpretations that we defined. The most remarkable aspect of these proof systems is the use of infinitary proof rules, which on the one hand allows for a better correspondence between the semantic notion of validity and its syntactic counterpart, and on the other hand forces one to generalise the usual notions of proof and theorem.

In the KARO-framework we proposed two definitions for the ability of agents to execute a sequentially composed action $\alpha_1;\alpha_2$ in cases where execution of α_1 leads to the counterfactual state of affairs. The simplicity of these definitions, both at a conceptual and at a technical level, may lead to counterintuitive situations. Recall that using the so-called optimistic approach it is possible that an agent is considered to be capable of performing $\alpha;\texttt{fail}$, whereas in the pessimistic approaches agents may be declared unable to perform $\alpha;\texttt{skip}$, for $\alpha \in \mathrm{Ac}$. A more realistic approach would be not to treat all actions equally, but instead to determine for each action individually whether it makes sense to declare an agent (un)able to perform the action in the counterfactual state of affairs. One way to formalise this consists of extending the models from $\mathbf{M}$ with an additional function $\mathtt{t} : \mathrm{A} \times \mathrm{Ac} \to \mathrm{S} \to \mathrm{S}$ which is such that $\mathtt{t}(i,\alpha)(s) = \mathtt{r}(i,\alpha)(s)$ whenever $\mathtt{r}(i,\alpha)(s) \neq \emptyset$. Hence in the case that $\mathtt{r}(i,\alpha)(s) \neq \emptyset$, $\mathtt{t}(i,\alpha)(s)$ equals $\mathtt{r}(i,\alpha)(s)$ and in other cases $\mathtt{t}(i,\alpha)(s)$ is definitely not empty. The function $\mathtt{t}$ denotes the outcome of actions when 'abstracting away' from opportunities, so to speak. The ability for the sequential composition is then defined by

$$\mathtt{c}(i,\alpha_1;\alpha_2)(s) = \mathbf{1} \Leftrightarrow \mathtt{c}(i,\alpha_1)(s) = \mathbf{1} \,\&\, \mathtt{c}(i,\alpha_2)(\mathtt{t}(i,\alpha_1)(s)) = \mathbf{1}$$

Applying this definition implies that $\mathbf{A}_i\alpha_1;\texttt{fail}$ is no longer satisfiable, and that $\mathbf{A}_i\alpha_1;\texttt{skip}$ holds in cases where $\mathbf{A}_i\alpha_1$ is true, regardless of the truth of $\langle \mathrm{do}_i(\alpha_1)\rangle\top$. A special instantiation of this approach corresponds to the idea that abilities of agents do not tend to change. Therefore it could seem reasonable to assume that agents retain their abilities when ending up in the counterfactual state of affairs. Formally this can be brought about by demanding $\mathtt{t}(i,\alpha)(s)$ to equate s in cases where $\mathtt{r}(i,\alpha)(s) = \emptyset$. Since this is but a special case of the general idea discussed above, it also avoids the counterintuitive situations where agents are declared to be able to do $\alpha;\texttt{fail}$ or unable to do $\alpha;\texttt{skip}$.

Acknowledgements

The comments of an anonymous referee on a previous version of this paper were very useful. This work was partially supported by ESPRIT WG No. 23531 ('FIREworks').

A A proof of soundness and completeness

Below we prove the soundness and completeness of deducibility in $\mathrm{LCap}_{\mathbf{b}}$ for $\models^{\mathbf{b}}$-validity in $\mathbf{M}$. As far as we know, this is one of the very few proofs of completeness that concerns a proof system in which both knowledge and actions are dealt with, and it is probably the very first in which abilities are also taken into consideration.

Rather than restricting ourselves to $\mathrm{LCap}_{\mathbf{b}}$ we will for the greater part consider general logics, culminating in a very general and rather powerful result from which the soundness and completeness proof for $\mathrm{LCap}_{\mathbf{b}}$ can be derived as a corollary. Globally, the proof given below can be split into three parts. In the first part of the proof, canonical models are constructed for the logics induced by the proof systems $\Sigma_{\mathbf{1}}$ and $\Sigma_{\mathbf{0}}$. The possible worlds of these canonical models are given by so-called maximal theories. In the second part, the truth-theorem is proved, which states that truth in a possible world of a canonical model corresponds to being an element of the maximal theory that constitutes the possible world. In the last, and almost trivial, part of the proof it is shown how the general truth-theorem implies soundness and completeness of $\mathrm{LCap}_{\mathbf{b}}$ for $\mathbf{b}$-validity in $\mathbf{M}$.

The definition of canonical models as we give it is, as far as actions and dynamic constructs are concerned, based on the construction given by Goldblatt [10]. The proof of the truth-theorem is inspired by the one given by Spruit [47] to show completeness of the Segerberg axiomatisation for propositional dynamic logic. Due to the fact that formulas and actions are strongly related, the subformula or subaction relation does not provide an adequate support for induction in the proof of the truth-theorem. Instead a fairly complex ordering is used, well-foundedness of which is proved using some very powerful (and partly automated) techniques that are well-known from the theory of Term Rewriting Systems [7, 23].

Some preliminary definitions, propositions and lemmas are needed before the canonical models can be constructed.

Proposition A.1 For all $\mathrm{M} \in \mathbf{M}$ with state s and all $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$, $\varphi \in \mathrm{L}$ and $\phi \in \mathrm{Afm}(\mathrm{L})$ we have:

- $\mathrm{M}, s \models^{\mathbf{b}} \phi([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$ iff for all $l \in \mathbb{N}$, $\mathrm{M}, s \models^{\mathbf{b}} \phi(\psi_l(i, \alpha, \varphi))$
- $\mathrm{M}, s \models^{\mathbf{b}} \phi(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})$ iff for all $l \in \mathbb{N}$, $\mathrm{M}, s \models^{\mathbf{b}} \phi(\neg(\varphi_l(i, \alpha)))$

Proof: We prove both items by induction on the structure of ϕ.

- Let $\mathrm{M} \in \mathbf{M}$ with state s, and $i \in \mathrm{A}$, $\varphi, \psi \in \mathrm{L}$ and $\alpha \in \mathrm{Ac}$ be arbitrary.

  1. $\phi = \#$:

$$\begin{aligned}
& \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \psi \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \psi \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha)^l; \mathtt{confirm}\,\neg\varphi)(s) \\
& \qquad \text{for all } l \in \mathbb{N} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{confirm}\,\varphi; \alpha)^l; \mathtt{confirm}\,\neg\varphi)]\psi \text{ for all } l \in \mathbb{N} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} \psi_l(i, \varphi, \alpha) \text{ for all } l \in \mathbb{N}
\end{aligned}$$

  2. $\phi = \mathbf{K}_i\phi'$:

$$\begin{aligned}
& \mathrm{M}, s \models^{\mathbf{b}} (\mathbf{K}_i\phi')([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi) \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} \mathbf{K}_i(\phi'([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \phi'([\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi) \text{ for all } t \in \mathrm{S} \text{ such that } (s, t) \in \mathrm{R}(i) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \phi'(\psi_l(i, \alpha, \varphi)) \text{ for all } l \in \mathbb{N}, \\
& \qquad \text{for all } t \in \mathrm{S} \text{ such that } (s, t) \in \mathrm{R}(i) \text{ (by induction hypothesis)} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} \mathbf{K}_i\phi'(\psi_l(i, \alpha, \varphi)) \text{ for all } l \in \mathbb{N} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} (\mathbf{K}_i\phi')(\psi_l(i, \alpha, \varphi)) \text{ for all } l \in \mathbb{N}
\end{aligned}$$

3. The cases where $\phi = [\mathrm{do}_i(\beta)]\phi'$ and $\phi = \psi' \to \phi'$ are analogous to the case where $\phi = \mathbf{K}_i\phi'$.

- Let again $\mathrm{M} \in \mathbf{M}$ with state s, $i \in \mathrm{A}$, $\varphi \in \mathrm{L}=$ and $\alpha \in \mathrm{Ac}$ be arbitrary.

1. $\phi = \#$:

$$\begin{aligned}
& \mathrm{M}, s \models^{\mathbf{b}} \neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \\
\Leftrightarrow\ & \mathrm{not}(\mathrm{M}, s \models^{\mathbf{b}} \mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}) \\
\Leftrightarrow\ & \mathrm{not}(\exists k \in \mathbb{N}(\mathtt{c}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s) = \mathbf{1})) \\
\Leftrightarrow\ & \forall k \in \mathbb{N}(\mathrm{not}(\mathtt{c}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s) = \mathbf{1})) \\
\Leftrightarrow\ & \forall k \in \mathbb{N}(\mathrm{not}(\mathrm{M}, s \models^{\mathbf{b}} \varphi_k(i, \alpha))) \\
\Leftrightarrow\ & \forall k \in \mathbb{N}(\mathrm{M}, s \models^{\mathbf{b}} \neg(\varphi_k(i, \alpha_1)))
\end{aligned}$$

2. $\phi = [\mathrm{do}_i(\beta)]\phi'$:

$$\begin{aligned}
& \mathrm{M}, s \models^{\mathbf{b}} ([\mathrm{do}_i(\beta)]\phi')(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}) \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\beta)](\phi'(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \phi'(\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}) \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, \beta)(s) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \phi'(\neg(\varphi_l(i, \alpha))) \text{ for all } l \in \mathbb{N}, \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, \beta)(s) \\
& \qquad\qquad \text{(by induction hypothesis)} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\beta)](\phi'(\neg(\varphi_l(i, \alpha)))) \text{ for all } l \in \mathbb{N} \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} ([\mathrm{do}_i(\beta)]\phi')(\neg(\varphi_l(i, \alpha))) \text{ for all } l \in \mathbb{N}
\end{aligned}$$

3. The cases where $\phi = \mathbf{K}_i\phi'$ and $\phi = (\psi' \to \phi')$ are analogous to the case where $\phi = [\mathrm{do}_i(\beta)]\phi'$.

⊠

Proposition A.2 If $\mathrm{M} \in \mathbf{M}$ is a well-defined model from $\mathbf{M}$, then $\Lambda^{\mathrm{M}}_{\mathbf{b}} =^{\mathrm{def}} \{\varphi \in \mathrm{L} \mid \mathrm{M} \models^{\mathbf{b}} \varphi\}$ is a **b**-logic.

