Small steps in dynamics of information

Fernando Raymundo Velázquez Quesada

Small steps
in dynamics of information

ILLC Dissertation Series DS-2011-02

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 904
1098 XH Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

Small steps in dynamics of information

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.dr. D.C. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op dinsdag 22 februari 2011, te 12.00 uur

door

Fernando Raymundo Velázquez Quesada

geboren te Mexico-Stad, Verenigde Mexicaanse Staten.

Promotiecommissie

Promotor: Prof. dr. J. F. A. K. van Benthem

Overige leden:
Dr. N. Alechina
Prof. dr. A. Aliseda Llera
Dr. H. P. van Ditmarsch
Prof. dr. P. van Emde Boas
Dr. D. Grossi
Prof. dr. A. Nepomuceno Fernández
Prof. dr. F. J. M. M. Veltman

Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam
Science Park 904
1098 XH Amsterdam

The investigations were partially supported by the **Consejo Nacional de Ciencia y Tecnología** (**CONACyT**), México (scholarship # 167693).

Cover design by Inés and Mercedes Paulina Velázquez Quesada.
Printed and bound by Ponsen & Looijen B.V., Wageningen.

ISBN: 978-90-5776-222-2
ILLC Dissertation Series: DS-2011-02

Contents

Chapter 1

Introduction

Look at the following situations of a relatively common day. You wake up in the morning and observe that the day is slightly windy and cold. Being summer, you assume that the weather will get better, and decide to go to work without a jacket. But after five minutes biking, the sky gets filled with dark clouds; then you change your mind and, expecting rain, go back to pick a raincoat.

Once you arrive at your workplace, you look at the notes on the blackboard (whiteboard nowadays) to remember the activities that your team should finish. From the five activities for the week, two have been completed and your colleagues are working on two of the others; then you realize you should take care of the remaining one, and start working on it.

At 17hrs you are about to go home, and you have planned to visit your bank to make a payment. On your way out, a colleague tells you that it might be possible to do a transfer via internet. Now that you consider this possibility, you make a call to your bank and happily find out that indeed an internet transfer is possible, taking care of it immediately.

When arriving at home, you see the bedroom window open. You assume that you left it open in the morning, and then you realize that the bookshelf below it should be wet after today's rain. Later that night, after having dinner, you remember to set up the alarm. Then you go to sleep, hoping to have enough rest and be ready for the next day.

The previous example shows how every single day of our life is filled with small actions that change our information. We observe new facts, draw inferences from them, make assumptions, become aware of new possibilities, acknowledge what we do and do not know, forget some things and remember others. All these actions change our knowledge, beliefs, opinions, desires, intentions and other attitudes in a small but decisive way, and they are precisely the main interest of the present dissertation. Our main goal is to provide a formal logical framework in which we can not only represent, but also reason about small steps in dynamics of information.

In order to achieve this goal, we should start from the beginning. If we want to represent and reason about *small steps* in dynamics of information, we need a setting that allows us to represent and reason about *dynamics* of information. And in order to represent and reason about *dynamics* of information, we should first find an adequate framework in which we can represent *information*, and reason about it.

One of the most well-known systems for this, *Epistemic Logic* (Hintikka 1962; Fagin et al. 1995), provides us with a compact and powerful framework that allows us to deal not only with an agent's information about propositional facts, but also with her information about her own (and eventually other agents') information. This system has a very simple language and its usual semantic model, possible worlds, is very intuitive. On top of this, simple specific properties of the model allow us to deal with different attitudes, like knowledge, safe, conditional and plain beliefs, and several others. For these and other reasons, Epistemic Logic is widely used in many areas in which information representation is needed, like Computer Science (security and distributed systems), Philosophy (Epistemology), Economics (Game Theory) and others. All these reasons make it very appealing for our purposes.

Nevertheless, with possible worlds as semantic model, the system has an important drawback. An agent represented in this framework is *logically omniscient*: her information is closed under logical consequence. This property, useful in some areas, has been widely criticized in some others, and there is an extensive literature discussing it (e.g., Sim (1997), Moreno (1998) and Halpern and Pucella (2007)). Most people agree that omniscience is an excessive idealization for 'human' agents; after all, we have disciplines like Mathematics and Computer Science whose purpose is to fill in the logical consequences of the information we already have. But omniscience is also a strong assumption for *computational* agents who may lack the required time and/or space (Ågotnes and Alechina 2009). For our purposes of representing small steps in dynamics of information, omniscience is also undesirable since it makes irrelevant some of the actions we are interested in: for example, an act of truth-preserving inference does not give new information to an omniscient agent since she already has all logical consequences of her information.

If we want to use Epistemic Logic, we should take care of this omniscience problem. Many approaches have been presented in order deal with it, and most of them do it by weakening the properties of the agent's information. Some of them use syntactic representations of information; some use *impossible worlds*. Some others use variants of neighbourhood models and even non-standard logical approaches (see the mentioned surveys for summaries).

There is, nevertheless, an important observation to take into account when discussing omniscience. As several authors have mentioned (Konolige (1984); Levesque (1984); Lakemeyer (1986); Vardi (1986); Fagin and Halpern (1988);

Barwise (1988); van Benthem (2006) among many others), Epistemic Logic really describes the agent's implicit semantic information, a notion that is definitely closed under logical consequence. But this closure does not need to hold for weaker attitudes, like 'actual' or *explicit* information. And it is precisely for these finer notions for which the finer actions that we want to represent are actually meaningful.

What we need first is, then, to define a model in which we can represent finer notions of information, and we will begin by reviewing the existing literature.

1.1 Structure of information

Information is a widely used term, and therefore there are several definitions and theories about it in many different fields, ranging from natural to social sciences and humanities. Even when restricting ourselves to logical frameworks, we can find several accounts of this notion (van Benthem and Martínez 2008). Fortunately, we can divide them in two main groups, according to the way information is understood and, therefore, represented.

1.1.1 Semantic representations

Semantic approaches associate an agent's information with the collection of situations she considers possible. In other words, semantic approaches encode information by means of a range possibilities: the different ways the real situation might be from the agent's point of view.

In fact, semantic approaches do not focus on the agent's information, but on her *uncertainty*. Instead of representing directly the information the agent has, these approaches encode it by representing the situations the agent cannot rule out. A great range indicates a big uncertainty, and therefore less information about the real situation. On the other hand, a small range indicates less uncertainty, that is, more information. For example, an agent is informed that a given p is the case when her range contains only possibilities where p holds, and she is not informed about whether a given q is the case when she considers as possible at least one situation where q fails (so she cannot affirm that q is true) and another in which q holds (so she cannot affirm that q is false).

Semantical approaches provide us with a very compact representation of information. An agent that considers only two possibilities, one in which p and q hold, and another in which p holds but q does not, is indeed informed about p. At this point the range (two possibilities) may seem larger than just writing down p, but it encodes much more. An agent with such range is also informed about *"p and p"*, *"q or not q"*, *"p or q"*, *"q or p"*, *"p or not q"* and many more simply because her range does not contain possibilities in which these statements fail. All this is encoded in a two-possibilities range; instead of listing exhaustively

all the agent knows about the real situation, we just need to list the situations she considers possible, given the information she currently has .

There are several approaches for representing information as a collection of possible situations. Among them, the best-known is the already mentioned Epistemic Logic with possible worlds as semantic model. Since this framework will play a prominent role in the rest of our work, we will devote some time here to present it properly.

Epistemic Logic

Epistemic Logic (*EL*) was first introduced in Hintikka (1962), and it has been further developed by many authors from different disciplines. Its most common semantic model, the *possible worlds model*, is formally defined as follows.

Definition 1.1 (Possible worlds model) Let $\mathtt{P}$ be a set of atomic propositions. A *possible worlds model* is a tuple $M = \langle W, R, V \rangle$ where

- W is a non-empty set whose elements are called *possible worlds (situations, states, possibilities, points)*;
- $V : W \to \wp(\mathtt{P})$ is an *atomic valuation function*, indicating the atomic propositions in $\mathtt{P}$ that are true at each possible world;
- $R \subseteq (W \times W)$ is an accessibility relation, indicating which worlds the agent considers possible from each one of them.

Among the possible worlds, we usually distinguish one called the *evaluation point*. The pair (M, w), consisting of a possible worlds model M and this distinguished world w, is called a *pointed possible worlds model*. ◀

A possible worlds model $M = \langle W, R, V \rangle$ is a collection of situations (W), each one of them associated to an atomic valuation that indicates which atomic propositions are true in it (V). The model represents an agent's information by indicating which situations the agent considers possible from each world (R). More precisely, from each $w \in W$, the agent considers as possible all the situations u that she can R-access from w, that is, she considers possible those situations in $R[w] := \{u \in W \mid Rwu\}$. Note how the information range of the agent is not defined globally but rather locally, since the worlds the agent considers possible may vary from world to world.

Epistemic Logic is more than just a semantic model for representing information. It has an associated language that allows us to talk about the real situation and the information an agent has about it, therefore allowing us *to reason* about the agent's information. This language extends the propositional one with a modal operator $\Box$. With it we can build formulas of the form $\Box\varphi$, read as "*the agent is informed about* φ". The formal definition is the following.

Definition 1.2 (Epistemic Logic language) Let $\mathtt{P}$ be a set of atomic propositions. The language of *Epistemic Logic* contains exactly those formulas built according to the following rules.

1. An atomic proposition $p \in \mathtt{P}$ is a formula in the language.
2. If φ and ψ are formulas in the language, so are $\neg\varphi$, $\varphi \vee \psi$ and $\Box\varphi$.
3. Nothing else is a formula in the language.

The definition can be abbreviated with the following statement. Formulas φ, ψ of the *EL* language are built according to the following rule, where p is in $\mathtt{P}$:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \Box\varphi$$

Other connectives, like conjunction ($\wedge$), implication ($\rightarrow$) and biconditional ($\leftrightarrow$), can be defined from negation ($\neg$) and disjunction ($\vee$) in the standard way:

$$\varphi \wedge \psi := \neg(\neg\varphi \vee \neg\psi), \quad \varphi \rightarrow \psi := \neg\varphi \vee \psi, \quad \varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi).$$

Similarly, the constants $\top$ and $\bot$ can be defined as $p \vee \neg p$ and $p \wedge \neg p$, respectively. The 'diamond' modal operator $\Diamond$ is defined as the dual of $\Box$:

$$\Diamond\varphi := \neg\Box\neg\varphi$$

We can get the reading of 'diamond' formulas by unfolding its definition: $\Diamond\varphi$ corresponds to $\neg\Box\neg\varphi$, that is, *"it is not the case that the agent is informed about* $\neg\varphi$*"* or, in other words, *"the agent considers* φ *possible"*. ◂

As we mentioned before, the range of an agent is defined locally, and therefore formulas of the *EL* language are evaluated in pointed possible worlds models. The formal definition of this 'truth' relation between pointed models and formulas is as follows.