Proof: We need to check for a given model $\mathrm{M} \in \mathbf{M}$ that the axioms of $\Sigma_{\mathbf{b}}$ are valid in M and that M is validity-preserving for the proof rules of $\Sigma_{\mathbf{b}}$. The validity of the axioms A1–A9 and A12–A14 is easily checked. Axiom A10 follows from the determinism of all actions as stated in Corollary 4.12. Axiom $\mathrm{A15}_{\mathbf{1}}$ is shown in Proposition 4.3, and $\mathrm{A15}_{\mathbf{0}}$ is shown analogously. The validity-preservingness of M for the rules R1 and R2 follows from Proposition A.1; M is easily seen to be validity-preserving for the other rules. As an example we show here the validity of axiom A10.

$$\begin{aligned}
& \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \psi \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) \\
\Leftrightarrow\ & \mathrm{M}, t \models^{\mathbf{b}} \psi \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, \mathtt{confirm}\,\neg\varphi)(s) \text{ and} \\
& \mathrm{M}, t \models^{\mathbf{b}} \psi \text{ for all } t \in \mathrm{S} \text{ such that } t = \mathtt{r}^{\mathbf{b}}(i, (\mathtt{confirm}\,\varphi; \alpha); \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi)]\psi \text{ and} \\
& \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i((\mathtt{confirm}\,\varphi; \alpha); \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \\
\Leftrightarrow\ & \mathrm{M}, s \models^{\mathbf{b}} [\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi)]\psi \wedge \\
& \qquad\qquad [\mathrm{do}_i(\mathtt{confirm}\,\varphi; \alpha)][\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi
\end{aligned}$$

⊠

Proposition A.3 *Let* Λ *be a logic. The following properties are shared by all* Λ*-theories* $,$ *, for all* $\varphi, \psi \in \mathrm{L}$*,* $i \in \mathrm{A}$*,* $\alpha \in \mathrm{Ac}$ *and all* $\phi \in \mathrm{Afm}(\mathrm{L})$*:*

1. $\top \in ,$
2. *if* $, \vdash^{\Lambda} \varphi$ *then* $\varphi \in ,$
3. *if* $\vdash^{\Lambda} (\varphi \to \psi)$ *and* $\varphi \in ,$ *then* $\psi \in ,$
4. $,$ *is* Λ*-consistent iff* $\bot \notin ,$ *iff* $, \neq \mathrm{L}$
5. $(\varphi \wedge \psi) \in ,$ *iff* $\varphi \in ,$ *and* $\psi \in ,$
6. *if* $\varphi \in ,$ *or* $\psi \in ,$ *then* $(\varphi \vee \psi) \in ,$
7. $\phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi) \in ,$ *iff* $\{\phi(\psi_l(i, \varphi, \alpha)) \mid l \in \mathbb{N}\} \subseteq ,$
8. $\phi(\neg \mathbf{A}_i \texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}) \in ,$ *iff* $\{\phi(\neg(\varphi_l(i, \alpha))) \mid l \in \mathbb{N}\} \subseteq ,$

Proof: The items 1 to 6 are fairly standard, and are proved by Goldblatt [10]. The cases 7 and 8 follow from the fact that theories contain the axioms A10 and A15$_{\mathrm{b}}$ and are closed under ΩI and ΩIA.
⊠

Definition A.4 *Let* Λ *be a logic. A maximal* Λ*-theory is a consistent* Λ*-theory* $,$ *such that* $\varphi \in ,$ *or* $\neg\varphi \in ,$ *for all* $\varphi \in \mathrm{L}$*.*

Proposition A.5 *The following properties are shared by all maximal* Λ*-theories* $,$ *, for* Λ *some logic, and* $\varphi, \psi \in \mathrm{L}$*.*

1. $\bot \notin ,$
2. *exactly one of* φ *and* $\neg\varphi$ *belongs to* $,$ *, for all* $\varphi \in \mathrm{L}$
3. $(\varphi \vee \psi) \in ,$ *iff* $\varphi \in ,$ *or* $\psi \in ,$

Proposition A.6 *For* Λ *a logic and all* $\varphi, \psi \in \mathrm{L}$*,* $\Phi, \Psi \subseteq \mathrm{L}$*,* $i \in \mathrm{A}$*,* $\alpha \in \mathrm{Ac}$ *and* $\phi \in \mathrm{Afm}(\mathrm{L})$ *we have:*

1. *if* $\varphi \in \Phi$ *then* $\Phi \vdash^{\Lambda} \varphi$
2. *if* $\Phi \vdash^{\Lambda} \varphi$ *and* $\Phi \subseteq \Psi$ *then* $\Psi \vdash^{\Lambda} \varphi$
3. $\vdash^{\Lambda} \varphi$ *iff* $\emptyset \vdash^{\Lambda} \varphi$
4. *if* $\Phi \vdash^{\Lambda} (\varphi \to \psi)$ *and* $\Phi \vdash^{\Lambda} \varphi$ *then* $\Phi \vdash^{\Lambda} \psi$
5. *if* $\Phi \vdash^{\Lambda} \phi(\psi_l(i, \varphi, \alpha))$ *for all* $l \in \mathbb{N}$ *then* $\Phi \vdash^{\Lambda} \phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi)$
6. *if* $\Phi \vdash^{\Lambda} \phi(\neg(\varphi_l(i, \alpha)))$ *for all* $l \in \mathbb{N}$ *then* $\Phi \vdash^{\Lambda} \phi(\neg \mathbf{A}_i \texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})$

Theorem A.7 (The deduction theorem) *For* Λ *some logic and all* $\varphi, \psi \in \mathrm{L}$ *and* $\Phi \subseteq \mathrm{L}$ *we have that* $\Phi \cup \{\varphi\} \vdash^{\Lambda} \psi$ *iff* $\Phi \vdash^{\Lambda} (\varphi \to \psi)$*.*

Proof: We will prove the 'iff' by proving two implications:

- '$\Leftarrow$' This case follows directly from items 1, 2, and 4 of Proposition A.6.
- '$\Rightarrow$' Assume that $\Phi \cup \{\varphi\} \vdash^{\Lambda} \psi$. Let , $=^{\text{def}} \{\rho \in \mathrm{L} \mid \Phi \vdash^{\Lambda} (\varphi \to \rho)\}$. We have to show that $\psi \in$, . For this it suffices to show that , is a Λ-theory containing $\Phi \cup \{\varphi\}$. We show here that , is closed under ΩIA; the proof of the other properties is easy and left to the reader. Assume that $\{\phi(\neg(\varphi'_l(i,\alpha))) \mid l \in \mathbb{N}\} \subseteq$, . Then $\Phi \vdash^{\Lambda} (\varphi \to \phi(\neg(\varphi'_l(i,\alpha))))$ for all $l \in \mathbb{N}$. Applying case 6 of Proposition A.6 to the set $\{(\varphi \to \phi(\neg(\varphi'_l(i,\alpha)))) \mid l \in \mathbb{N}\}$ yields $\Phi \vdash^{\Lambda} (\varphi \to \phi(\neg\mathbf{A}_i\mathtt{while}\,\varphi'\,\mathtt{do}\,\alpha\,\mathtt{od}))$, hence $\phi(\neg\mathbf{A}_i\mathtt{while}\,\varphi'\,\mathtt{do}\,\alpha\,\mathtt{od}) \in$, . Thus , is closed under ΩIA.

⊠

Corollary A.8 *For Λ some logic and all $\varphi \in \mathrm{L}$ and $\Phi \subseteq \mathrm{L}$ we have:*

- $\Phi \cup \{\varphi\}$ *is Λ-consistent iff* $\Phi \not\vdash^{\Lambda} \neg\varphi$
- $\Phi \cup \{\neg\varphi\}$ *is Λ-consistent iff* $\Phi \not\vdash^{\Lambda} \varphi$

Definition A.9 *For $\Phi \subseteq \mathrm{L}$, $i \in \mathcal{A}$ and $\alpha \in \mathrm{Ac}$ we define:*

- $\Phi/\mathbf{K}_i =^{\text{def}} \{\varphi \in \mathrm{L} \mid \mathbf{K}_i\varphi \in \Phi\}$
- $\mathbf{K}_i\Phi =^{\text{def}} \{\mathbf{K}_i\varphi \in \mathrm{L} \mid \varphi \in \Phi\}$
- $\Phi/[\mathrm{do}_i(\alpha)] =^{\text{def}} \{\varphi \in \mathrm{L} \mid [\mathrm{do}_i(\alpha)]\varphi \in \Phi\}$
- $[\mathrm{do}_i(\alpha)]\Phi =^{\text{def}} \{[\mathrm{do}_i(\alpha)]\varphi \in \mathrm{L} \mid \varphi \in \Phi\}$

Proposition A.10 *For Λ some logic and all $\varphi \in \mathrm{L}$, $i \in \mathcal{A}$ and $\Phi \subseteq \mathrm{L}$ we have:*

- *if* $\Phi \vdash^{\Lambda} \varphi$ *then* $\mathbf{K}_i\Phi \vdash^{\Lambda} \mathbf{K}_i\varphi$
- *if* $\Phi \vdash^{\Lambda} \varphi$ *then* $[\mathrm{do}_i(\alpha)]\Phi \vdash^{\Lambda} [\mathrm{do}_i(\alpha)]\varphi$

Proof: We show the first case; the second case is completely analogous. So let , be a Λ-theory such that $\mathbf{K}_i\Phi \subseteq$, . We need to show that $\mathbf{K}_i\varphi \in$, . Let $\Delta =^{\text{def}}$, $/\mathbf{K}_i$. Since $\Phi \vdash^{\Lambda} \varphi$, it suffices to show that Δ is a Λ-theory containing Φ. Then $\varphi \in \Delta$ and hence $\mathbf{K}_i\varphi \in$, .

1. $\Phi \subseteq \Delta$: If $\psi \in \Phi$, then $\mathbf{K}_i\psi \in$, and hence $\psi \in \Delta$.
2. Δ contains Λ: If $\vdash^{\Lambda} \psi$, then by NK, $\vdash^{\Lambda} \mathbf{K}_i\psi$ and, since , is a Λ-theory, then $\mathbf{K}_i\psi \in$, , which implies $\psi \in \Delta$.
3. Δ is closed under MP, ΩI and ΩIA.
   - MP: If $\psi \in \Delta$ and $(\psi \to \psi_1) \in \Delta$, then $\mathbf{K}_i\psi \in$, and $\mathbf{K}_i(\psi \to \psi_1) \in$, . Since , contains axiom A2, this implies $\mathbf{K}_i\psi_1 \in$, and hence $\psi_1 \in \Delta$.
   - ΩI: If $\{\phi(\psi_l(j,\varphi',\alpha)) \mid l \in \mathbb{N}\} \subseteq \Delta$, then $\{\mathbf{K}_i\phi(\psi_l(j,\varphi',\alpha)) \mid l \in \mathbb{N}\} \subseteq$, . Applying ΩI to the set $\{\mathbf{K}_i\phi(\psi_l(j,\varphi',\alpha)) \mid l \in \mathbb{N}\}$ yields $\mathbf{K}_i\phi([\mathrm{do}_j(\mathtt{while}\,\varphi'\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi) \in$, , and hence $\phi([\mathrm{do}_j(\mathtt{while}\,\varphi'\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi) \in \Delta$.