Definition 1.3 (Semantic interpretation) Let the pair (M, w) be a pointed possible worlds model, with $M = \langle W, R, V\rangle$. Then,

$(M, w) \Vdash p$	iff	$p \in V(w)$
$(M, w) \Vdash \neg\varphi$	iff	it is not the case that $(M, w) \Vdash \varphi$
$(M, w) \Vdash \varphi \vee \psi$	iff	$(M, w) \Vdash \varphi$ or $(M, w) \Vdash \psi$
$(M, w) \Vdash \Box\varphi$	iff	for all $u \in W$, Rwu implies $(M, u) \Vdash \varphi$

When $(M, w) \Vdash \varphi$, we say that φ is true (holds) at w in M. ◂

The semantic interpretation of atomic propositions is given directly by the atomic valuation, and that of negation and disjunction is the classical one. The interesting one here is the semantic interpretation of $\Box\varphi$: the agent is informed about φ at w in M, $(M, w) \Vdash \Box\varphi$, if and only if φ is true in all the worlds that are

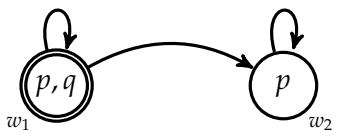
R-reachable from w. In other words, the agent is informed about φ if and only if φ is true in all the worlds she considers possible.

We can use the semantic interpretation of our primitive connectives and modalities $\neg$, $\vee$ and $\Box$ in order to get that of the defined ones $\wedge$, $\rightarrow$, $\leftrightarrow$ and $\Diamond$. In particular, $\Box$ works as a universal quantifier restricted to the worlds the agent considers possible from the evaluation point, so the semantic interpretation of $\Diamond$ corresponds to a restricted *existential* quantifier:

$$(M, w) \Vdash \Diamond \varphi \quad \text{iff} \quad \text{there is a } u \in W \text{ such that } Rwu \text{ and } (M, u) \Vdash \varphi$$

The following example shows how a small possible worlds model allows us to represent a huge amount of information.

Example 1.1 Consider the following possible worlds model M. It has two possible worlds, w_1 and w_2, with their respective valuation indicated: both p and q are true at w_1, p is true and q is false at w_2. When considering w_1 as the evaluation (double circled) point, the model describes a situation in which w_1 is the real world, but the agent considers possible both w_1 and w_2. Then, (1) the agent is informed about p, but (2) she is informed about neither q nor $\neg q$.



(1) $(M, w_1) \Vdash \Box p$

(2) $(M, w_1) \Vdash \neg\Box q \wedge \neg\Box \neg q$

But the model represents much more than just the agent's information about p and lack of information about whether q. It also indicates that the agent is informed about $p \vee q$ ($\Box(p \vee q)$ is true at w_1), $q \rightarrow p$ ($\Box(q \rightarrow p)$ is true w_1) and many other propositional formulas. More importantly, the model also represents *high-order* information, that is, information the agent has about her own information. For example, while the agent is informed that she is informed about p ($\Box\Box p$) and she is informed about her lack of information about q ($\Box\neg\Box q$), she is not informed that she is informed about q ($\neg\Box\Box q$). ◀

Identifying a piece of information with the situations in which it is true allows us to encode a huge amount of information within a small model. But there is a price to pay.

Consequences of semantic representations

Semantic approaches identify a piece of information with the situations in which it holds. Then, the agent cannot make a difference between formulas that are true exactly in the same situations: she is informed about one of them if and only if she is informed about the other. In the pointed model (M, w_1) of Example 1.1, the agent is informed about $p \vee q$, but also about the logically equivalent $\neg(\neg p \wedge \neg q)$, $\neg p \rightarrow q$, $\neg q \rightarrow p$ and so on.

In particular, associating formulas with the situations in which they are true implies that all tautologies are informationally equivalent simply because all of them are true in every possible situation. This implies that 'obvious' tautologies like $p \to p$ are, from the agent's perspective, identical to more 'illuminating' ones like $(p \wedge (p \to q)) \to q$.

This informational equivalence of logically equivalent formulas is inherent to any semantic approach. But the combination of the *EL* language and possible worlds models has a stronger effect: the described omniscience property that makes the agent's information closed under logical consequence. The key reason for this property is that each possible world encodes an infinite set of *EL*-formulas: exactly the ones that are true at it. This fact becomes evident when we look at the general Henkin model (Henkin 1950) of this case: the *canonical* possible worlds model.

The canonical possible worlds model has as domain the collection of all maximal consistent sets of *EL*-formulas, that is, each possible world is defined as a set of *EL*-formulas with two properties: consistency (the contradiction $\bot$ cannot be finitely derived) and maximality (no further formula can be added without making the set inconsistent). The model is called *canonical* because it is a 'universal' model: any satisfiable set of *EL*-formulas can be satisfied in it.

The important observation for us is what is called the *Truth Lemma*: the *EL*-formulas that are true at each world of this canonical model (i.e., the *EL*-formulas that are true at each maximal consistent set) are exactly those that belong to it. In other words, for the *EL* language, the canonical possible worlds model is a syntactic construction that puts explicitly in each possible world exactly all the *EL*-information the world itself provides. But this information turns out to be a maximal consistent (and therefore infinite) set of *EL*-formulas!

Now it is easier to see where the omniscience property comes from. Recall that an agent is informed about φ at w if and only if φ is true in all the worlds she considers possible. As the canonical possible worlds model shows, each possible world stands for a maximal consistent set of *EL*-formulas, and maximal consistent sets are closed under logical consequence: if $\varphi \to \psi$ and φ are in the set, then consistency gives us the right to add ψ and maximality actually puts it in. But then the information each possible world provides is closed under logical consequence, and hence so is the agent's information: if the agent is informed about both $\varphi \to \psi$ and φ at w (if $\Box(\varphi \to \psi)$ and $\Box\varphi$ are true at w), then both $\varphi \to \psi$ and φ are true in all worlds R-reachable from w. But then ψ also holds in each one of such worlds, and therefore the agent is also informed about ψ at w ($\Box\psi$ is true at w).

How omniscience affects inference As mentioned before, the omniscience property has a consequence that is important for us: truth-preserving inference, which guarantees the truth of the conclusion from the truth of the premises, becomes irrelevant. It does not provide any new information, since anything

that follows logically from the agent's information is already part of it. In other words, the possible worlds approach does not account for the informative nature of truth-preserving inference. Information is represented as the agent's range of possibilities, but there is the tacit assumption that the agent has available every single piece of information each one of these possibilities encode. Therefore truth-preserving inference, which extends the agent's information about each possibility, does not provide anything new.

This is a serious drawback because, for human beings, truth-preserving inference is clearly informative in many cases. It may not *create* new information in the sense that what we infer was already present in some implicit form, but it definitely *gives us* new information that we did not have before the inference. A simple and clear example is the proof of a theorem. When we state the assumptions (*"Let x be y and suppose z holds"*), we are merely reducing the possible situations that should be considered. Then, after that, there is the proof of the theorem, which is nothing but a sequence of truth-preserving reasoning steps showing that the conclusion indeed holds. We need the proof because, even after discarding the irrelevant possibilities, we may not have the needed information about the remaining ones to see the truth of the conclusion; we need these steps to bring the conclusion into the light.

So pure semantic approaches are not suitable for our purposes. Which other alternatives do we have for representing information?

1.1.2 Syntactic representations

On the other extreme of the coarse semantic representation, we have syntactic approaches. They follow the most natural way of representing information: by means of symbols of a given formal language, encoding information in formulas at some abstract level. After all, human information is most obviously expressed in written or spoken language, and the use of a formal one has the advantage of avoiding most ambiguities and other obscurities. These approaches have the advantages of clarity (the encoding is merely a translation from a natural to a formal language) and being fine-grained enough to allow us to even represent possible differences in idiosyncracies and formulation. In this view, syntactic approaches can be seen as *"... little more than a streamlined and regimented version of an ordinary language"* (Hintikka and Sandu 2007).

The simplest variant of this approach is the one in which information is represented by a plain set of formulas of a given formal language, an idea originated when looking for representations of knowledge and beliefs that are adequate for more 'real' beings like humans (Eberle 1974) or computers (Moore and Hendrix 1979). For example, if some agent is informed that some fact p is the case, then we simply add p to her correspond set of formulas. If, on the other hand, she does not have information about whether p is the case or not, we do not add anything to her set. The idea is to represent an agent's information

simply by an explicit, plain and exhaustive listing of what the agent knows about the real situation. The more the information, the bigger the set.

Other variants assume a set of formulas with certain properties. One of the most representative examples are the belief sets of classical *Belief Revision* (see, e.g., Gärdenfors (1992); Williams and Rott (2001)): consistent sets of formulas closed under logical consequence. Note how such assumptions produce omniscient agents, and it is generally accepted that these properties are not realistic for describing the actual beliefs of individuals.

Some other syntactic approaches consider sets of formulas without particular properties, but with further internal structure. Ryan (1992) considers *ordered theory representations*: multi-sets of formulas with a partial order among them. Then we have the *labelled deductive systems* of Gabbay (1996) that enrich formulas with labels (terms of an algebra, formulas of another logic, resources or databases) providing further information. Again in Belief Revision, there is also the distinction between belief sets, the mentioned consistent set of formulas closed under logical consequence, and the *belief base*, a simple set of formulas which serves as a basis for generating the belief set (Makinson 1985).

In a syntactic representation, the meaning of each piece of information is completely determined by the formula representing it. Then, two pieces of information have the same meaning if and only if they are represented by the same sequence of symbols. This could be excessive in some cases: for example, though the linguistic conjunction 'and' is usually understood as the affirmation of both conjuncts, the formulas $p \wedge q$ and $q \wedge p$ are syntactically different and therefore express different information. One could argue that in some situations the order of the conjuncts does make a difference (think about the different emphasis in the phrases *"she is smart and beautiful"* and *"she is beautiful and smart"*), but there are more extreme cases. It is difficult to find a situation in which p and $p \wedge p$ have different meaning and yet the two formulas, being syntactically different, are understood as different pieces of information. Syntactic approaches have been criticized as being too fine-grained, making differences in meaning where there seems to be none.

Another criticism to syntactic approaches is that the typically used formal languages can express the information the agent can get but not what information the agent has. To indicate that the agent is informed about p we add the formula to the corresponding set, but usually there is no formula that expresses the fact that the agent is informed about p. In such cases, the reasoning about the properties of the agent's information should take place in a metalanguage, usually a non-logical one. This also prevents the agent from having high-order information, that is, information about her own information.[1]

[1]Note that we do not mean that high-order information cannot be represented syntactically (the *EL* language shows it can be done), but rather that few purely syntactic approaches have used this possibility.

1.1.3 Intermediate points

The problem of choosing an adequate representation of information is that our intuition pulls in opposite directions. Paraphrasing a sentence of Moore (1989), it seems that in order to be the object of attitudes like knowledge, beliefs and so on, information should be individuated almost as narrowing as sentences of natural language, and yet, it seems that it should not be represented specifically with linguistic entities, but rather 'semantic' objects of some special kind.

Several authors have looked for intermediate representations. In fact, *"much of the discussion of 'propositions' and 'meanings' in the philosophical literature [. . .] might be seen as the search for a level of information in between mere sets of models and every last detail of syntax"* (van Benthem and Martínez 2008).

Carnap already worked with syntactic descriptions of possible situations in his inductive logic (Carnap 1952). His *state descriptions* are conjunctions describing the atomic valuation of the situation, that is, conjunctions containing, for each atomic proposition, either the atom itself or else its negation.

Lewis (1970) also looked for a balance when defining *meaning*. He argued that the *intension*, a function that returns the truth-value of a sentence based on a series of arguments like a possible world, time, place, speaker, audience and others, does not provide the sentence's meaning. The reason is that sentences with the same intension may have different meanings: for example, *"it would be absurd to say that all tautologies have the same meaning, but they have the same intension; the constant function having [for every argument] the value [*true*]"* (Lewis 1970). He proposes that, for atomic sentences, we can identify meaning with intension, but the meaning of composed sentences should be given by the intensions of the constituents. With this idea, the tautology "Snow is white or it is not" differs in meaning from the similarly structured "Grass is green or it is not" because their respective components, "Snow is white" and "Grass is green", have different intentions.

More recently, Moore (1989) suggested that the simplest approach with some hope of success is the 'Russellian' view. Different from the 'Fregean' perspective that claims that a proposition consists of a relation and *the concepts of the related objects*, Russell (1903) defines a proposition as a relation *and the related objects* themselves. Moore argues that, since they are structured objects, Russellian propositions can mirror syntax in order to distinguish propositions that are true in the same situations. Nevertheless, they are no linguistic entities since they are defined by means of objects and relations.

The mentioned approaches, based on philosophical discussions about what a proposition means and what kind of information it conveys, attack the problem at the very foundations of the theory of meaning. There are also approaches that aim at intermediate points by looking at existing proposals and then abstracting existing differences (if the original proposal is syntactic) or imposing

further ones (if the original proposal is semantic). Among the latter we find *neighbourhood models*, generalizations of possible worlds models developed independently in Scott (1970) and Montague (1970). Similar to syntactic approaches, a neighbourhood model represents an agent's information by listing all the formulas the agent is informed about; similar to semantic approaches, each one of these formulas is not presented as a string of symbols, but as the set of situations in which it is true.

More precisely, in a neighbourhood model $M = \langle W, N, V \rangle$, the accessibility relation is replaced by a neighbourhood function $N : W \to \wp(\wp(W))$ that assigns, to each possible world, a set of *sets of worlds*. Then it is said that the agent is informed about some formula φ at a world w if and only if the set of worlds in which φ is true in M (denoted by $[\![\varphi]\!]^M$) is in $N(w)$.[2] This allows us to represent agents whose information is not closed under logical consequence, since $N(w)$ does not need to have any particular closure property, and therefore having $[\![\varphi]\!]^M$ and $[\![\varphi \to \psi]\!]^M$ does not imply to have $[\![\psi]\!]^M$. The textbook Chellas (1980) and the lecture notes Pacuit (2007) provide extensive information and important results about neighbourhood models.

As appealing as it may be, a neighbourhood model is still not fine enough to provide important differences in meaning. It still associates pieces of information with the set of situations in which they are true, and therefore the agent cannot make a difference between formulas that are true in exactly the same situations, that is, she cannot make a difference between logically equivalent formulas. This has again the unpleasant consequence of making the tautologies $p \to p$ and $(p \wedge (p \to q)) \to q$ the same in the eyes of the agent.

Despite all the efforts, there is no clear consensus about a proper intermediate representation. Lewis himself mentions that, though some approaches can be found, he *"doubt[s] that there is any unique natural way to do so"* (Lewis 1970).

1.1.4 Combining the two extremes

There is, however, an important observation about the way semantic and syntactic approaches understand information; an observation that is useful for deciding where to look for an appropriate information representation.

Semantic approaches are based on a *universal* quantification: the agent has a piece of information if and only if this information is true *in all* the situations she considers possible. Syntactic approaches, on the other hand, are based on *existential* quantification: the agent has a piece of information if and only if in her information set *there is* a formula standing for it. These two views are not opposite; for example, completeness results establish correspondence

[2] There is also an alternative approach in which the neighbourhood of w should contain not the set of worlds in which φ is true, but simply a subset of it. The two approaches are compared in Areces and Figueira (2009).

between semantic validity as truth in every model and syntactic validity as the existence of a derivation/proof. As mentioned in van Benthem and Martínez (2008), semantic and syntactic approaches can be seen as the *dual* or the *complement* of each other, and therefore they can be put together, just like the use of complementary colors is an important aspect in art and graphic design or like sound and silence complement each other in musical pieces. Looking for an approach standing in between semantics and syntax is not the only possibility; from the perspective of this alternative methodology, we can also look at the different ways these two extremes can 'complete the circle' and work together.

Some authors have looked at this duality. van Benthem (1993) suggested a merging between notions of information as range with some sort of 'calculus' of justifications, and in the literature there are already several proposals for this, like combinations of Epistemic Logic with *Logics of Proofs* (Artemov and Nogina 2005) and with modal representations of inference (van Benthem 2008d).

1.2 Dynamics

When discussing what information is and how it should be represented, there is an important fact to keep in mind: information is not static. Our knowledge, beliefs, opinions, desires, intentions and other attitudes change as the result of many different informational activities, including not only those given by the interaction with our environment (reading newspapers, conversations, asking questions) but also those that stand for our own internal reasoning (inferences, changes in awareness, acts of introspection, remembering and forgetting). Information states are just stages in a dynamic process, and the actions that produce the changes should be taken into account when looking for a proper representation of an agent's information.

The importance of studying structures together with the actions that transform them has been recognized in many fields. Actions, in fact, play the key role in Computer Science, a field frequently described as the systematic study of algorithmic processes that create, describe, and transform information. Many logicians and philosophers have also given a first-class status to the acts that transform information. Lewis (1970) mentions that *"in order to say what a meaning* is*, we may first ask what a meaning* does*"*; Gärdenfors (1988) emphasizes that *"the problem of finding an appropriate knowledge representation is a key problem for artificial intelligence. But a solution to this problem is of little help unless one also understands how to* update *the epistemic states in the light of new information"*; van Benthem and Martínez (2008) also agrees in that *"[t]here are structures representing the information, but these only make sense as vehicles for various processes of information flow"*. Emphasizing actions is the main idea behind the *dynamic* turn in Epistemic Logic: notions of information should be studied together with the informational actions that modify them.

When looking for a adequate representation of information, we should also take into account which are the actions we are interested in. Let us review what semantics and syntax offer us.

1.2.1 Semantic dynamics

With semantic approaches that represent information as a range of possibilities, the relevant informational actions are those that modify this range. The most natural of such operations is the one that reduces the range, standing for an action of observation, but there are many other options, like introducing new possibilities, or modifying the further internal structure the range may have.

All in all, ranges of information change as the agent makes observations or engages in communication. And it is not strange that the actions that make sense for semantic representations have this 'external-interaction' flavor. After all, since the agent's semantic information has strong closure properties, she usually has already all the information she can extract from each one of the possibilities she considers. No further 'internal' reflexive act will lead to new information, so the only way she can get to know more about the real situation is by means of *interaction*, either with her environment (acts of observation) or with other agents (acts of communication).

Just like Epistemic Logic is the best-known paradigm based on a semantic representation of information, its dynamic counterpart, *Dynamic Epistemic Logic* (*DEL*; van Ditmarsch et al. (2007)), has become an important paradigm for representing changes in an agent's range. Some of the most interesting *DEL*-incarnations, like *action models* for representing uncertainty about the observation (Baltag et al. 1999), or order-changing operations for representing changes in preference and/or beliefs (van Benthem 2007; van Benthem and Liu 2007; Baltag and Smets 2008) will be discussed in different chapters of this dissertation. Here we will present the proper definitions of the simplest of them: *Public Announcement Logic* (*PAL*; Plaza (1989); Gerbrandy (1999)), which allows us to represent public announcements and the effect they have in the agent's information. Since *PAL* announcements are not associated to any announcer, we will refer to them with another name: we will call them acts of *observation*.

Observation Logic

The act of observation is the simplest one of the 'external' actions an agent can perform. In order to express the effects of such an act, the *EL* language is extended with an existential modality $\langle \chi! \rangle$, the *observation* modality, where χ is a formula of the language. More precisely,

Definition 1.4 (Observation Logic language) Let $\mathtt{P}$ be a set of atomic propositions. Formulas φ, ψ, χ of the language of *Observation Logic* are given by:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \Box\varphi \mid \langle\chi!\rangle\varphi$$

with p an atomic proposition in $\mathtt{P}$. Formulas of the form $\langle\chi!\rangle\varphi$ are read as *"χ can be observed and after that φ will be the case"*. The universal counterpart of the $\langle\chi!\rangle$ modality is defined as its dual, as usual:

$$[\chi!]\varphi := \neg\langle\chi!\rangle\neg\varphi$$

This formula is read as *"after any observation of χ, φ will be the case"*. ◀

Besides being the simplest 'external' action an agent can perform, the act of observation also has a very natural interpretation as an operation that reduces the agent's range of possibilities. By observing certain χ, the agent realizes that χ is true, and therefore she can discard those possibilities in which χ does not hold, hence keeping just those in which χ is the case. The formal definition of this operation over a possible worlds model is as follow.

Definition 1.5 (Observation operation) Let $M = \langle W, R, V\rangle$ be a possible worlds model, and let χ be a formula in the Observation Logic language. The possible worlds model $M_{\chi!} = \langle W', R', V'\rangle$ is given by

- $W' := \left\{w \in W \mid (M, w) \Vdash \chi\right\}$;
- $R' := R \cap (W' \times W')$;
- for every $w \in W'$, $V'(w) := V(w)$.

In words, the *observation operation* restricts the model by keeping only the possible worlds where the observed χ holds, restricting the accessibility relation to the new domain and retaining the atomic valuation of the preserved worlds.◀

The observation formula is related to the observation operation by means of its semantic interpretation.

Definition 1.6 (Semantic interpretation) Let the pair (M, w) be a pointed possible worlds model and χ a formula of the observation language.

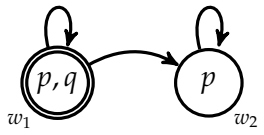
$$(M, w) \Vdash \langle\chi!\rangle\varphi \quad \text{iff} \quad (M, w) \Vdash \chi \text{ and } (M_{\chi!}, w) \Vdash \varphi$$

By unfolding its definition, the semantic interpretation of the universal observation modality becomes

$$(M, w) \Vdash [\chi!]\varphi \quad \text{iff} \quad (M, w) \Vdash \chi \text{ implies } (M_{\chi!}, w) \Vdash \varphi$$ ◀

Note how the observation modality comes with a precondition. Indeed, χ can be observed and after doing it φ will be the case, $(M, w) \Vdash \langle \chi! \rangle \varphi$, if and only if χ can be observed, $(M, w) \Vdash \chi$, and after observing it, φ is the case, $(M_{\chi!}, w) \Vdash \varphi$. This precondition has a technical reason. We need for the evaluation point to satisfy χ; otherwise, it will not survive the observation operation. But there is also an intuitive and very natural reason: in order for χ to be observed, χ *has to be true*.

Example 1.2 Recall the possible worlds model M of Example 1.1, representing a situation where the agent is informed about p, but not informed about whether q. Suppose that she observes that indeed q is the case. The resulting model $M_{q!}$ and formulas indicating the observation's effect appear below (note that the formulas are still evaluated in the original pointed model (M, w_1)).