- ΩIA: If $\{\phi(\neg(\varphi'_l(j,\alpha))) \mid l \in \mathbb{N}\} \subseteq \Delta$, then $\{\mathbf{K}_i\phi(\neg(\varphi'_l(j,\alpha))) \mid l \in \mathbb{N}\} \subseteq$, . Applying ΩIA to $\{\mathbf{K}_i\phi(\neg(\varphi'_l(j,\alpha))) \mid l \in \mathbb{N}\}$ yields $\mathbf{K}_i\phi(\neg\mathbf{A}_j\texttt{while}\,\varphi'\,\texttt{do}\,\alpha\,\texttt{od}) \in$, , and hence $\phi(\neg\mathbf{A}_j\texttt{while}\,\varphi'\,\texttt{do}\,\alpha\,\texttt{od}) \in \Delta$.

It follows that Δ is closed under MP, ΩI and ΩIA.

Since Δ contains Λ and is closed under MP, ΩI and ΩIA it follows that Δ is a Λ-theory.
⊠

Corollary A.11 *Let Λ be some logic. For all Λ-theories , , and for $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$ and $\varphi \in \mathrm{L}$ we have:*

- $\mathbf{K}_i\varphi \in$, *iff* , $/\mathbf{K}_i \vdash^{\Lambda} \varphi$
- $[\mathrm{do}_i(\alpha)]\varphi \in$, *iff* , $/[\mathrm{do}_i(\alpha)] \vdash^{\Lambda} \varphi$

Proposition A.12 *Let Λ be some logic. For all maximal Λ-theories , we have that if , $/[\mathrm{do}_i(\alpha)]$ is Λ-consistent then , $/[\mathrm{do}_i(\alpha)]$ is a maximal Λ-theory.*

Proof: Suppose that , $/[\mathrm{do}_i(\alpha)]$ is Λ-consistent. We show that , $/[\mathrm{do}_i(\alpha)]$ is a Λ-theory and that for all $\varphi \in \mathrm{L}$, either $\varphi \in$, $/[\mathrm{do}_i(\alpha)]$ or $\neg\varphi \in$, $/[\mathrm{do}_i(\alpha)]$. Since by assumption , $/[\mathrm{do}_i(\alpha)]$ is consistent, this suffices to conclude that , $/[\mathrm{do}_i(\alpha)]$ is a maximal Λ-theory.

1. , $/[\mathrm{do}_i(\alpha)]$ contains Λ: If $\vdash^{\Lambda} \varphi$ then by NA, $\vdash^{\Lambda} [\mathrm{do}_i(\alpha)]\varphi$, and, since , is a Λ-theory, $[\mathrm{do}_i(\alpha)]\varphi \in$, . This implies that $\varphi \in$, $/[\mathrm{do}_i(\alpha)]$.
2. , $/[\mathrm{do}_i(\alpha)]$ is closed under MP, ΩI and ΩIA:
    - MP: Assume that $(\varphi \to \psi) \in$, $/[\mathrm{do}_i(\alpha)]$ and $\varphi \in$, $/[\mathrm{do}_i(\alpha)]$. Then $[\mathrm{do}_i(\alpha)](\varphi \to \psi) \in$, and $[\mathrm{do}_i(\alpha)]\varphi \in$, , which implies, since , contains A6 and is closed under MP, that $[\mathrm{do}_i(\alpha)]\psi \in$, . This implies that $\psi \in$, $/[\mathrm{do}_i(\alpha)]$.
    - ΩI: If $\{\phi(\psi_l(j,\varphi,\beta)) \mid l \in \mathbb{N}\} \subseteq$, $/[\mathrm{do}_i(\alpha)]$, then $\{[\mathrm{do}_i(\alpha)]\phi(\psi_l(j,\varphi,\beta)) \mid l \in \mathbb{N}\} \subseteq$, . Applying ΩI to the set of afms $\{[\mathrm{do}_i(\alpha)]\phi(\psi_l(j,\varphi,\beta)) \mid l \in \mathbb{N}\}$, yields that $[\mathrm{do}_i(\alpha)]\phi([\mathrm{do}_j(\texttt{while}\,\varphi\,\texttt{do}\,\beta\,\texttt{od})]\psi) \in$, , and $\phi([\mathrm{do}_j(\texttt{while}\,\varphi\,\texttt{do}\,\beta\,\texttt{od})]\psi) \in$, $/[\mathrm{do}_i(\alpha)]$.
    - ΩIA: Let $\{\phi(\neg(\varphi_l(j,\beta))) \mid l \in \mathbb{N}\} \subseteq$, $/[\mathrm{do}_i(\alpha)]$. Then $\{[\mathrm{do}_i(\alpha)]\phi(\neg(\varphi_l(j,\beta))) \mid l \in \mathbb{N}\} \subseteq$, . By applying ΩI to the set $\{[\mathrm{do}_i(\alpha)]\phi(\neg(\varphi_l(j,\beta))) \mid l \in \mathbb{N}\}$ it follows that $[\mathrm{do}_i(\alpha)]\phi(\neg\mathbf{A}_j\texttt{while}\,\varphi\,\texttt{do}\,\beta\,\texttt{od}) \in$, . Hence $\phi(\neg\mathbf{A}_j\texttt{while}\,\varphi\,\texttt{do}\,\beta\,\texttt{od}) \in$, $/[\mathrm{do}_i(\alpha)]$.
3. Since , is a theory, , contains axiom A11: $[\mathrm{do}_i(\alpha)]\varphi \vee [\mathrm{do}_i(\alpha)]\neg\varphi$ for all i, α and φ. Since , is maximal, $[\mathrm{do}_i(\alpha)]\varphi \in$, or $[\mathrm{do}_i(\alpha)]\neg\varphi \in$, for all i, α and φ. But this implies that $\varphi \in$, $/[\mathrm{do}_i(\alpha)]$ or $\neg\varphi \in$, $/[\mathrm{do}_i(\alpha)]$, for all $\varphi \in \mathrm{L}$.

By items 1, 2, and 3 it follows that , $/[\mathrm{do}_i(\alpha)]$ is a maximal Λ-theory if , $/[\mathrm{do}_i(\alpha)]$ is Λ-consistent.
⊠

Definition A.13 *For Λ some logic, the set S_Λ is defined by $\mathrm{S}_\Lambda =^{\mathrm{def}} \{, \subseteq \mathrm{L} \mid$, is a maximal Λ-theory$\}$.*

Proposition A.14 *For Λ some logic and all $\Phi \subseteq \mathrm{L}$ and $\varphi \in \mathrm{L}$ we have:*

- $\Phi \vdash^{\Lambda} \varphi$ *iff for all* , $\in \mathrm{S}_{\Lambda}$ *such that* $\Phi \subseteq$, *holds that* $\varphi \in$,
- $\vdash^{\Lambda} \varphi$ *iff for all* , $\in \mathrm{S}_{\Lambda}$ *holds that* $\varphi \in$,

Proof: The second item follows by instantiating the first item with $\Phi = \emptyset$ and using item 3 of Proposition A.6. We show the first item by proving two implications.

'$\Rightarrow$' By definition of $\Phi \vdash^{\Lambda} \varphi$.

'$\Leftarrow$' We show: if $\Phi \not\vdash^{\Lambda} \varphi$ then some , $\in \mathrm{S}_{\Lambda}$ exists such that $\Phi \subseteq$, and $\varphi \notin$, . We construct a , that satisfies this demand. To this end, we start by making an enumeration $\rho_0, \rho_1, \ldots$ of the formulas of L. Using this enumeration, the increasing sequence of sets $,_l \subseteq \mathrm{L}$ is for $l \in \mathbb{N}$ inductively defined as follows:

1. $,_0 = \Phi \cup \{\neg\varphi\}$
2. Assume that $,_k$ has been defined. The set $,_{k+1}$ is defined by the following algorithm, written in a high-level programming language pseudocode:

```
if     $,_k \vdash^{\Lambda} \rho_k$ then $,_{k+1} = ,_k \cup \{\rho_k\}$
elsif  $\rho_k$ is of the form $\phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi)$
then   $,_{k+1} = ,_k \cup \{\neg\phi(\psi_j(i,\varphi,\alpha))\} \cup \{\neg\rho_k\}$,
       where $j$ is the least number such that $,_k \not\vdash^{\Lambda} \phi(\psi_j(i,\varphi,\alpha))$
       (this $j$ exists since otherwise application of $\Omega$I would yield $,_k \vdash^{\Lambda} \rho_k$)
elsif  $\rho_k$ is of the form $\phi(\neg\mathbf{A}_i\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})$
then   $,_{k+1} = ,_k \cup \{\neg\phi(\neg(\varphi_j(i,\alpha)))\} \cup \{\neg\rho_k\}$,
       where $j$ is the least number such that $,_k \not\vdash^{\Lambda} \phi(\neg(\varphi_j(i,\alpha)))$
       (this $j$ exists since otherwise application of $\Omega$IA would yield $,_k \vdash^{\Lambda} \rho_k$)
else   $,_{k+1} = ,_k \cup \{\neg\rho_k\}$
fi
```

Now , is defined by , $=^{\mathrm{def}} \cup_{l \in \mathbb{N}} ,_l$. We show that , is a maximal Λ-theory.

Lemma A.15 *The set* $,_l$ *is* Λ*-consistent for all* $l \in \mathbb{N}$*.*

Proof: We prove the lemma by induction on l. Since $\Phi \not\vdash^{\Lambda} \varphi$, we have that $\Phi \cup \{\neg\varphi\} = ,_0$ is Λ-consistent by Corollary A.8. Now assume that $,_k$ is consistent. Consider the four possibilities for the definition of $,_{k+1}$:

1. If $,_k \vdash^{\Lambda} \rho_k$, then, since $,_k$ is assumed to be Λ-consistent, $,_k \not\vdash^{\Lambda} \neg\rho_k$, and hence, by Corollary A.8, $,_{k+1} = ,_k \cup \{\rho_k\}$ is Λ-consistent.
2. If $,_k \cup \{\neg\phi(\psi_j(i,\varphi,\alpha))\} \cup \{\neg\rho_k\}$ were to be Λ-inconsistent, we would have $,_k \cup \{\neg\phi(\psi_j(i,\varphi,\alpha))\} \vdash^{\Lambda} \rho_k$, where $\rho_k = \phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi)$. Since we have $\vdash^{\Lambda} \rho_k \to \phi(\psi_l(i,\varphi,\alpha))$ for all $l \in \mathbb{N}$, we also have $,_k \cup \{\neg\phi(\psi_j(i,\varphi,\alpha))\} \vdash^{\Lambda} \phi(\psi_j(i,\varphi,\alpha))$, which implies that $,_k \cup \{\neg\phi(\psi_j(i,\varphi,\alpha))\}$ is Λ-inconsistent. But then, by Corollary A.8, $,_k \vdash^{\Lambda} \phi(\psi_j(i,\varphi,\alpha))$ which contradicts the fact that $,_k \not\vdash^{\Lambda} \phi(\psi_j(i,\varphi,\alpha))$. Hence $,_{k+1}$ is Λ-consistent.