M

$M_{q!}$

$(M, w_1) \Vdash \neg\Box q \wedge \langle q! \rangle \Box q$

$(M, w_1) \Vdash \Box p \wedge \langle q! \rangle \Box p$

$(M, w_1) \Vdash \Box \neg\Box q \wedge \langle q! \rangle (\neg\Box \neg\Box q \wedge \Box\Box q)$

The observation changes the agent's information by adding q to it, as the first formula expresses, preserving the agent's information about p, as the second formula indicates. But not all the agent's previous information is preserved. This may look counterintuitive at first sight, but this is because we usually consider information about plain propositional facts (which is not affected by an observation, since the atomic valuation in the surviving worlds is not modified), but not information *about information*. Before the observation, the agent was informed about her lack of knowledge about q, as we indicated before. But after the observation this information has gone! Even more: now she is informed about her being informed about q, as the third formula expresses. ◀

Note two important facts about the observation operation. First, the new incoming information reduces the possibilities the agent considers; as mentioned before, more information leads to a smaller range. Second, by observing q, the agent gets much more than just q. For example, after the observation, she is also informed about $p \wedge q$ and $\Box q$; the first one is reasonable since she already had p, and the second one is also reasonable from a *conscious* observation. But there is more. After the observation the agent is also informed about any propositional logical consequence of p and q together and, moreover, every epistemic logical consequence of being informed about p and being informed about q. This is a consequence of the information's closure under logical consequence in possible worlds models. Observing q discards the possibilities where q does not hold, but then the agent's information is 'recomputed', producing not only q but also all logical consequences of q and the closed-under-logical-consequence information the agent already had before.

1.2.2 Syntactic dynamics

Syntactic approaches represent information as a set of formulas, so syntactic dynamics of information are nothing but operations that modify such sets. More precisely, since syntactic approaches represent information with symbols, dynamics in these approaches can be seen as operations that manipulate an existing collection of such symbols in order to produce a new collection.

The major syntactic dynamic paradigm is *inference*, the process of drawing a conclusion from some given assumptions (premises) in a general sense. Logic itself has been traditionally identified with the study of inferences and inferential relations, with *Proof Theory* (Troelstra and Schwichtenberg 2000) and *Theory of Computation* (including lambda calculus, recursive functions and Turing machines) as the relevant sub-disciplines. Also traditionally, Logic has focused on the development of mechanisms for *truth-preserving* inference (also called *deduction*): inference in which the truth of the premises guarantees the truth of the conclusion. Being closely tied to details of syntactic representation, there is a great variety of logical systems for truth-preserving inference, with very different formats. While Proof Theory itself is based on Hilbert-style proof systems and natural deduction, there are other well-established approaches like *Logic Programming* (Kowalski 1979), *Resolution Theorem Proving* and *Unification*-based mechanisms (Doets 1994).

Truth-preserving inference is essentially cumulative or, to use a more common term, monotonic. This means that truth-preserving inference with true premises, besides allowing us to add the derived conclusion to the collection of things we know are true, it also assures us that this collection of true facts will never need to be revised, since no further information can affect their status of 'true'. This focus on monotonic inference has slowly changed, and the 1980s witnessed the development of many proposals for non-monotonic (i.e., non-truth-preserving) inferences: those in which what we have accepted as true may become false in the light of further information or, in other words, those in which the conclusion can be false even if the premises are true. These proposals emerged with the aim to formalize the more 'human' reasoning processes that we use in our every-day life. Indeed, non-monotonic inferences seem close to our common reasoning, and the classical examples are typical situations in which we 'adventurously' make extra assumptions, given our lack of complete information about the real world. If someone tells us that Chilly Willy is a bird, we will probably think it can fly. Nevertheless, further information about Chilly Willy (being hurt, being a penguin, etc.) could make us reconsider its flying abilities. Among the most important works on non-monotonic inference, we can mention *Default Logic* (Reiter 1980), *Circumscription* (McCarthy 1980), *Auto-epistemic Logic* (Moore 1985) and the extensively studied *Belief Revision* (Alchourrón et al. 1985; Gärdenfors 1992; Gärdenfors and Rott 1995; Rott 2001; Williams and Rott 2001).

Inference can be seen as an operation over a set of formulas, but it can also be seen as a transition between information states. Recent works (Duc (1995); Ågotnes (2004); Jago (2006a) among others) have represented inference in a modal style, with states standing for sets of formulas and transitions between them standing for (rule-based) inference steps. Then modal languages can be used to describe such structures; this has the advantage of allowing us to express the information the agent has at the current state and, moreover, how it will change after a given sequence of reasoning acts.

But inference is not the only dynamic paradigm of syntactic approaches. Inference indeed corresponds to operations that add formulas with certain degree of truth: certainly true in truth-preserving inferences, probably but no definitely true in the case of non-monotonic ones. But we can also look at operations that remove formulas from the agent's information set, representing in this way actions of *forgetting* or *rejecting* certain information.

Syntactic approaches are also fine enough to represent *changes in awareness*, that is, changes in the possibilities an agent considers. For example, consider an agent whose information set contains only formulas built from the atoms p and q. The action of introducing the formula $r \vee \neg r$ does not give the agent any real information, but it can be seen as introducing the topic r to the conversation, therefore making the agent aware of that possibility.

Even more. If the language from which formulas are built allows us to express whether the agent has some given piece of information or not, then syntactic dynamics can also represent acts of *introspection*: the action through which an agent realizes that she has or does not have some piece of information. If the formula A p is read as *"the agent is informed about p"*, then an operation that puts such formula in the information set represents an action through which the agent becomes informed of being informed about p.

Though the actions of change in awareness and introspection are conceptually different from the act of inference, in syntactic representations the formers could be seen as a particular case of the latter. After all, all of them can be represented by operations that add formulas/remove formulas to/from the agent's information set. In other words, in syntactic representations, a general notion of inference can be seen as the 'normal form' for the mentioned actions.

1.3 A summary and a choice

We have recalled several approaches for representing information from both semantic and syntactic perspectives. Semantic approaches, representing information as a range of possible situations, have the advantage of a clear and compact representation of factual and often high-order information, but are too coarse in the sense that they cannot make fine distinctions on pieces of

information. Syntactic approaches, representing information by an exhaustive listing of formulas of a formal language encoding information at some abstract level, have the advantage of being fine-grained enough to allow us to represent possible differences in idiosyncracies and formulation, but frequently do not provide further structure to this information and often suffer from a language that lacks enough expressivity to reason about the information the agent has. Though there are several proposals that look for intermediate points with a proper granularity, there is no clear consensus about an ideal representation.

Nevertheless, semantic and syntactic approaches are not opposite. They simply look at the notion of information from different, but nevertheless complementary perspectives: information as what is true *in all* possible situations in the semantic case, and information as the *existence* of a string of symbols representing it in the syntactic one. This observation shows that there is an alternative methodology: we can take these two extreme approaches and 'close the circle', that is, we can put them together.

Now recall our main goal: we are interested in representing and reasoning about small steps in dynamics of information. What is important for us are the informational actions that can be defined and studied in a given representation of information. As we have seen, while semantic approaches are natural frameworks for representing actions that correspond to the agent's external interaction, syntactic approaches are natural frameworks for representing actions that correspond to the agent's internal and introspective reasoning. And while our focus is mainly the internal actions, it is the combination of all of them what matters. Our opening example shows how our every-day life is full of informational activities that work together with each other, transforming our information in small but decisive ways.

Given our interest in dynamics of information, we will follow the idea of combining the two extremes. This will allow us to represent 'internal' and 'external' informational actions together, therefore giving us the possibility to express not only the isolated effect of each one of them, but also the way they intertwine with each other in real-life scenarios. Through this dissertation, we will work with possible worlds models extended with functions that indicates explicitly the information the agent has at each possible world. One natural way of understanding this combination is the following.

While semantic approaches represent the agent's information about the real situation by encoding it in terms of all the situations the agent considers possible, syntactic approaches represent the agent's information about the real situation by an explicit enumeration of formulas encoding it. Then, by putting the two approaches together, what we get is a model in which we can represent the information an agent has about the real situation *by explicitly listing the information the agent has about each one of the situations she considers possible*.

1.4 Outline of the dissertation

This dissertation is organized as follows.

We start in Chapter 2 from the already mentioned observation: the $\Box$ operator should not be understood as 'full-blooded information', representing what the agent actually has, but as a notion of *implicit* information, representing what she can eventually get. In order to define the agent's *explicit* information, we follow two systems in the *Dynamic Syntactic Epistemic Logic* style, and we associate to each possible world a set of formulas and a set of rules. While the first is interpreted as the formulas the agent has acknowledged as true in each possible world, the second is interpreted as the rules the agent can apply in each one of them. Then, by asking for extra model properties, we focus on notions of *true* information, that is, implicit and explicit *knowledge*. This setting already allows us to represent non-omniscient agents.

But our point is not only to represent agents that are not ideal, but also to represent the actions that lead such agents to change (and possibly improve) their information. Following this idea, we define model operations that represent two of the most important informational actions: (rule-based) truth-preserving inference and explicit observations. Moreover, we also provide model operations that mimic the application of *structural rules*, allowing the agent to extend the inference rules she can apply. All these operations are introduced semantically and syntactically, and in the three cases a complete axiomatization is provided following the reduction axioms technique. Once the framework has been defined and some of its properties discussed, we show how it allows us to describe real-life situations.

In Chapter 3 we explore another reason for which an agent may not be explicitly informed about her implicit information: lack of *awareness*. We recall the existing *awareness logic*: a setting that extends a possible worlds model by associating a set of formulas to each possible world. Different from the previous chapter, these sets do not indicate what the agent has acknowledged as true, but only the formulas she is aware of (what she entertains), without specifying any attitude pro or con. The framework gives us several options for defining explicit information, and we discuss some of them.

Then we explore the dynamics of the introduced notions. We present actions that produce changes in awareness and in implicit information, therefore producing changes in explicit information too. In all cases, we provide semantic and syntactic definitions as well as complete axiomatizations.

Though the actions that change awareness have an 'internal' feeling, they become public when we move to a multi-agent environment. Then we take the *action models* idea and extend it in order to deal with the syntactic component of our models; this allows us to provide private and even unconscious versions of the awareness-changing actions.

In Chapter 4 we combine the two different ingredients that in the previous chapters made explicit the agent's implicit information: awareness of the formula and acknowledgement of it as true. In particular, the notion of awareness we work with is not given by an arbitrary set of formulas anymore: it is now given by the formulas generated from the atoms the agent can use in all the worlds she considers possible. Asking for equivalence indistinguishability relations allows us to turn the notions of implicit and explicit information into implicit and explicit knowledge, and we discuss several of their properties.

On the dynamic side, we adapt the already provided actions of raising awareness, truth-preserving inference and explicit observation to the new richer setting, again stating syntactic and semantic definitions as well as complete axiom systems. For the action of explicit observation, we briefly sketch a version that fits better the non-omniscient spirit of our work. Then we show how the developed setting allows us to describe the way information flows during agents' interaction.

The previous chapters deal either with an abstract notion of information or else with the notion of knowledge. Nevertheless, most of the behaviour of 'real' agents is based not on what they know, but rather on what they *believe*. Based on *DEL* ideas for representing this notion and our previous ideas for representing non-omniscient agents, we introduce in Chapter 5 a framework for representing implicit and explicit beliefs, and we discuss some of the properties of these notions.

Then we look at the dynamics. We recall the existing notion of *upgrade*, close to the notion of revision, and we adapt it to our non-omniscient setting. But, just like a setting with implicit and explicit knowledge suggest the action of deduction, the current setting suggest different forms of inferences that involve not only knowledge but also beliefs. We argue that such inferences should allow the agent to create more possibilities, and with that aim we combine existing plausibility models with the richer action models defined in Chapter 3. This yields a framework in which we can represent several forms of inference, including not only combinations of known/believed premises/rules, but also weak and strong forms of local reasoning. We also provide a completeness result that extends to the multi-agent system of Chapter 3, and then we present an example of the situations the setting can describe.

In Chapters 6 and 7 we present links of the developed framework with different areas. The first one focuses on known forms of inference, and it shows how our framework allows us to represent some forms of deduction, default and abductive reasoning; it also discusses connections with belief bases and how our setting deals with contradictions. The second one focuses on connections with other fields, including Linguistics, Cognitive Science as well as Game Theory.

Finally, Chapter 8 concludes this dissertation, presenting a summary of the developed work, and mentioning further interesting questions that deserve additional investigation.

Sources of the chapters The material presented in this dissertation is based on the following papers.

The framework of Chapter 2 for representing inference and explicit observations has evolved from Velázquez-Quesada (2008a) and Velázquez-Quesada (2008b), and appears in its final version in Velázquez-Quesada (2009a).

The material on which the dynamization of awareness of Chapter 3 is based appears in van Benthem and Velázquez-Quesada (2010).

The analysis of awareness, implicit and explicit knowledge of Chapter 4 is the final installment of a work whose previous versions appear in Grossi and Velázquez-Quesada (2009) and Grossi and Velázquez-Quesada (2010).

Chapter 5 extends the work on implicit and explicit beliefs that appears in Velázquez-Quesada (2009c), Velázquez-Quesada (2010b) and Velázquez-Quesada (2010a).

Parts of Chapter 6 appear in Soler-Toscano and Velázquez-Quesada (2010).

Chapter 2

Truth-preserving inference and observation

As mentioned before, Logic itself has been traditionally identified with the study of inferences and inferential relations. Also traditionally, Logic has focused on *truth-preserving* inference: inference in which the truth of the premises guarantees the truth of the conclusion. Nevertheless, there are not so many approaches that study inference from an agent's point of view. Our first step towards a framework for representing small steps in dynamics of information is the development of a framework in which we can represent the action of truth-preserving inference and reason about it.

But our goal is not to represent each one of the different discussed actions in isolation; our goal is to represent them as different components of the same framework, so we can study not only the particular effect of each one of them, but also the way they interact and work together. Then, in this chapter, we will present a framework in which we can represent not only the act of *truth-preserving inference*, but also a non-omniscient version of the act of observation, that is, *explicit observation*.

2.1 The Restaurant example

Consider the following situation, from van Benthem (2008a):

> *You are in a restaurant with your parents, and you have ordered three dishes: fish, meat, and vegetarian, for you, your father and your mother, respectively. Knowing that each person gets one dish, a new waiter comes from the kitchen with the full order. What can he do to get to know which dish corresponds to which person?*

The waiter can ask "Who has the fish?"; then he can ask "Who has the meat?". Now he does not have to ask anymore: *"two questions plus one inference are all that is needed"* (van Benthem 2008a).

This example shows the two mentioned logical processes at work. On one hand we have acts of *observation*, represented by the answers the second waiter receives for his questions. Acts of this kind reflect the agent's interaction with her environment, and provide her with new arbitrary (and yet truthful) information. On the other hand, we also have an act of *inference*, more precisely, an act of *deduction*, represented by the reasoning step the waiter performs to realize that the remaining person should get the remaining dish. Acts of this kind are more 'internal', and allow the agent to derive new information based on what she already has.

Like van Benthem mentions, these two phenomena fall directly within the scope of modern Logic, since *"asking a question and giving an answer is just as 'logical' as drawing a conclusion!"* (van Benthem 2008b). Indeed, both processes are equally important in their own right, but so is their interaction. This chapter is devoted to the development of a logical framework that allows us to represent inference and observation *together*.

The approach of the present chapter results from combining ideas for representing inference in a modal framework with ideas for representing observations in the same setting. The key notions for the latter have been already introduced (the *Observation Logic* of Section 1.2.1), so now we will provide a brief summary of two modal approaches for inference: *Dynamic Syntactic Epistemic Logic* and *Logic for Rule-Based Agents*.

2.2 Modal truth-preserving inference

The two approaches discussed in this section emerged as proposals for solving the *logical omniscience problem*. Though most of the proposals for solving this problem focus on weakening the properties of the agent's information, some authors (Drapkin and Perlis (1986); Duc (1995) among others) have argued that solutions of this kind are not acceptable. Their main reasons can be summarized in the following two points.

1. The agent's information can be weakened in many ways, and there is no clear method to decide which restrictions produce reasonable agents and which ones make them too strong/weak.

2. These approaches do not look at the heart of the matter: they still describe the agent's information at a single (probably final) stage, without looking at how such state is reached.

Dynamic Syntactic Epistemic Logic and *Logic for Rule-Based Agents* are based on the idea of *dynamizing* Epistemic Logic. As it is argued, by saying that an agent knows the laws of Logic, we do not mean that she knows some facts about the world, but rather that she is able to use these laws in the proper

situations to draw conclusions from what she already knows. This idea allows *"a good trademark between logical omniscience and logical ignorance: the agent is surely not omniscient with respect to her actual or explicit knowledge, but neither is she logically ignorant"* because she can always extend her information by means of the proper reasoning steps (Duc 1997).

2.2.1 Dynamic Syntactic Epistemic Logic

In Duc (1995, 1997, 2001), Ho Ngoc Duc proposes a *Dynamic Syntactic Epistemic Logic* to represent truth-preserving inference in a modal framework. His main goal is to represent an agent that is not logically omniscient, but nevertheless has enough reasoning abilities to extend what she knows. In other words, the agent's knowledge does not appear automatically; it is the result of an action.

Duc proposes several languages, the main difference between them being the sub-language used for expressing what the agent can get to know (just propositional formulas, propositional and epistemic formulas, etc.). His approach is mainly syntactic in the sense that he mostly focuses on presenting different axiom systems and discussing how reasonable are some postulates about the agent's reasoning abilities (he discusses questions like "Should the agent have perfect recall?", "Can she reach an omniscient state?", "Is the reasoning linear or branching?"). Among his proposals, the most interesting for us is the one in which he presents not only a language but also a semantic model: the logic $\mathcal{L}_{BDE}$ (Duc 1995).

The language for this logic is built in two stages. There is an internal language, the propositional one, that allows us to express the knowledge the agent can get. Then there is another language to reason about this knowledge and how it evolves.

Definition 2.1 (Language $\mathcal{L}_{BDE}$) Let At denote the set of formulas of the form $\Box\gamma$ for γ a propositional formula. The language $\mathcal{L}_{BDE}$ is the smallest set of formulas that contains At and is closed under negation, conjunction and the modal operator $\langle F\rangle$. More precisely, the language $\mathcal{L}_{BDE}$ is given by

$$\gamma ::= p \mid \neg\gamma \mid \gamma \vee \delta$$
$$\varphi ::= \Box\gamma \mid \neg\varphi \mid \varphi \vee \psi \mid \langle F\rangle\varphi$$

Formulas of the form $\Box\gamma$ and $\langle F\rangle\varphi$ are read as *"γ is known"* and *"after some course of thought of the agent, φ is true"*, respectively. Note how the agent's knowledge is restricted to propositional formulas and how $\mathcal{L}_{BDE}$ itself does not allow us to talk about the real world.

A *BDE*-model is a small variation of a possible worlds model satisfying some special requirements.

Definition 2.2 (*BDE*-model) A general model M is a tuple (W, R, V) where $W \neq \emptyset$ is the set of *possible worlds*, $R \subseteq (W \times W)$ is a transitive binary relation and $V : W \to \wp(At)$ associates a set of formulas of At to each possible world. Then, a *BDE-model* is a general model M in which

1. for all $w \in W$, if $\Box\gamma \in V(w)$ and Rwu, then $\Box\gamma \in V(u)$;
2. for all $w \in W$, if $\Box\gamma$ and $\Box(\gamma \to \delta)$ are in $V(w)$, then there is a world u such that Rwu and $\Box\delta \in V(u)$.
3. if γ is a propositional tautology, then for all $w \in W$ there is a world u such that Rwu and $\Box\gamma \in V(u)$. ◀

Definition 2.3 (Semantic interpretation) Given a *BDE*-model, the semantic interpretation for negation and conjunctions is as usual. For formulas in At and 'course-of-thought' formulas, we have

$$\begin{array}{lll} (M,w) \Vdash \Box\gamma & \text{iff} & \Box\gamma \in V(w) \\ (M,w) \Vdash \langle F\rangle\varphi & \text{iff} & \text{there is a } u \in W \text{ such that } Rwu \text{ and } (M,u) \Vdash \varphi \end{array}$$ ◀

Duc's strategy is now clear. In order to avoid logical omniscience, he represents the agent's knowledge in a syntactic form, with the function V returning those formulas the agent knows at each world. In order to avoid logical ignorance he introduces inference, representing it as a relation that stands for the reasoning steps the agent can perform in order to change her knowledge.

Given the semantic interpretation, we can now see the meaning of the three semantic requirements. The first guarantees that the agent's knowledge will only grow as she reasons, and the second and third guarantee that this knowledge will be closed under modus ponens and will contain all tautologies at some point in the future. So though a $\mathcal{L}_{BDE}$-agent is not omniscient, she has the logical resources to eventually derive all that follows logically from what she currently knows (i.e., she can reach an omniscient state).

Though the important part of the language, the modality $\langle F\rangle$, is similar to the 'future' modality in *Tense Logic* (Prior 1957), Duc discusses a more interesting interpretation of these "course of thoughts". If the set of actions the agent can perform is explicitly given by $\{r_1, \ldots, r_n\}$, then F actually stands for $(r_i \cup \cdots \cup r_n)^+$: the transitive closure of the non-deterministic application of actions in the set. This makes the approach closer to the ideas in *Propositional Dynamic Logic* (*PDL*; Harel et al. (2000)). In fact, we could look at a more appealing *PDL*-style language that states explicitly which are the actions that the agent performs. To quote one of Duc's examples, assume that the agent knows the conjunction of p and $p \to q$, that is, $\Box(p \wedge (p \to q))$. In classical Epistemic Logic, it follows that the agent knows $p \wedge q$, that is, $\Box(p \wedge q)$. But there is no guarantee that a realistic agent will know $p \wedge q$ automatically. What we should say instead is that *if* she

knows p and $p \to q$, *and* she reasons appropriately, *then* she will get to know $p \wedge q$. In this concrete case, let *CE*, *MP* and *CI* stand for the rules of *conjunction elimination*, *modus ponens* and *conjunction introduction*, respectively. Then, by using *PDL*-style notation (';' stands for sequential composition), instead of the omniscient $\Box(p \wedge (p \to q)) \to \Box(p \wedge q)$, we get the more realistic

$$\Box\Big(p \wedge (p \to q)\Big) \to \langle CE; MP; CI\rangle \Box\Big(p \wedge q\Big)$$

But semantically, in order to formalize this example, we need more than just the abstract relation R of before. We need specific relations R_{CE}, R_{MP}, R_{CI} and so on for each one of the inference steps the agent can perform. More importantly, we need to be sure that each one of these relations follows the intuition behind it: if R_{MP} relates world w with world u, then at w the agent should know an implication and its antecedent, and at u her knowledge should be extended with the implication's consequent.