3. If $,_k \cup \{\neg(\phi(\neg\varphi_j(i,\alpha)))\} \cup \{\neg\rho_k\}$ were to be Λ-inconsistent, we would have $,_k \cup \{\neg\phi(\neg(\varphi_j(i,\alpha)))\} \vdash^\Lambda \rho_k$, where $\rho_k = \phi(\neg\mathbf{A}_i\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})$. Since we have $\vdash^\Lambda \rho_k \to \phi(\neg(\varphi_l(i,\alpha)))$ for all $l \in \mathbb{N}$, we also have $,_k \cup \{\neg\phi(\neg(\varphi_j(i,\alpha)))\} \vdash^\Lambda \phi(\neg(\varphi_j(i,\alpha)))$, which implies that $,_k \cup \{\neg\phi(\neg(\varphi_j(i,\alpha)))\}$ is Λ-inconsistent. Then $,_k \vdash^\Lambda \phi(\neg(\varphi_j(i,\alpha)))$ which contradicts the fact that $,_k \nvdash^\Lambda \phi(\neg(\varphi_j(i,\alpha)))$. Hence $,_{k+1}$ is Λ-consistent.
4. If $,_k \nvdash^\Lambda \rho_k$ then $,_k \cup \{\neg\rho_k\}$ is Λ-consistent by Corollary A.8.

⊠

Lemma A.16 *The set , as constructed above is maximal, i.e. for all $\varphi \in$ L, exactly one of φ and $\neg\varphi$ is an element of , .*

Proof: Let $\psi \in$ L be arbitrary, then $\psi = \rho_k$ for some $k \in \mathbb{N}$. By construction, now either $\rho_k \in ,_{k+1}$ or $\neg\rho_k \in ,_{k+1}$, hence either $\psi \in ,$ or $\neg\psi \in ,$. Suppose both ψ and $\neg\psi$ in , . Then for some $k \in \mathbb{N}$, $\{\psi, \neg\psi\} \subseteq ,_k$, which would make $,_k$ inconsistent. Since this contradicts the result of Lemma A.15 given above, it follows that ψ and $\neg\psi$ are not both in , .
⊠

Lemma A.17 *The set , as constructed above is a Λ-theory.*

Proof: We need to show that , contains Λ and is closed under MP, ΩI, and ΩIA. So let $\varphi, \psi \in$ L, $i \in$ A and $\alpha \in$ Ac be arbitrary.

1. , contains Λ: If $\vdash^\Lambda \varphi$, where $\varphi = \rho_k$ for some $k \in \mathbb{N}$, then $,_k \vdash^\Lambda \rho_k$ and hence $\varphi = \rho_k \in ,_{k+1} \subseteq ,$.
2. Closure under MP, ΩI, and ΩIA:
   - MP: Suppose that φ, $\varphi \to \psi \in ,$. If $\psi \notin ,$, then $\neg\psi \in ,$, since , is maximal by Lemma A.16. Hence $\{\varphi, \varphi \to \psi, \neg\psi\} \in ,_k$ for some $k \in \mathbb{N}$, which would make $,_k$ Λ-inconsistent. This leads to a contradiction with Lemma A.15, hence $\psi \in ,$.
   - ΩI: Suppose $\{\phi(\psi_l(i,\varphi,\alpha)) \mid l \in \mathbb{N}\} \subseteq ,$. Let $\phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi) = \rho_k$, for some $k \in \mathbb{N}$. If $\rho_k \notin ,$, then $,_k \nvdash^\Lambda \rho_k$, and, by case 2 of the construction of $,_{k+1}$, this implies that $\neg\phi(\psi_j(i,\varphi,\alpha)) \in ,_{k+1}$, where $j \in \mathbb{N}$ is the least number such that $,_k \nvdash^\Lambda \phi(\psi_j(i,\varphi,\alpha))$. Hence $\neg\phi(\psi_j(i,\varphi,\alpha)) \in ,$, and by Lemma A.16, $\phi(\psi_j(i,\varphi,\alpha)) \notin ,$, which contradicts the assumption that $\{\phi(\psi_l(i,\varphi,\alpha)) \mid l \in \mathbb{N}\} \subseteq ,$. Hence $\phi([\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi) \in ,$.
   - ΩIA: Suppose $\{\phi(\neg(\varphi_l(i,\alpha))) \mid l \in \mathbb{N}\} \subseteq ,$. Let $\phi(\neg\mathbf{A}_i\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}) = \rho_k$, for some $k \in \mathbb{N}$. If $\rho_k \notin ,$, then $,_k \nvdash^\Lambda \rho_k$, and by case 2 of the construction of $,_{k+1}$, this implies that $\neg(\phi(\neg\varphi_j(i,\alpha))) \in ,_{k+1}$, for $j \in \mathbb{N}$ the least number such that $,_k \nvdash^\Lambda \phi(\neg(\varphi_j(i,\alpha)))$. Hence $\neg\phi(\neg(\varphi_j(i,\alpha))) \in ,$, and $\phi(\neg(\varphi_j(i,\alpha))) \notin ,$, which contradicts the assumption that $\{\phi(\neg(\varphi_l(i,\alpha))) \mid l \in \mathbb{N}\} \subseteq ,$. Hence $\phi(\neg\mathbf{A}_i\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od}) \in ,$.

   We conclude that , is closed under MP, ΩI and ΩIA.

Since , contains Λ and is closed under MP, ΩI and ΩIA, we conclude that , is a Λ-theory.

⊠

Now if , is Λ-inconsistent, then , $\vdash^{\Lambda} \bot$. Since, by Lemma A.17, , is a Λ-theory, it follows by Proposition A.3(2) that $\bot \in$, . Then $\bot \in$,$_k$ for some $k \in \mathbb{N}$, which contradicts the Λ-consistency of ,$_k$ which was shown in Lemma A.15. Hence , is a Λ-theory (Lemma A.17) which is maximal (Lemma A.16) and Λ-consistent, thus , is a maximal Λ-theory. Note that by construction of , , $\Phi \subseteq$, and $\neg\varphi \in$, , which suffices to conclude the right-to-left implication.

⊠

Definition A.18 *Let Λ be some logic. The canonical model* M_Λ *for Λ is defined by* $\mathrm{M}_\Lambda =^{\mathrm{def}} \langle \mathrm{S}_\Lambda, \pi_\Lambda, \mathrm{R}_\Lambda, \mathtt{r}_\Lambda, \mathtt{c}_\Lambda \rangle$ *where*

1. S_Λ *is the set of maximal Λ-theories*
2. $\pi_\Lambda(p, s) = \mathbf{1}$ *iff* $p \in s$, *for* $p \in \Pi$ *and* $s \in \mathrm{S}_\Lambda$
3. $(s, t) \in \mathrm{R}_\Lambda(i)$ *iff* $s/\mathbf{K}_i \subseteq t$, *for* $s, t \in \mathrm{S}_\Lambda$ *and* $i \in \mathrm{A}$
4. $t = \mathtt{r}_\Lambda(i, a)(s)$ *iff* $s/\mathrm{do}_i(a) \subseteq t$, *for* $i \in \mathrm{A}$, $a \in \mathrm{At}$ *and* $s, t \in \mathrm{S}_\Lambda$
5. $\mathtt{c}_\Lambda(i, a)(s) = \mathbf{1}$ *if* $\mathbf{A}_i a \in s$ *and* $\mathtt{c}_\Lambda(i, a)(s) = \mathbf{0}$ *if* $\mathbf{A}_i a \notin s$ *for* $i \in \mathrm{A}$, $a \in \mathrm{At}$ *and* $s \in \mathrm{S}_\Lambda$

Proposition A.19 *Let Λ be some logic. The canonical model* M_Λ *for Λ as defined above is a well-defined model from* $\mathbf{M}$.

Proof: Let Λ be some logic. In order to show that M_Λ is a well-defined model from $\mathbf{M}$ we have to show that the demands determining well-definedness of models are met by M_Λ. It is easily seen that S_Λ, π_Λ, and $\mathtt{c}_\Lambda$ are well-defined, which leaves to show that R_Λ and $\mathtt{r}_\Lambda$ are. To prove that $\mathrm{R}(i)$ is an equivalence relation, assume that $i \in \mathrm{A}$ and that $\{s, t, u\} \subseteq \mathrm{S}_\Lambda$. We show:

1. $(s, s) \in \mathrm{R}(i)$, i.e. $\mathrm{R}(i)$ is reflexive.
2. if $(s, t) \in \mathrm{R}(i)$ and $(s, u) \in \mathrm{R}(i)$ then $(t, u) \in \mathrm{R}(i)$, i.e. $\mathrm{R}(i)$ is Euclidean.

To show the reflexivity of $\mathrm{R}(i)$, note that $(s, t) \in \mathrm{R}(i)$ iff $s/\mathbf{K}_i \subseteq t$. Now since s contains axiom A3: $\mathbf{K}_i\varphi \to \varphi$, we have for $\varphi \in s/\mathbf{K}_i$ that $\mathbf{K}_i\varphi \in s$ and hence $\varphi \in s$ by MP. Thus $s/\mathbf{K}_i \subseteq s$, hence $(s, s) \in \mathrm{R}(i)$. To show that $\mathrm{R}(i)$ is Euclidean assume that $\varphi \in t/\mathbf{K}_i$, i.e., $\mathbf{K}_i\varphi \in t$. To prove: $\varphi \in u$. Suppose $\varphi \notin u$. Then since $(s, u) \in \mathrm{R}(i)$, $\varphi \notin s/\mathbf{K}_i$, i.e., $\mathbf{K}_i\varphi \notin s$. Since s is a maximal Λ-theory this implies that $\neg\mathbf{K}_i\varphi \in s$, and since s contains axiom A5: $\neg\mathbf{K}_i\varphi \to \mathbf{K}_i\neg\mathbf{K}_i\varphi$, also $\mathbf{K}_i\neg\mathbf{K}_i\varphi \in s$. Since $(s, t) \in \mathrm{R}(i)$, $s/\mathbf{K}_i \subseteq t$ and thus $\neg\mathbf{K}_i\varphi \in t$. But then $\mathbf{K}_i\varphi \in t$ and $\neg\mathbf{K}_i\varphi \in t$ which contradicts the consistency of t. Thus $\mathrm{R}(i)$ is Euclidean, and, combined with the reflexivity, this ensures that $\mathrm{R}(i)$ is an equivalence relation.