The approach that we will recall now attacks the second problem, providing formal definitions of what a relation should satisfy in order to represent properly a rule-application.

2.2.2 Logic for Rule-Based Agents

Based on the prominent case of rule-based agents of the *Artificial Intelligence* (*AI*) literature, Mark Jago proposes in Jago (2006a,b, 2009) a system for agents whose reasoning steps are given by a generalized version of modus ponens.

In his approach, the agent can have *beliefs* about two different entities: literals and rules. A *literal* λ is an atomic proposition or its negation. A rule, denoted usually as ρ, has the form $\lambda_1, \ldots, \lambda_n \Rightarrow \lambda$, with λ and all λ_is literals. In particular λ, the *rule's conclusion*, is usually denoted by $\mathsf{cn}(\rho)$.

The important concept, that of a *rule application*, has the following form:

$$\frac{\lambda_1, \ldots, \lambda_n, (\lambda_1, \ldots, \lambda_n \Rightarrow \lambda)}{\lambda}$$

In words, if the agent has the rule and all its premises, then she can apply it, obtaining the rule's conclusion.

The language used to reason about the agent's beliefs, called $\mathcal{ML}$, is based on formulas of the form $\mathsf{B}\lambda$ and $\mathsf{B}\rho$ for λ a literal and ρ a rule, and it is closed under negation, conjunction and the existential modal operator $\Diamond$.

Definition 2.4 (Language $\mathcal{ML}$) Let $\mathtt{P}$ be a set of atomic propositions. The collection of literals λ based on $\mathtt{P}$ and rules ρ based on such literals is called the *agent's internal language*. Then, formulas ϕ, ϑ of the $\mathcal{ML}$ language are given by

$$\phi ::= \mathsf{B}\lambda \mid \mathsf{B}\rho \mid \neg\phi \mid \phi \vee \vartheta \mid \Diamond\phi$$

Formulas of the form $\mathsf{B}\lambda$ ($\mathsf{B}\rho$) are read as *"the agent believes the literal λ (the rule ρ)"*. Formulas of the form $\Diamond\phi$ are read as *"after a rule application, ϕ is true"*. Similar to Duc's $\mathcal{L}_{BDE}$, the agent's beliefs are restricted, this time to literals and rules based on them; also, the language can express the agent's beliefs and how they change, but cannot express what happens in the real world. ◀

A model for this language is again a small variation of a possible worlds model. Each world has now associated a subset of the agent's internal language (i.e., a set of literals and rules) representing what she believes at it, and each transition between worlds represents a change in the agent's belief state.

Definition 2.5 ($\mathcal{ML}$-model) A $\mathcal{ML}$-model for the $\mathcal{ML}$ language is a tuple $M = \langle W, R, V\rangle$ where W is a non-empty set of states, R is a binary relation on W, and V is a *labelling function*, assigning a subset of the agent's internal language to each possible world. ◀

The semantic interpretation is the usual one, with formulas of the form $\mathsf{B}\lambda$ and $\mathsf{B}\rho$ simply looking at the contents of the subset of the internal language associated to the evaluation point.

Definition 2.6 (Semantic interpretation) Let (M, w) be a pointed model for the $\mathcal{ML}$ language, with $M = \langle W, R, V\rangle$. The semantic interpretation for negation and disjunction are standard. For the rest,

$$
\begin{array}{lll}
(M,w) \Vdash \mathsf{B}\lambda & \text{iff} & \lambda \in V(w) \\
(M,w) \Vdash \mathsf{B}\rho & \text{iff} & \rho \in V(w) \\
(M,w) \Vdash \Diamond\phi & \text{iff} & \text{there is a } u \in W \text{ such that } Rwu \text{ and } (M,u) \Vdash \phi
\end{array}
$$ ◀

The defined class of models is too general, so it needs to be restricted in order to faithfully represent rule-based reasoning. The following definitions are used to state formally the properties such models should satisfy.

Definition 2.7 (Matching rule) Let w be a state of a $\mathcal{ML}$-model $M = \langle W, R, V\rangle$. A rule ρ of the form $\lambda_1, \ldots, \lambda_n \Rightarrow \lambda$ is *w-matching* if and only if at w the agent believes the rule and all its premises but not its conclusion, that is, $\{\rho, \lambda_1, \ldots, \lambda_n\} \subseteq V(w)$ but $\lambda \notin V(w)$. ◀

Definition 2.8 (ρ-extension of a state) Let w and u be states of a $\mathcal{ML}$-model $M = \langle W, R, V\rangle$, and let ρ be a rule. The state u *ρ-extends the state w* if and only if $V(u)$ extends $V(w)$ with ρ's conclusion, that is, $V(u) = V(w) \cup \{\mathsf{cn}(\rho)\}$. ◀

Definition 2.9 (Terminating state) A state w in a $\mathcal{ML}$-model M is said to be *terminating* if and only if no rule is w-matching. ◀

With these concepts, we can now present the four requirements a $\mathcal{ML}$-model should satisfy in order to represent rule-based reasoning.

1. For every state w, if a rule ρ is w-matching, then there is a state u such that Rwu and u is a ρ-extension of w.

2. For every terminating state w there is a state u such that Rwu and, moreover, $V(w) = V(u)$.

3. For every states w, u, we have Rwu only if w and u satisfy one of the previous two points, that is, either $V(u) = V(w) \cup \{\mathsf{cn}(\rho)\}$ for some rule ρ, or else $V(u) = V(w)$.

4. For all rules ρ and states w, u, we have $\rho \in V(w)$ if and only if $\rho \in V(u)$.

The first requirement tells us that if the agent can apply a rule, then there should always be a state that results from the application. The second one says that if there are no applicable rules, the agent should still be able to perform reasoning steps, but they should not change her beliefs. The third one states that no other reasoning steps are allowed and the fourth one, following a standard *AI* practice, says that the rules the agent believes should not change: they are neither learnt nor forgotten.

The two discussed approaches have interesting proposals: the representation of inference as a modal relation (Duc's *Dynamic Syntactic Epistemic Logic*), and the use of a generalized version of modus ponens and the formalization of the precondition and the effect of a rule application (Jago's *Logic for Rule-Based Agents*). Within these two frameworks we can represent agents that are, indeed, non-omniscient. And not only that; the represented agents are also capable of extending their information by performing the adequate reasoning steps.

But there is still room for improvement. From an 'inference' perspective, we can be more precise about the agent's reasoning steps by specifying, syntactically and semantically, which one is the action (i.e., the rule) that the agent is actually performing (i.e., applying). From a 'meta-inference' perspective, the rules the agent can apply do not need to be given by the same set at any time: we can dynamize once more by looking at possible ways in which the agent's *rules* can change. Even from an 'expressiveness' perspective, we can extend the language in order to be able to express not only the agent's information and how it evolves, but also what happens in the real world.

Still, the most important point has to do with the representation of inference. The two discussed works represent inference as a modal relation, and in order to get a proper representation, the relation should satisfy several requirements (Definition 2.2 for Duc's case; the paragraph below Definition 2.9 for Jago's one). All these requirements make the frameworks not as clear as we would like, and we could expect for it to be really confusing when we incorporate more actions to the picture.

There is another possibility. Actions can be represented not only as relations within the model (the *Propositional Dynamic Logic* style); they can also be represented as *operations* that change the model (the *Dynamic Epistemic Logic* style). Instead of defining a model that represents not only the agent's information, but also *all* the possible paths the agent's reasoning steps can follow (what Duc and Jago do), we can define a model that represents exclusively the information the agent has at a given stage, and then define operations that change this model (and therefore the agent's information) in different ways. The advantage of the latter is that we do not need to ask for a relation to satisfy certain requirements; we just need to define reasonable operations. More importantly, representing inference as a model operation will facilitate the incorporation of our other relevant action, *explicit observation*.

2.3 Implicit and explicit information

Recall that our goal is to represent the way an agent's information evolves through the use of truth-preserving inferences and observations. As mentioned before, the *EL* framework with possible worlds models is one of the most widely used for representing and reasoning about agents' information. Nevertheless, in its traditional form, it is not fine enough for our purposes since, as we have discussed, agents whose information is represented with this framework are logically omniscient. Though this feature is useful in some applications, it is too much in some others and, more importantly, it *hides* the inference process. In fact, when representing the *restaurant* example with a standard possible worlds model, the answer to the second question makes the waiter know not only that your father should get the meat dish, but also that your mother should get the vegetarian one. In this case, the hidden inference is short and very simple, but in general this is not the case. Proving a theorem, for example, consists on successive applications of deductive inference steps to show that the conclusion indeed follows from the premises. Some theorems may be straightforward but, as we know, some are not. Moreover, the distinction does not correspond to immediate notions like the number of inference steps, and may be related with the 'complexity' of each one of them, whatever this 'complexity' is.

Now, truth-preserving inference over a notion of information that is already closed under logical consequence becomes irrelevant: it does not provide new information. So our goal should not be to represent inference over what the classical modal operator $\Box$ represents. In fact, this operator should not be understood as 'full-blooded knowledge', but as a more *implicit* notion, describing not the information the agent actually has, but rather the information she can eventually get. With this idea in mind, closure under logical consequence is not a problem anymore because we do expect for *implicit* information to have such property. What we need now is to extend *EL* to provide an adequate

representation for another 'weaker' notion in which truth-preserving inference is actually meaningful: *explicit* information. Only then we will be able to represent inferences and observations together in the proper way.

2.3.1 Formulas, rules and the implicit/explicit language

In our framework, the agent's explicit information will be given by a set of formulas and rules, with these rules being the mechanism through which the agent will be able to increase this explicit information. In other words, our agent can have information not only about the way the world is (given by the formulas), but also about the actions she can perform to increase her explicit information (given by the rules).

We start by defining the language to represent the explicit information the agent can have, and by indicating what a rule in that language is.

Definition 2.10 (Formulas and rules in $\mathcal{L}_P$) Let $\mathtt{P}$ be a set of atomic propositions. Formulas γ, δ of the propositional language $\mathcal{L}_P$ are given by the rule

$$\gamma ::= p \mid \neg\gamma \mid \gamma \vee \delta$$

with p an atomic proposition in $\mathtt{P}$.

A *rule* ρ based on the propositional language is given by

$$\rho ::= (\{\gamma_1, \ldots, \gamma_{n_\rho}\}, \delta)$$

In words, a rule ρ is a pair, sometimes represented as $\{\gamma_1, \ldots, \gamma_{n_\rho}\} \Rightarrow \delta$, where $\{\gamma_1, \ldots, \gamma_{n_\rho}\}$ is a *finite* set of formulas and δ is a formula, all of them in $\mathcal{L}_P$. While formulas describe situations about the world, rules describe relations between such situations. Intuitively, the rule $(\{\gamma_1, \ldots, \gamma_{n_\rho}\}, \delta)$ tells us that if every $\gamma \in \{\gamma_1, \ldots, \gamma_{n_\rho}\}$ is true, so is δ. We denote by $\mathcal{R}_{\mathcal{L}_P}$ the set of rules based on formulas of $\mathcal{L}_P$, omitting the subindex when no confusion arises. ◄

When dealing with rules, the following definitions will be useful.