To show that $\mathtt{r}_\Lambda$ is well-defined, it needs to be shown that for all $i \in \mathrm{A}$, $a \in \mathrm{At}$ and $s \in \mathrm{S}_\Lambda$ it holds that $\mathtt{r}_\Lambda(i, a)(s) \in \mathrm{S}_\Lambda$ or $\mathtt{r}_\Lambda(i, a)(s) = \emptyset$. To this end it suffices to show for arbitrary $i \in \mathrm{A}$, $a \in \mathrm{At}$ and $s, t, u \in \mathrm{S}_\Lambda$ that if $t = \mathtt{r}_\Lambda(i, a)(s)$ and $u = \mathtt{r}_\Lambda(i, a)(s)$ then $t = u$. By definition it follows that $s/[\mathrm{do}_i(a)] \subseteq t$ and $s/[\mathrm{do}_i(a)] \subseteq u$ if both $t = \mathtt{r}_\Lambda(i, a)(s)$ and

$u = \mathtt{r}_\Lambda(i,a)(s)$. Since both t and u are maximal Λ-theories, both t and u are Λ-consistent, and hence $s/[\mathrm{do}_i(a)]$ is Λ-consistent. But then, by Proposition A.12, $s/[\mathrm{do}_i(a)]$ is a maximal Λ-theory, which is properly contained only in L. Hence $s/[\mathrm{do}_i(a)] = t$ and $s/[\mathrm{do}_i(a)] = u$, which suffices to conclude that $\mathtt{r}_\Lambda$ is well-defined.
⊠

Up till now, the two proof systems $\Sigma_\mathbf{0}$ and $\Sigma_\mathbf{1}$ were dealt with identically, i.e. in none of the definitions or propositions given above one needs to distinguish the proof systems or the logics based on these proof systems. From this point on, however, we need to treat the two systems, and thereby the logics, differently. We start with finishing the proof of soundness and completeness for **1**-logics, and indicate thereafter how this proof needs to be modified to end up with one for **0**-logics.

The presence of the confirmation action, which tightly links actions and formulas, prevents the subformula- or subaction-relation from being an adequate parameter for induction in the proof of the truth-theorem, the theorem which links satisfiability in a state of the canonical model to being an element of the maximal theory which constitutes the state. Instead we need a more elaborate relation, which is defined below.

Definition A.20 *The relation* $\prec$ *is the smallest relation on* $\{0,1\} \times \mathrm{L}$ *that satisfies for all* $\varphi, \psi \in \mathrm{L}$, $i \in \mathrm{A}$, *and* $\alpha, \alpha_1, \alpha_2 \in \mathrm{Ac}$ *the following constraints:*

1. $(0,\varphi) \prec (0, \varphi \vee \psi)$

2. $(0,\psi) \prec (0, \varphi \vee \psi)$

3. $(0,\varphi) \prec (0, \neg\varphi)$

4. $(0,\varphi) \prec (0, \mathbf{K}_i\varphi)$

5. $(0,\varphi) \prec (0, [\mathrm{do}_i(\alpha)]\varphi)$

6. $(1,[\mathrm{do}_i(\alpha)]\varphi) \prec (0, [\mathrm{do}_i(\alpha)]\varphi)$

7. $(1,[\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\varphi) \prec (1, [\mathrm{do}_i(\alpha_1;\alpha_2)]\varphi)$

8. $(1,[\mathrm{do}_i(\alpha_2)]\varphi) \prec (1, [\mathrm{do}_i(\alpha_1;\alpha_2)]\varphi)$

9. $(1,[\mathrm{do}_i(\texttt{confirm}\,\varphi;\alpha_1)]\psi) \prec (1, [\mathrm{do}_i(\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi})]\psi)$

10. $(1,[\mathrm{do}_i(\texttt{confirm}\,\neg\varphi;\alpha_2)]\psi) \prec (1, [\mathrm{do}_i(\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi})]\psi)$

11. $(1,\psi_l(i,\varphi,\alpha)) \prec (1, [\mathrm{do}_i(\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})]\psi)$ *for all* $l \in \mathbb{N}$

12. $(0,\neg\varphi) \prec (1, [\mathrm{do}_i(\texttt{confirm}\,\varphi)]\psi)$

13. $(1,\mathbf{A}_i\alpha) \prec (0, \mathbf{A}_i\alpha)$

14. $(1,\mathbf{A}_i\alpha_1) \prec (1, \mathbf{A}_i\alpha_1;\alpha_2)$

15. $(1,\mathbf{A}_i\alpha_2) \prec (1, \mathbf{A}_i\alpha_1;\alpha_2)$

16. $(1,[\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2) \prec (1, \mathbf{A}_i\alpha_1;\alpha_2)$

17. $(1,\mathbf{A}_i\texttt{confirm}\,\varphi;\alpha_1) \prec (1, \mathbf{A}_i\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi})$

18. $(1, \mathbf{A}_i\texttt{confirm}\,\neg\varphi; \alpha_2) \prec (1, \mathbf{A}_i\texttt{if}\,\varphi\,\texttt{then}\,\alpha_1\,\texttt{else}\,\alpha_2\,\texttt{fi})$

19. $(1, \varphi_l(i, \alpha)) \prec (1, \mathbf{A}_i\texttt{while}\,\varphi\,\texttt{do}\,\alpha\,\texttt{od})$ *for all* $l \in \mathbb{N}$

20. $(0, \varphi) \prec (1, \mathbf{A}_i\texttt{confirm}\,\varphi)$

Definition A.21 *The ordering* $<$ *is defined as the transitive closure of* $\prec$*, and* $\leq$ *is defined as the reflexive closure of* $<$*.*

Proposition A.22 *The ordering* $<$ *is well-founded.*

PROOF: The proof of this proposition is quite elaborate; it can be found in [19] where it takes over three pages. Basically, the idea is to use a powerful technique well-known from the theory of Term Rewriting Systems, viz. the lexicographic path ordering. Using this technique it suffices to select an appropriate well-founded precedence on the function symbols of the language in order to conclude that the ordering $\prec$ is well-founded. Since the actual proof is not only rather elaborate but also contains many details that are completely outside the scope of this paper, it is omitted here; those who are interested can find all details in [19]. ⊠

Having proved that the ordering $<$ is well-founded, we can use it in the proof of the truth-theorem.

Theorem A.23 (The truth-theorem) *Let* Λ *be some* **1**-*logic. For any* $\varphi \in \mathrm{L}$*, and any* $s \in \mathrm{S}_\Lambda$ *we have:* $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \varphi$ *iff* $\varphi \in s$*.*

PROOF: We prove the theorem by proving the following (stronger) properties for all $\varphi, \psi \in \mathrm{L}$, $i \in \mathrm{A}$, $\alpha \in \mathrm{Ac}$, and $s \in \mathrm{S}_\Lambda$:

1. For all $(0, \psi) \leq (0, \varphi)$ we have: $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi$ iff $\psi \in s$
2. For all $(1, [\mathrm{do}_i(\alpha)]\psi) < (0, \varphi)$ we have:
   (a) $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i, \alpha)(s) \Rightarrow [\mathrm{do}_i(\alpha)]\psi \in s$
   (b) if $t = \mathtt{r}^{\mathbf{1}}(i, \alpha)(s)$ and $[\mathrm{do}_i(\alpha)]\psi \in s$ then $\psi \in t$
3. For all $(1, \mathbf{A}_i\alpha) < (0, \varphi)$ we have: $\mathtt{c}^{\mathbf{1}}(i, \alpha)(s) = \mathbf{1}$ iff $\mathbf{A}_i\alpha \in s$

where $\mathtt{r}^{\mathbf{1}}$ and $\mathtt{c}^{\mathbf{1}}$ are the functions induced by $\mathtt{r}_\Lambda$ and $\mathtt{c}_\Lambda$ in the way described in Definition 3.4. The theorem then follows from the first item, since $(0, \varphi) \leq (0, \varphi)$. So let $\varphi \in \mathrm{L}$ be some fixed formula. We start by proving the first property. Let $\psi \in \mathrm{L}$ be such that $(0, \psi) \leq (0, \varphi)$. Consider the various cases for ψ:

- $\psi = p$, for $p \in \Pi$. By definition of π_Λ we have that $\pi_\Lambda(p, s) = \mathbf{1}$ iff $p \in s$.
- $\psi = \psi_1 \wedge \psi_2$. Since $(0, \psi_1) < (0, \psi_1 \wedge \psi_2)$ and $(0, \psi_2) < (0, \psi_1 \wedge \psi_2)$, we have that $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1 \wedge \psi_2$ iff $(\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1$ and $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_2)$ iff $\psi_1 \in s$ and $\psi_2 \in s$ (by induction on (1)) iff $\psi_1 \wedge \psi_2 \in s$ (since s is a (maximal) theory).
- $\psi = \neg\psi_1$. Since $(0, \psi_1) < (0, \neg\psi_1)$ we have that $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \neg\psi_1$ iff not$(\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1)$ iff not$(\psi_1 \in s)$ (by induction on (1)) iff $\neg\psi_1 \in s$ since s is (a) maximal (theory).

- $\psi = \mathbf{K}_i\psi_1$. We will prove two implications:

  '⇐' Suppose $\mathbf{K}_i\psi_1 \in s$. Then by definition of $\mathrm{R}_\Lambda(i)$, $\psi_1 \in t$ for all t such that $(s,t) \in \mathrm{R}_\Lambda(i)$. Since $(0,\psi_1) < (0,\mathbf{K}_i\psi_1)$, this implies that $\mathrm{M}_\Lambda, t \models^{\mathbf{1}} \psi_1$ for all $t \in \mathrm{S}_\Lambda$ with $(s,t) \in \mathrm{R}_\Lambda(i)$, hence $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \mathbf{K}_i\psi_1$.

  '⇒' Suppose $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \mathbf{K}_i\psi_1$. Now if for $t \in \mathrm{S}_\Lambda$, $(s,t) \in \mathrm{R}_\Lambda(i)$, then $\mathrm{M}_\Lambda, t \models^{\mathbf{1}} \psi_1$. Since $(0,\psi_1) < (0,\mathbf{K}_i\psi_1)$, we have by induction on (1) that $\psi_1 \in t$, for all $t \in \mathrm{S}_\Lambda$ with $(s,t) \in \mathrm{R}_\Lambda(i)$. This implies that ψ_1 belongs to every maximal theory containing $s/\mathbf{K}_i$, and by Proposition A.14 we conclude that $s/\mathbf{K}_i \vdash^\Lambda \psi_1$. By Corollary A.11 we conclude that $\mathbf{K}_i\psi_1 \in s$.