Definition 2.11 (Premises, conclusion and translation) Let ρ be a rule of the form $(\{\gamma_1, \ldots, \gamma_{n_\rho}\}, \delta)$. We define

$$\begin{array}{ll} \mathsf{pm}(\rho) := \{\gamma_1, \ldots, \gamma_{n_\rho}\} & \text{the } \textit{set of premises} \text{ of } \rho \\ \mathsf{cn}(\rho) := \delta & \text{the } \textit{conclusion} \text{ of } \rho \end{array}$$

Moreover, we define a rule's *translation*, $\mathsf{tr}(\rho)$, as an implication in $\mathcal{L}_P$ whose antecedent is the (finite) conjunction of the rule's premises and whose consequent is the rule's conclusion:

$$\mathsf{tr}(\rho) := \Big(\bigwedge_{\gamma \in \mathsf{pm}(\rho)} \gamma \Big) \rightarrow \mathsf{cn}(\rho)$$

◄

The rules we have defined are simple implications with a special notation. We could have defined them as rules schemas (based on meta-variables to be substituted by formulas) and then their application would be a non-deterministic operation (instantiating the rule and then accepting the instantiated rule's conclusion). But our approach is not a limitation since, as we will see, rules will be applied in a generalized modus ponens way: if all the premises have been accepted, then the conclusion can be accepted. Then we can mimic the application of (instances of) other rules, like *conjunction elimination* ($p \wedge q \Rightarrow p$) or *disjunction introduction* ($p \Rightarrow p \vee q$).

Note also that we have defined the premises of a rule as a *set*, and not as a more general notion like an *ordered sequence* or a *multi-set*. With such a generalized definition, it would be possible to analyze inference in 'resource-conscious' sub-structural logics where order and multiplicity matters, like *Linear Logic* (Girard 1987) or *Categorial Grammar* (Moortgat 1997). Nevertheless, our restricted definition is good enough for dealing with the process we are interested in, *truth-preserving inference*, which will be explored in Section 2.4.[1]

Finally, we could simplify the approach by defining formulas as rules with empty premises. Nevertheless, we will stick to the formulas-and-rules setting since it emphasizes the difference between 'factual' information (the formulas; what the agent already has) and 'procedural' information (the rules; the tools to perform derivations).

The language to reason about the agent's information extends that of *EL* by adding two kinds of formulas: one for expressing the agent's explicit information ($\mathrm{A}\,\gamma$) and the other for expressing the rules she can apply ($\mathrm{R}\,\rho$).

Definition 2.12 (Language $\mathcal{IE}$) Let P be a set of atomic propositions. Formulas φ, ψ of the *implicit/explicit language* $\mathcal{IE}$ are given by

$$\varphi ::= p \mid \mathrm{A}\,\gamma \mid \mathrm{R}\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \Box\varphi$$

with $p \in \mathsf{P}$, $\gamma \in \mathcal{L}_P$ and $\rho \in \mathcal{R}$. Formulas of the form $\mathrm{A}\,\gamma$, access formulas, are read as *"the agent is explicitly informed about γ"*, and formulas of the form $\mathrm{R}\,\rho$, rule formulas, are read as *"the agent can apply rule ρ"*. The universal modal operator, $\Box$, is now interpreted as implicitly information, with formulas of the form $\Box\varphi$ being read as *"the agent is implicitly informed about φ"*. Other boolean connectives ($\wedge$, $\rightarrow$ and $\leftrightarrow$), logical constants ($\top$ and $\bot$) as well as the existential modal operator $\Diamond$ are defined as usual. ◄

Our agent can have explicit information about facts, but not about her own (or, eventually, other agents') information. This is indeed a limitation, but it allows us to define one of the two processes we are interested in: *observation*

[1] In fact, sets are good enough even for dealing some forms of non-monotonic reasoning, as shown in Chapters 5 and 6.

(Section 2.5). In Section 2.7 we discuss the reasons for this limitation, leaving a deeper analysis and further proposals for the next chapters.

The semantic model extends a possible worlds model by assigning two new sets to each possible world: one indicating the *formulas* the agent is explicitly informed about, and other indicating the *rules* she can apply. We still have just one relation between worlds, the accessibility relation, indicating which the worlds the agent considers possible from a given one.

Definition 2.13 (Implicit/explicit model) Let $\mathtt{P}$ be a set of atomic propositions. An *implicit/explicit model* is a tuple $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$ where $\langle W, R, V\rangle$ is a possible worlds model over $\mathtt{P}$ and

- $\mathsf{A} : W \to \wp(\mathcal{L}_P)$ is the *access set function*, indicating the agent's explicit information at each possible world. The set $\mathsf{A}(w)$ will be called the agent's *access set at* w, and should be preserved by the accessibility relation: if $\gamma \in \mathsf{A}(w)$ and Rwu, then $\gamma \in \mathsf{A}(u)$ (the *coherence* property for formulas);
- $\mathsf{R} : W \to \wp(\mathcal{R})$ is the *rule set function*, indicating the rules the agent can apply at each possible world. The set $\mathsf{R}(w)$ will be called the agent's *rule set at* w, and should be also preserved by the accessibility relation: if $\rho \in \mathsf{R}(w)$ and Rwu, then $\rho \in \mathsf{R}(u)$ (the *coherence* property for rules).

We denote by **IE** the class of implicit/explicit models. Note again how, just as in the definition of the premises of a rule, the agent's explicit information about formulas and rules is given by a set. ◄

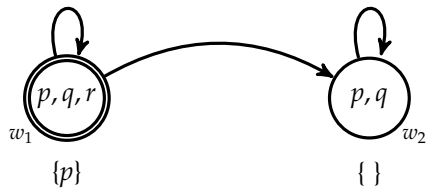
The two model requirements, coherence for formulas and rules, reflect the following idea. At each world w, the sets $\mathsf{A}(w)$ and $\mathsf{R}(w)$ represent the formulas and rules the agent is explicitly informed about. Then, if at w the agent considers u possible, it is natural to ask for u to preserve the agent's explicit information at w. Note also how formulas (rules) in the A-sets (R-sets) are not required to be true (truth-preserving) in the corresponding world. This requirement will be imposed in order to deal with *true* information (Subsection 2.3.2).

The semantic interpretation of formulas in $\mathcal{IE}$ has an immediate definition.

Definition 2.14 (Semantic interpretation) Let (M, w) be a pointed **IE**-model with $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$. The semantic interpretation for negations and disjunctions is given as usual. The case of atomic propositions p and implicit information formulas $\Box\varphi$ is just like in Epistemic Logic. For access and rule formulas, we just look at the corresponding sets:

$$\begin{array}{lll} (M, w) \Vdash \mathrm{A}\,\gamma & \text{iff} & \gamma \in \mathsf{A}(w) \\ (M, w) \Vdash \mathrm{R}\,\rho & \text{iff} & \rho \in \mathsf{R}(w) \end{array}$$ ◄

Example 2.1 In the leftmost world w_1 of following model, the agent has (1) implicit and explicit information about p, (2) implicit but not explicit information about q and (3) neither implicit nor explicit information about r, as indicated by the formulas on the right. Access sets are drawn below the corresponding world, and rule sets are not indicated.



(1) $(M, w_1) \Vdash \Box p \ \wedge \ \mathrm{A}\, p$
(2) $(M, w_1) \Vdash \Box q \ \wedge \ \neg\, \mathrm{A}\, q$
(3) $(M, w_1) \Vdash \neg \Box r \ \wedge \ \neg\, \mathrm{A}\, r$

◀

Axiom system So which properties does the agent's information get under this representation? A standard approach to find them is to look for those formulas that are *valid* in the given class of models, that is, formulas that are true at every world of every model in the given class. There are several ways to look for such formulas, and one of the most commonly used is to look for their syntactic characterization, that is, a *derivation or axiom system*. Such system is a sort of calculus that gives us basic formulas and then operations to derive more formulas. An axiom system is interesting when it only derives formulas valid in a given class of models (a *sound* axiom system), and it becomes even more interesting when, additionally, it derives every valid formula of the given class (a *complete* axiom system).

In order to provide a sound and complete axiom system for formulas in $\mathcal{IE}$ with respect to **IE**-models, it is helpful to look at that for the underlying system: Epistemic Logic (Subsection 1.1.1). The well-known axioms and rules of Table 2.1 provide us with a sound and strongly complete axiom system for the *EL* language with respect to possible worlds models.

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$
K	$\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	*Nec*	If $\vdash \varphi$, then $\vdash \Box\varphi$
Dual	$\vdash \Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$		

Table 2.1: Axiom system for *EL* w.r.t. possible worlds models.

Now we can provide our particular axiom system.

Theorem 2.1 (Axiom system for $\mathcal{IE}$ w.r.t. IE) *The axioms and rules of Tables 2.1 and 2.2 form a sound and strongly complete axiom system for formulas in $\mathcal{IE}$ with respect to* **IE***-models. While axioms and rules of Table 2.1 provide us with validities of the language in possible worlds models, axioms $Coh_{\mathcal{L}_P}$ and $Coh_{\mathcal{R}}$ describe the particular requirements of access and rule formulas for* **IE***-models: the* coherence *property for formulas and rules, respectively. This axiom system is denoted by* IE*.*

$Coh_{\mathcal{L}_P}$	$\vdash \mathrm{A}\,\gamma \to \Box\,\mathrm{A}\,\gamma$
$Coh_{\mathcal{R}}$	$\vdash \mathrm{R}\,\rho \to \Box\,\mathrm{R}\,\rho$

Table 2.2: Axioms for the coherence properties.

Proof. Soundness follows from axioms being valid and rules being validity-preserving. Completeness is proved by a standard modal canonical model construction with an adequate definition of the access and rule set functions. In order to obtain the crucial coherence properties, we use the $Coh_{\mathcal{L}_P}$ and $Coh_{\mathcal{R}}$ axioms. See Appendix A.1 for details. ■

Observe how axioms of Table 2.1 say that the agent's implicit information is omniscient: it contains all validities (rule *Nec*) and it is closed under logical consequence (axiom *K*). Nevertheless, the agent's explicit information does not have these properties: the validity of γ does not imply the validity of $\mathrm{A}\,\gamma$, and $\mathrm{A}\,(\gamma \to \delta) \to (\mathrm{A}\,\gamma \to \mathrm{A}\,\delta)$ is not valid.

2.3.2 The case of *true* information

The way it is defined, an **IE**-model allows us to represent an agent whose information is not necessarily true. We do not ask for any property for the accessibility relation, so there are no constraints for implicit information, others than those given by the representation itself, like closure under modus ponens (the *K* axiom) and the inclusion of validities (the *Gen* rule). In particular, the actual world does not need to be among the ones the agent considers possible, so the agent can be implicitly informed about certain φ without it being true. Moreover, formulas in access sets do not need to be true and rules in rule sets do not need to be truth-preserving at the corresponding world; therefore the agent can be explicitly informed about a formula γ or a rule ρ without they being true and truth-preserving, respectively.

By asking for the adequate model properties, we can represent different notions of information. Here we will focus on the case of *true* information, that is, *knowledge*.[2]

[2]For literature about information that can be true or false, we refer to Dretske (1981) and Floridi (2005).