- $\psi = [\mathrm{do}_i(\alpha)]\psi_1$. We will prove two implications:

  '⇐' Let $[\mathrm{do}_i(\alpha)]\psi_1 \in s$. Let $t = \mathtt{r}^{\mathbf{1}}(i,\alpha)(s)$. Since $(1,[\mathrm{do}_i(\alpha)]\psi_1) < (0,[\mathrm{do}_i(\alpha)]\psi_1)$, we find by induction on (2b) that $\psi_1 \in t$. Since $(0,\psi_1) < (0,[\mathrm{do}_i(\alpha)]\psi_1)$, we find by induction on (1) that $\mathrm{M}_\Lambda, t \models^{\mathbf{1}} \psi_1$, if $t = \mathtt{r}^{\mathbf{1}}(i,\alpha)(s)$. But this implies that $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} [\mathrm{do}_i(\alpha)]\psi_1$.

  '⇒' Suppose $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} [\mathrm{do}_i(\alpha)]\psi_1$. This implies that $\mathrm{M}_\Lambda, t \models^{\mathbf{1}} \psi_1$ if $t = \mathtt{r}^{\mathbf{1}}(i,\alpha)(s)$. Since $(0,\psi_1) < (0,[\mathrm{do}_i(\alpha)]\psi_1)$ we have by induction on (1) that $\psi_1 \in t$ if $t = \mathtt{r}^{\mathbf{1}}(i,\alpha)(s)$. Now since $(1,[\mathrm{do}_i(\alpha)]\psi_1) < (0,[\mathrm{do}_i(\alpha)]\psi_1)$, we conclude by induction on (2a) that $[\mathrm{do}_i(\alpha)]\psi_1 \in s$.

- $\psi = \mathbf{A}_i\alpha$. Since $(1,\mathbf{A}_i\alpha) < (0,\mathbf{A}_i\alpha)$ we find by induction on (3) that $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \mathbf{A}_i\alpha$ iff $\mathtt{c}^{\mathbf{1}}(i,\alpha)(s) = \mathbf{1}$ iff $\mathbf{A}_i\alpha \in s$.

Next we prove (2a). Let $(1,[\mathrm{do}_i(\alpha)]\psi) < (0,\varphi)$. Consider the various possibilities for α.

- $\alpha = a$, for $a \in \mathrm{At}$. Assume that $\psi \in t$ if $t = \mathtt{r}_\Lambda(i,a)(s)$. By definition of $\mathtt{r}_\Lambda$ this implies that ψ is in every maximal theory containing $s/[\mathrm{do}_i(a)]$, i.e. $s/[\mathrm{do}_i(a)] \vdash^\Lambda \psi$. By Corollary A.11 we conclude that $[\mathrm{do}_i(a)]\psi \in s$.

- $\alpha = \mathtt{confirm}\,\psi_1$. Assume that $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i,\mathtt{confirm}\,\psi_1)(s)$. If $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1$ we have that $s = t$, by definition of $\mathtt{r}^{\mathbf{1}}$. Then $\psi \in s$, and, since s is a theory, this implies $\neg\psi_1 \vee \psi \in s$, which in turn implies $[\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi \in s$. If $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \neg\psi_1$ then, since $(0,\neg\psi_1) < (1,[\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi)$, we have by induction on (1) that $\neg\psi_1 \in s$, hence $\neg\psi_1 \vee \psi \in s$, and thus $[\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi \in s$.

- $\alpha = \alpha_1;\alpha_2$. Assume that $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i,\alpha_1;\alpha_2)(s)$. By definition of $\mathtt{r}^{\mathbf{1}}$ this implies that $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i,\alpha_2)(u)$ for $u = \mathtt{r}^{\mathbf{1}}(i,\alpha_1)(s)$. Since $(1,[\mathrm{do}_i(\alpha_2)]\psi) < (1,[\mathrm{do}_i(\alpha_1;\alpha_2)]\psi)$ we have that $[\mathrm{do}_i(\alpha_2)]\psi \in u$ for $u = \mathtt{r}^{\mathbf{1}}(i,\alpha_1)(s)$. Since furthermore $(1,[\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\psi) < (1,[\mathrm{do}_i(\alpha_1;\alpha_2)]\psi)$ we have that $[\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\psi \in s$. Since s is closed under the axioms of $\Sigma_{\mathbf{1}}$ and MP, this implies that $[\mathrm{do}_i(\alpha_1;\alpha_2)]\psi \in s$.

- $\alpha = \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi}$. Let $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i,\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})(s)$. Then $\psi \in t$ for all $t = \mathtt{r}^{\mathbf{1}}(i,\mathtt{confirm}\,\varphi;\alpha_1)(s)$ and $\psi \in t$ for all $t = \mathtt{r}^{\mathbf{1}}(i,\mathtt{confirm}\,\neg\varphi;\alpha_2)(s)$. Since we have both $(1,[\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha_1)]\psi) < (1,[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi)$ and $(1,[\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi;\alpha_2)]\psi) < (1,[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi)$ we have by induction that $[\mathrm{do}_i(\mathtt{confirm}\,\varphi;\alpha_1)]\psi \in s$ and $[\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi;\alpha_2)]\psi \in s$, and, since s is a theory, $[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi \in s$.

- $\alpha = \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$. Assume that $\psi \in t$ for $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s)$. Since $\mathtt{r}^{\mathbf{1}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) = \cup_{k \in \mathbb{N}} \mathtt{r}^{\mathbf{1}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s)$, we have that $\psi \in t$ for all $t = \mathtt{r}^{\mathbf{1}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s)$, for all $k \in \mathbb{N}$. Now since we have that $(1, [\mathrm{do}_i((\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)]\psi) < (1, [\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$ for all $k \in \mathbb{N}$ we have by induction on (2a) that $\psi_k(i, \varphi, \alpha) \in s$ for all $k \in \mathbb{N}$, and since s is closed under ΩA this implies that $[\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \in s$.

We continue with proving (2b). So let again $(1, [\mathrm{do}_i(\alpha)]\psi) < (0, \varphi)$, and consider the various possibilities for α.

- $\alpha = a$, for $a \in \mathrm{At}$. If $t = \mathtt{r}_\Lambda(i, a)(s)$ and $[\mathrm{do}_i(a)]\psi \in s$, then by definition of $\mathtt{r}_\Lambda$, $\psi \in t$.

- $\alpha = \mathtt{confirm}\,\psi_1$. Let $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{confirm}\,\psi_1)(s)$ and $[\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi \in s$. By definition of $\mathtt{r}^{\mathbf{1}}$, $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1$ and $s = t$. Since $(0, \psi_1) < (0, \neg\psi_1) < (1, [\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi)$, we find by induction on (1) that $\psi_1 \in s$. Since s is a theory, $[\mathrm{do}_i(\mathtt{confirm}\,\psi_1)]\psi \in s$ implies that $\neg\psi_1 \vee \psi \in s$, and, since s is maximal, we conclude that $\psi \in s$.

- $\alpha = \alpha_1; \alpha_2$. Let $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_1; \alpha_2)(s)$ and $[\mathrm{do}_i(\alpha_1; \alpha_2)]\psi \in s$. Then, by definition of $\mathtt{r}^{\mathbf{1}}$, we have that $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_2)(u)$ for some $u \in \mathrm{S}_\Lambda$ such that $u = \mathtt{r}^{\mathbf{1}}(i, \alpha_1)(s)$. Since s is closed under the axioms and proof rules of $\Sigma_{\mathbf{1}}$ we have that $[\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\psi \in s$, and hence, since $(1, [\mathrm{do}_i(\alpha_1)][\mathrm{do}_i(\alpha_2)]\psi) < (1, [\mathrm{do}_i(\alpha_1; \alpha_2)]\psi)$, we have by induction on (2b) that $[\mathrm{do}_i(\alpha_2)]\psi \in u$. But this implies, since $(1, [\mathrm{do}_i(\alpha_2)]\psi) < (1, [\mathrm{do}_i(\alpha_1; \alpha_2)]\psi)$, that $\psi \in t$.

- $\alpha = \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi}$. Let $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})(s)$ and let furthermore $[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi \in s$. Then either $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{confirm}\,\varphi; \alpha_1)(s)$ or $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{confirm}\,\neg\varphi; \alpha_2)(s)$. If $[\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi \in s$,then, since s is a theory, both $[\mathrm{do}_i(\mathtt{confirm}\,\varphi; \alpha_1)]\psi \in s$ and $[\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi; \alpha_2)]\psi \in s$. Since it holds that both $(1, [\mathrm{do}_i(\mathtt{confirm}\,\varphi; \alpha_1)]\psi) < (1, [\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi)$ and $(1, [\mathrm{do}_i(\mathtt{confirm}\,\neg\varphi; \alpha_2)]\psi) < (1, [\mathrm{do}_i(\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})]\psi)$ we have by induction on (2b) that $\psi \in t$.

- $\alpha = \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$. Let $t = \mathtt{r}^{\mathbf{1}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s)$ and $[\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi \in s$. Since s is a theory, we have that $\psi_l(i, \varphi, \alpha) \in s$, for all $l \in \mathbb{N}$. By definition of $\mathtt{r}^{\mathbf{1}}$, it holds that $t = \mathtt{r}^{\mathbf{1}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s)$ for some $k \in \mathbb{N}$. Now since $(1, \psi_l(i, \varphi, \alpha)) < (1, [\mathrm{do}_i(\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})]\psi)$ for all $l \in \mathbb{N}$, we conclude by induction on (2b) that $\psi \in t$.

Finally we come to the proof of item (3). Let $(1, \mathbf{A}_i\alpha) < (0, \varphi)$. Consider the various cases for α.

- $\alpha = a$, where $a \in \mathrm{At}$. Now $\mathtt{c}_\Lambda(i, a)(s) = \mathbf{1}$ iff $\mathbf{A}_i a \in s$, by definition of $\mathtt{c}_\Lambda$.

- $\alpha = \mathtt{confirm}\,\psi_1$. By definition, $\mathtt{c}^{\mathbf{1}}(i, \mathtt{confirm}\,\psi_1)(s) = \mathbf{1}$ iff $\mathrm{M}_\Lambda, s \models^{\mathbf{1}} \psi_1$ iff, since $(0, \psi_1) < (0, \mathbf{A}_i\mathtt{confirm}\,\psi_1)$, $\psi_1 \in s$ iff $\mathbf{A}_i\mathtt{confirm}\,\psi_1$ in s, since s is a theory.

- $\alpha = \alpha_1; \alpha_2$. We prove two implications:

  '$\Leftarrow$' Since s is a theory, $\mathbf{A}_i\alpha_1; \alpha_2 \in s$ iff $\mathbf{A}_i\alpha_1 \in s$ and $[\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2 \in s$. Since $(1, \mathbf{A}_i\alpha_1) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, we find by induction on (3) that $\mathtt{c}^{\mathbf{1}}(i, \alpha_1)(s) = \mathbf{1}$. Now suppose $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_1)(s)$. Since $(1, [\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, we find by induction on (2b) that $\mathbf{A}_i\alpha_2 \in t$. Furthermore, since $(1, \mathbf{A}_i\alpha_2) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, the latter implies that $\mathtt{c}^{\mathbf{1}}(i, \alpha_2)(t) = \mathbf{1}$, for all $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_1)(s)$, which, together with $\mathtt{c}^{\mathbf{1}}(i, \alpha_1)(s) = \mathbf{1}$, suffices to conclude that $\mathtt{c}^{\mathbf{1}}(i, \alpha_1; \alpha_2)(s) = \mathbf{1}$.

'$\Rightarrow$' By definition, $\mathtt{c}^{\mathbf{1}}(i, \alpha_1; \alpha_2)(s) = \mathbf{1}$ iff $\mathtt{c}^{\mathbf{1}}(i, \alpha_1)(s) = \mathbf{1}$ and $\mathtt{c}^{\mathbf{1}}(i, \alpha_2)(t) = \mathbf{1}$ for all $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_1)(s)$. Now since $(1, \mathbf{A}_i\alpha_1) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, we conclude by induction on (3) that $\mathbf{A}_i\alpha_1 \in s$. Furthermore, since $(1, \mathbf{A}_i\alpha_2) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, we have that $\mathbf{A}_i\alpha_2 \in t$, for all $t = \mathtt{r}^{\mathbf{1}}(i, \alpha_1)(s)$. Now since $(1, [\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2) < (1, \mathbf{A}_i\alpha_1; \alpha_2)$, we find by induction on (2a) that $[\mathrm{do}_i(\alpha_1)]\mathbf{A}_i\alpha_2 \in s$. But then, since s is a theory, we conclude that $\mathbf{A}_i\alpha_1; \alpha_2 \in s$.

- $\alpha = \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi}$. By definition of $<$ we have that $(1, \mathbf{A}_i\mathtt{confirm}\,\varphi; \alpha_1) < (1, \mathbf{A}_i\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})$ and furthermore that $(1, \mathbf{A}_i\mathtt{confirm}\,\neg\varphi; \alpha_2) < (1, \mathbf{A}_i\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})$. This implies that $\mathtt{c}^{\mathbf{1}}(i, \mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi})(s) = \mathbf{1}$ iff $\mathtt{c}^{\mathbf{1}}(i, \mathtt{confirm}\,\varphi; \alpha_1)(s) = \mathbf{1}$ or $\mathtt{c}^{\mathbf{1}}(i, \mathtt{confirm}\,\neg\varphi; \alpha_2)(s) = \mathbf{1}$ iff — by induction on (3) — $\mathbf{A}_i\mathtt{confirm}\,\varphi; \alpha_1 \in s$ or $\mathbf{A}_i\mathtt{confirm}\,\neg\varphi; \alpha_2 \in s$ iff $\mathbf{A}_i\mathtt{if}\,\varphi\,\mathtt{then}\,\alpha_1\,\mathtt{else}\,\alpha_2\,\mathtt{fi} \in s$, since s is a theory.

- $\alpha = \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od}$. We prove two implications:

'$\Leftarrow$' Let $\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \in s$. Then, since s is maximal, $\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \notin s$, and, since s is closed under ΩIA, this implies that $\neg(\varphi_k(i, \alpha)) \notin s$, for some $k \in \mathbb{N}$, and, again since s is maximal, $\varphi_k(i, \alpha) \in s$. Since $(1, \varphi_l(i, \alpha)) < (1, \mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})$ for all $l \in \mathbb{N}$, we have by induction on (3) that $\mathtt{c}^{\mathbf{1}}(i, (\mathtt{confirm}\,\varphi; \alpha)^k; \mathtt{confirm}\,\neg\varphi)(s) = \mathbf{1}$, and, by definition of $\mathtt{c}^{\mathbf{1}}$, this implies $\mathtt{c}^{\mathbf{1}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) = \mathbf{1}$.

'$\Rightarrow$' If $\mathtt{c}^{\mathbf{1}}(i, \mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})(s) = \mathbf{1}$, then $\mathtt{c}^{\mathbf{1}}(i, (\mathtt{confirm}\,\varphi; \alpha_1)^k; \mathtt{confirm}\,\neg\varphi)(s) = \mathbf{1}$ for some $k \in \mathbb{N}$. Since $(1, \varphi_l(i, \alpha)) < (1, \mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od})$ for all $l \in \mathbb{N}$, this implies by induction on (3) that $\varphi_k(i, \alpha) \in s$. Then, since s is a theory, $\neg(\varphi_k(i, \alpha)) \notin s$, and, by item 7 of Proposition A.3, it follows that $\neg\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \notin s$. Now since s is maximal it follows that $\mathbf{A}_i\mathtt{while}\,\varphi\,\mathtt{do}\,\alpha\,\mathtt{od} \in s$.

Having proved the items (1), (2) and (3) suffices to prove that the truth-theorem holds.
⊠

The proof of the truth-theorem for $\mathbf{0}$-logics is almost identical to the one given for Theorem A.23. One just needs to change one clause in the definition of the $\prec$-relation, used to apply induction upon, and modify the proof of the truth-theorem accordingly.

Definition A.24 *The ordering $<'$ is defined as the transitive closure of the smallest relation on $\{0, 1\} \times \mathrm{L}$ satisfying the constraints 1 through 15 and 17 through 20 as given in Definition A.20 and the constraint*

16'. $(1, [\mathrm{do}_i(\alpha_1)]\neg\mathbf{A}_i\alpha_2) \prec (1, \mathbf{A}_i\alpha_1; \alpha_2)$

The ordering $\leq'$ is defined to be the reflexive closure of $<'$.

The only modification to the proof of the truth-theorem for $\mathbf{1}$-logics that is required to end up with a proof of a truth-theorem for $\mathbf{0}$-logics concerns the proof of property (3) for sequentially composed actions, i.e. the proof that $\mathtt{c}^{\mathbf{0}}(i, \alpha_1; \alpha_2)(s) = \mathbf{1}$ iff $\mathbf{A}_i\alpha_1; \alpha_2 \in s$, whenever $(1, \mathbf{A}_i\alpha_1; \alpha_2) <' (0, \varphi)$. We will show this by proving two implications:

'⇐' Since s is a Λ-theory, $\mathbf{A}_i\alpha_1;\alpha_2$ in s iff $\mathbf{A}_i\alpha_1 \in s$ and $\neg[\mathrm{do}_i(\alpha_1)]\neg\mathbf{A}_i\alpha_2 \in s$. Since $(1, \mathbf{A}_i\alpha_1) <' (1, \mathbf{A}_i\alpha_1;\alpha_2)$ we find by induction on (3) that $\mathtt{c}^{\mathbf{0}}(i,\alpha_1)(s) = \mathbf{1}$. Since $(1, [\mathrm{do}_i(\alpha_1)]\neg\mathbf{A}_i\alpha_2) <' (1, \mathbf{A}_i\alpha_1;\alpha_2)$, we find by induction on (2b), read in its contrapositive form, that for some $t \in \mathrm{S}_\Lambda$, $t = \mathtt{r}^{\mathbf{0}}(i,\alpha_1)(s)$ with $\neg\mathbf{A}_i\alpha_2 \notin t$. Now since t is maximal, this implies that $\mathbf{A}_i\alpha_2 \in t$. Since $(1, \mathbf{A}_i\alpha_2) <' (1, \mathbf{A}_i\alpha_1;\alpha_2)$ the latter implies that $\mathtt{c}^{\mathbf{0}}(i,\alpha_2)(t) = \mathbf{1}$. Together with $\mathtt{c}^{\mathbf{0}}(i,\alpha_1)(s) = \mathbf{1}$ this suffices to conclude that $\mathtt{c}^{\mathbf{0}}(i,\alpha_1;\alpha_2)(s) = \mathbf{1}$.

'⇒' By definition, $\mathtt{c}^{\mathbf{0}}(i,\alpha_1;\alpha_2)(s) = \mathbf{1}$ iff $\mathtt{c}^{\mathbf{0}}(i,\alpha_1)(s) = \mathbf{1}$ and for some $t \in \mathrm{S}_\Lambda$, $t = \mathtt{r}^{\mathbf{0}}(i,\alpha_1)(s)$ and $\mathtt{c}^{\mathbf{0}}(i,\alpha_2)(t) = \mathbf{1}$. Now since $(1, \mathbf{A}_i\alpha_1) <' (1, \mathbf{A}_i\alpha_1;\alpha_2)$ we have by induction on (3) that $\mathbf{A}_i\alpha_1 \in s$. Furthermore, since $(1, \mathbf{A}_i\alpha_2) <' (1, \mathbf{A}_i\alpha_1;\alpha_2)$, we have for the aforementioned t that $\mathbf{A}_i\alpha_2 \in t$. Hence we have some $t \in \mathrm{S}_\Lambda$ such that $t = \mathtt{r}^{\mathbf{0}}(i,\alpha_1)(s)$ and $\mathbf{A}_i\alpha_2 \in t$ while also $\mathbf{A}_i\alpha_1 \in s$. Since t is maximal, $\neg\mathbf{A}_i\alpha_2 \notin t$. And, rephrasing (2b) to 'if $t = \mathtt{r}^{\mathbf{0}}(i,\alpha)(s)$ and $\psi \notin t$ then $[\mathrm{do}_i(\alpha)]\psi \notin s$', we conclude by induction on (2b) that $[\mathrm{do}_i(\alpha_1)]\neg\mathbf{A}_i\alpha_2 \notin s$. Since s is maximal it follows that $\neg[\mathrm{do}_i(\alpha_1)]\neg\mathbf{A}_i\alpha_2 \in s$. Hence $\mathbf{A}_i\alpha_1 \in s$ and $\langle\mathrm{do}_i(\alpha_1)\rangle\mathbf{A}_i\alpha_2 \in s$, which, since s is a Λ-theory, implies that $\mathbf{A}_i\alpha_1;\alpha_2 \in s$, which was to be shown.

Having proved the truth-theorem both for **1**-logics and for **0**-logics, we can prove that deducibility for a logic Λ corresponds with validity in the canonical model M_Λ.

Proposition A.25 *For all* **b**-*logics* Λ *and all* $\varphi \in \mathrm{L}$ *we have:* $\vdash^\Lambda \varphi$ *iff* $\mathrm{M}_\Lambda \models^{\mathbf{b}} \varphi$.

Proof: Let $\varphi \in \mathrm{L}$ be arbitrary. Then we have:

$\vdash^\Lambda \varphi$ iff $\varphi \in s$, for all $s \in \mathrm{S}_\Lambda$
    iff $\mathrm{M}_\Lambda, s \models^{\mathbf{b}} \varphi$ for all $s \in \mathrm{S}_\Lambda$
    iff $\mathrm{M}_\Lambda \models^{\mathbf{b}} \varphi$

⊠

Using the propositions and theorems shown above, we can now prove those given in Section 6. Note that Proposition 6.10 is already shown as the third item of Proposition A.6.

6.7. Theorem. *For* **b** $\in$ bool *and all* $\varphi \in \mathrm{L}$ *we have:*

- $\vdash^{\mathbf{b}} \varphi \Leftrightarrow \models^{\mathbf{b}} \varphi$

Proof: We prove the theorem by proving two implications.

'⇐' If $\models^{\mathbf{b}} \varphi$ then $\mathrm{M} \models^{\mathbf{b}} \varphi$ for all $\mathrm{M} \in \mathbf{M}$. Since $\mathrm{M}_{\mathrm{LCap}_\mathbf{b}} \in \mathbf{M}$ it follows that $\mathrm{M}_{\mathrm{LCap}_\mathbf{b}} \models^{\mathbf{b}} \varphi$. By Proposition A.25 it then follows that $\vdash^{\mathbf{b}} \varphi$.

'⇒' Suppose $\vdash^{\mathbf{b}} \varphi$ and let $\mathrm{M} \in \mathbf{M}$. By Proposition A.2 we have that $\{\psi \in \mathrm{L} \mid \mathrm{M} \models^{\mathbf{b}} \psi\}$ is a **b**-logic. Since $\mathrm{LCap}_\mathbf{b}$ is the smallest **b**-logic, it follows that whenever $\varphi \in \mathrm{LCap}_\mathbf{b}$ also $\varphi \in \{\psi \in \mathrm{L} \mid \mathrm{M} \models^{\mathbf{b}} \psi\}$, and hence $\mathrm{M} \models^{\mathbf{b}} \varphi$. Since M is arbitrary, it follows that $\mathrm{M} \models^{\mathbf{b}} \varphi$ for all $\mathrm{M} \in \mathbf{M}$ and thus $\models^{\mathbf{b}} \varphi$, which was to be shown.

⊠

6.11. Proposition. *The proof systems* $\Sigma_\mathbf{1}$ *and* $\Sigma_\mathbf{0}$ *are strongly complete, i.e. every set* $\Phi \subseteq \mathrm{L}$ *that is* $\mathrm{LCap}_\mathbf{b}$*-consistent is* $\models^{\mathbf{b}}$*-satisfiable.*

Proof: The proposition follows, for arbitrary logics, directly from the proof of Proposition A.14. For if Φ is Λ-consistent, then by the procedure given in the proof of Proposition A.14 one constructs a maximal Λ-theory , that contains Φ. This , appears as a state in the canonical model for Λ, and by the truth-theorems, all formulas from , — and hence from Φ — are satisfied at this state. Hence every Λ-consistent set Φ is satisfied at some state of the canonical model for Λ, and since this canonical model is a well-defined one, the proposition follows.

⊠

References

[1] J. van Benthem. Correspondence theory. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 167–247. Reidel, Dordrecht, 1984.

[2] M.A. Brown. On the logic of ability. *Journal of Philosophical Logic*, 17:1–26, 1988.

[3] C. Castelfranchi. Guarantees for autonomy in cognitive agent architecture. In M. Wooldridge and N.R. Jennings, editors, *Intelligent Agents – Agent Theories, Architectures, and Languages*, volume 890 of *Lecture Notes in Computer Science (subseries LNAI)*, pages 56–70. Springer-Verlag, 1995.

[4] B.F. Chellas. *Modal Logic. An Introduction.* Cambridge University Press, Cambridge, 1980.

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

[6] *Communications of the ACM*, vol. 37, nr. 7. Special Issue on Intelligent Agents.

[7] N. Dershowitz and J.-P. Jouannaud. Rewrite systems. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 243–320. Elsevier, 1990.

[8] J.L. Fiadeiro and P.-Y. Schobbens, editors. *Proceedings of the 2nd Workshop of the ModelAge Project*, 1996.

[9] L.N. Foner. What's an agent, anyway? A sociological case study. Technical report, MIT Media Laboratory, 1993.

[10] R. Goldblatt. *Axiomatising the Logic of Computer Programming*, volume 130 of *LNCS*. Springer-Verlag, 1982.

[11] R. Goldblatt. The semantics of Hoare's iteration rule. *Studia Logica*, 41:141–158, 1982.

[12] R. Goldblatt. *Logics of Time and Computation*, volume 7 of *CSLI Lecture Notes*. CSLI, Stanford, 1992. Second edition.

[13] J. Halpern and J. Reif. The propositional dynamic logic of deterministic, well-structured programs. *Theoretical Computer Science*, 27:127–165, 1983.

[14] J.Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[15] D. Harel. Dynamic logic. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, chapter 10, pages 497–604. D. Reidel, Dordrecht, 1984.

[16] D. Hilbert. Die Grundlegung der elementaren Zahlenlehre. *Mathematische Annalen*, 104:485–494, 1931.

[17] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, NY, 1962.

[18] C.A.R. Hoare. An axiomatic basis for computer programming. *Communications of the ACM*, 12:576–580, 1969.

[19] W. van der Hoek, B. van Linder, and J.-J. Ch. Meyer. A logic of capabilities. Technical Report IR-330, Vrije Universiteit Amsterdam, July 1993.

[20] G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic.* Routledge, London, 1968.

[21] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic.* Methuen & Co. Ltd., London, 1984.

[22] A. Kenny. *Will, Freedom and Power.* Basil Blackwell, Oxford, 1975.

[23] J.W. Klop. Term rewriting systems. In S. Abramsky, D.M. Gabbay, and T.S.E. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 2, pages 1–116. Oxford University Press, New York, 1992.

[24] D. Kozen and J. Tiuryn. Logics of programs. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 789–840. Elsevier, 1990.

[25] S. Kripke. Semantic analysis of modal logic. *Zeitschrift für Mathematische Logik und Grundslagen der Mathematik*, 9:67–96, 1963.

[26] F. Kröger. Infinite proof rules for loops. *Acta Informatica*, 14:371–389, 1980.

[27] Y. Lespérance, H. Levesque, F. Lin, D. Marcu, R. Reiter, and R. Scherl. Foundations of a logical approach to agent programming. In M. Wooldridge, J.P. Müller, and M. Tambe, editors, *Intelligent Agents Volume II – Agent Theories, Architectures, and Languages*, volume 1037 of *Lecture Notes in Computer Science (subseries LNAI)*, pages 331–347. Springer-Verlag, 1996.

[28] V. Lesser, editor. *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95).* MIT Press, 1995.

[29] B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. Communicating rational agents. In B. Nebel and L. Dreschler-Fischer, editors, *KI-94: Advances in Artificial Intelligence*, volume 861 of *Lecture Notes in Computer Science (subseries LNAI)*, pages 202–213. Springer-Verlag, 1994.

[30] B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. Tests as epistemic updates. In A.G. Cohn, editor, *Proceedings of the 11th European Conference on Artificial Intelligence (ECAI'94)*, pages 331–335. John Wiley & Sons, 1994.

[31] B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. Actions that make you change your mind. In A. Laux and H. Wansing, editors, *Knowledge and Belief in Philosophy and Artificial Intelligence*, pages 103–146. Akademie Verlag, 1995.

[32] B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. Formalising motivational attitudes of agents: On preferences, goals and commitments. In M. Wooldridge, J.P. Müller, and M. Tambe, editors, *Intelligent Agents Volume II – Agent Theories, Architectures, and Languages*, volume 1037 of *Lecture Notes in Computer Science (subseries LNAI)*, pages 17–32. Springer-Verlag, 1996.

[33] B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer. The dynamics of default reasoning. *Data and Knowledge Engineering*, 21(3):317–346, 1997.

[34] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40, July 1994.

[35] P. Maes. Intelligent software. *Scientific American*, 273(3):66–68, September 1995. Special Issue on Key Technologies for the 21st Century.

[36] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science.* Cambridge University Press, 1995.

[37] R.C. Moore. A formal theory of knowledge and action. In J.R. Hobbs and R.C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex, Norwood, NJ, 1985.

[38] I. Pörn. *Action Theory and Social Science.* Reidel, Dordrecht, 1977.

[39] A.S. Rao and M.P. Georgeff. Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In J. Mylopoulos and R. Reiter, editors, *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI'91)*, pages 498–504. Morgan Kaufmann, 1991.

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

[41] A.S. Rao and M.P. Georgeff. A model-theoretic approach to the verification of situated reasoning systems. In R. Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 318–324. Morgan Kaufmann, 1993.

[42] D. Riecken. Intelligent agents. *Communications of the ACM*, 37(7):18–21, July 1994.

[43] K. Schütte. *Beweistheorie.* Springer-Verlag, Berlin-Göttingen-Heidelberg, 1960.

[44] K. Segerberg. Bringing it about. *Journal of philosophical logic*, 18:327–347, 1989.

[45] T. Selker. Coach: A teaching agent that learns. *Communications of the ACM*, 37(7):92–99, July 1994.

[46] Y. Shoham. Agent-oriented programming. *Artificial Intelligence*, 60:51–92, 1993.

[47] P.A. Spruit. Henkin-style completeness proofs for Propositional Dynamic Logic. Manuscript.

[48] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.

[49] M. Wooldridge and N.R. Jennings, editors. *Intelligent Agents – Agent Theories, Architectures, and Languages*, volume 890 of *Lecture Notes in Computer Science (subseries LNAI)*. Springer-Verlag, 1995.

[50] M. Wooldridge, J.P. Müller, and M. Tambe, editors. *Intelligent Agents Volume II – Agent Theories, Architectures, and Languages*, volume 1037 of *Lecture Notes in Computer Science (subseries LNAI)*. Springer-Verlag, 1996.

[51] G.H. von Wright. *Norm and Action.* Routledge & Kegan Paul, London, 1963.

[52] G.H. von Wright. The logic of action: A sketch. In N. Rescher, editor, *The Logic of Decision and Action.* University of Pittsburgh Press, 1967.