Among models in **IE**, we distinguish those where implicit and explicit information are true and the rules are truth-preserving. For implicit information, we consider equivalence accessibility relations, as it is usually done in *EL*.[3] For explicit information, we ask for every formula in an access set to be true in the corresponding world. Finally, for the case of rules, we ask for its translation to be true in the corresponding world.

Definition 2.15 (The class $\mathbf{IE}_K$) We denote by $\mathbf{IE}_K$ the class of models $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$ in **IE** satisfying the following properties.

- *Equivalence*: R is an equivalence relation.
- *Truth for formulas*: for every world w, if $\gamma \in \mathsf{A}(w)$, then $(M, w) \Vdash \gamma$.
- *Truth for rules*: for every world w, if $\rho \in \mathsf{R}(w)$, then $(M, w) \Vdash \mathsf{tr}(\rho)$. ◀

Recall the *coherence* property for formulas and rules: if $\gamma \in \mathsf{A}(w)$ ($\rho \in \mathsf{R}(w)$) and Rwu, then we have $\gamma \in \mathsf{A}(u)$ ($\rho \in \mathsf{R}(u)$). Note how, when the accessibility relation is an equivalence relation, we get the same information and rule set for all the worlds that belong to the same equivalence class.

In the rest of this chapter, we will use the term "information" for the general class of **IE**-models, and the term "knowledge" for the class of models with *true* information, that is, $\mathbf{IE}_K$-models.

Axiom system In order to provide a sound and complete axiom system with respect to the just defined class of $\mathbf{IE}_K$-models, we just need to provide axioms that characterize the properties of models in the class.

Theorem 2.2 (Axiom system for $\mathcal{IE}$ w.r.t. $\mathbf{IE}_K$) *The axioms of Tables 2.1, 2.2 and 2.3 form a sound and strongly complete axiom system for formulas in $\mathcal{IE}$ with respect to $\mathbf{IE}_K$-models. In particular, axioms of Table 2.3 characterize the properties that distinguish models in $\mathbf{IE}_K$ from models in* **IE***: equivalence and truth for formulas and rules. This axiom system is denoted by IE_K.*

$T \ \vdash \Box\varphi \to \varphi$	$Tth_{\mathcal{L}_P} \ \vdash \mathrm{A}\,\gamma \to \gamma$
$4 \ \vdash \Box\varphi \to \Box\Box\varphi$	$Tth_{\mathcal{R}} \ \vdash \mathrm{R}\,\rho \to \mathsf{tr}(\rho)$
$5 \ \vdash \neg\Box\varphi \to \Box\neg\Box\varphi$	

Table 2.3: Extra axioms for $\mathbf{IE}_K$-models

[3] Given our understanding of *knowledge* as *true* information, we actually just need for the relation to be reflexive. Nevertheless, we will stick to the standard approach of using equivalence relations that give the agent (implicit) positive and negative introspection.

Proof. Soundness is again simple. Completeness is proved by showing that the canonical model for IE_K satisfies *equivalence* (from axioms *T*, *4*, *5*), *truth for formulas* (from axiom $Tth_{\mathcal{L}_P}$) and *truth for rules* (from axiom $Tth_{\mathcal{R}}$). Details can be found in Appendix A.2. ■

The new axioms provide new properties. Axioms *T*, *4* and *5* tell us that implicit knowledge is *true* (*T*) and has the *positive and negative introspection* properties (*4* and *5*). Axioms $Tth_{\mathcal{L}_P}$ and $Tth_{\mathcal{R}}$ indicate that the agent's explicit knowledge about formulas and rules is also *true*. Note that from the coherence and the truth axioms, $Coh_{\mathcal{L}_P}$ and $Tth_{\mathcal{L}_P}$, we get the following validity, expressing that the agent's explicit knowledge is also implicit knowledge.

$$\mathrm{A}\,\gamma \rightarrow \Box\,\gamma$$

It is now time to turn our attention to the *dynamics* of implicit and explicit knowledge. In the following sections we will define our intended informational actions: truth-preserving inference and observation.

2.4 Inference

The agent can extend her explicit information by applying the rules she has available. Intuitively, a rule (Γ, δ) indicates that if every $\gamma \in \Gamma$ is true, so is δ. However, so far, we have not indicated or restricted the way the agent can use a rule. She can use it to get the conclusion without having all the premises, or even deriving the premises whenever she has the conclusion. In the previous section we focused on models for true information; in the same spirit, this section will deal with truth-preserving inference.

2.4.1 Truth-preserving inference

The inference process aims at extending the agent's *explicit* information by means of a rule-application. Technically, this boils down to adding formulas to the agent's access sets. In order to represent truth-preserving inference, we will restrict the way in which the rule can be applied.

Definition 2.16 (Deduction operation) Let $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$ be an **IE**-model, and let σ be a rule in $\mathcal{R}$. The model $M_{\hookrightarrow_\sigma} = \langle W, R, V, \mathsf{A}', \mathsf{R}\rangle$ differs from M just in the access set function, which is given for every $w \in W$ by

$$\mathsf{A}'(w) := \begin{cases} \mathsf{A}(w) \cup \{\mathsf{cn}(\sigma)\} & \text{if } \mathsf{pm}(\sigma) \subseteq \mathsf{A}(w) \text{ and } \sigma \in \mathsf{R}(w) \\ \mathsf{A}(w) & \text{otherwise} \end{cases}$$

The operation $(\cdot)_{\hookrightarrow_\sigma}$ is called the *deduction operation* with rule σ. ◄

The conclusion of the rule will be added to a world only when all the premises and the rule are already present. In other words, after the inference the agent's explicit information will be extended with the rule's conclusion only if she already has the rule and all its premises.

Note how the deduction operation preserves models in $\mathbf{IE}_K$.

Proposition 2.1 *Let σ be a rule. If M is an $\mathbf{IE}_K$-model, so is $M_{\hookrightarrow_\sigma}$.*

Proof. Equivalence and both properties of rules are immediate since neither the accessibility relation nor the rule set function are modified. The properties of formulas can be verified easily; details can be found in Appendix A.3. ■

The language $\mathcal{IE}_{\mathcal{D}}$ extends $\mathcal{IE}$ by closing it under existential deduction modalities $\langle\hookrightarrow_\sigma\rangle$ for σ a rule: if φ is a formula in $\mathcal{IE}_{\mathcal{D}}$, so is $\langle\hookrightarrow_\sigma\rangle\varphi$. These new formulas are read as *"there is a deductive inference with σ after which φ is the case"* and their universal duals, defined as usual,

$$[\hookrightarrow_\sigma]\varphi := \neg\langle\hookrightarrow_\sigma\rangle\neg\varphi$$

are read as *"after any deductive inference with σ, φ is the case"*.

For the semantic interpretation, note that the agent cannot preform a truth-preserving inference with σ in any situation. In order to do it, she should know explicitly the rule and its premises. The abbreviation

$$\mathrm{Pre}_{\hookrightarrow_\sigma} := \Big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \mathrm{A}\,\gamma\Big) \wedge \mathrm{R}\,\sigma$$

indicates precisely this requirement.

Definition 2.17 (Semantic interpretation) Let (M, w) be a pointed **IE**-model.

$$(M,w) \Vdash \langle\hookrightarrow_\sigma\rangle\varphi \quad \text{iff} \quad (M,w) \Vdash \mathrm{Pre}_{\hookrightarrow_\sigma} \text{ and } (M_{\hookrightarrow_\sigma}, w) \Vdash \varphi$$

By unfolding the definition of the universal deduction modality, we get

$$(M,w) \Vdash [\hookrightarrow_\sigma]\varphi \quad \text{iff} \quad (M,w) \Vdash \mathrm{Pre}_{\hookrightarrow_\sigma} \text{ implies } (M_{\hookrightarrow_\sigma}, w) \Vdash \varphi \qquad \blacktriangleleft$$

The semantic interpretation of deduction modalities simply reflects our intuition about a rule application: the agent can perform a deductive inference with σ after which φ is the case, $(M, w) \Vdash \langle\hookrightarrow_\sigma\rangle\varphi$, if and only if she knows explicitly the rule and its premises, $(M, w) \Vdash \mathrm{Pre}_{\hookrightarrow_\sigma}$, and, after the inference, φ is the case, $(M_{\hookrightarrow_\sigma}, w) \Vdash \varphi$. Note how the new pointed model $(M_{\hookrightarrow_\sigma}, w)$ is defined even if the precondition of the operation $\mathrm{Pre}_{\hookrightarrow_\sigma}$ does not hold at the original (M, w). This is different from the observation operation (Definition 1.5) that makes $(M_{\chi!}, w)$ undefined if the precondition χ fails at (M, w). Nevertheless, the behaviour of the modalities is the same in the two operations: the existential

ones, $\langle \hookrightarrow_\sigma \rangle \varphi$ and $\langle \chi! \rangle \varphi$, are true if and only if the respective actions can be executed and the (unique) resulting pointed model satisfies φ, and the universal ones, $[\hookrightarrow_\sigma] \varphi$ and $[\chi!] \varphi$, are true if and only if either the action can be executed with the specified results, or else the action cannot be executed at all.

Now we can see that in $\mathbf{IE}_K$-models, the deduction operation behaves as expected: if the agent can perform a truth-preserving inference with σ, then after doing it she will know explicitly σ's conclusion. This is because if she can perform the inference at w in M, then the precondition tells us that she knows explicitly the rule and its premises. Because of the truth properties, the premises are true and the rule is truth-preserving at w in M, so the conclusion is also true *at w in M*. But since the conclusion is a propositional formula, its truth-value depends only on the atomic valuation of w; since w's atomic valuation in the new model is exactly the same as in the original one, the conclusion will also be true at w in $M_{\hookrightarrow_\sigma}$. Moreover, this conclusion will be in the access set of w in $M_{\hookrightarrow_\sigma}$; therefore, the agent will know explicitly the rule's conclusion. The following validity express this:

$$\langle \hookrightarrow_\sigma \rangle \, \mathrm{A}\, \mathsf{cn}(\sigma)$$

Note also how, in $\mathbf{IE}_K$-models, if the agent can apply a σ-inference, then σ's conclusion is already in the agent's implicit information. This is because if σ is applicable at w in M, then the agent knows the rule and all its premises, that is, $\mathrm{A}\, \mathsf{pm}(\sigma) \wedge \mathrm{R}\, \sigma$ holds at w. Then, the *coherence* properties put σ and its premises in all worlds R-reachable from w. But then, because of the truth properties, σ's conclusion holds in all of them. Hence, $\Box\, \mathsf{cn}(\sigma)$ is true in the current world. The following validity express this:

$$\langle \hookrightarrow_\sigma \rangle \top \to \Box\, \mathsf{cn}(\sigma)$$

In other words, an act of deduction extends the agent's explicit knowledge by making explicit what was already implicit.

Axiom system In order to provide a sound and complete axiom system for formulas in $\mathcal{IE}$ plus deduction modalities, we will review the sound and complete axiom system for *Observation Logic* (Definition 1.4).

Recall that the idea of a sound and complete axiom system is to characterize validities of a language with respect to a given class of models. We already have a characterization of the validities of the 'static' epistemic language in possible worlds models (Table 2.1); what we need now are axioms and rules describing the relevant properties of the observation modalities.

First, note that the observation operation preserves possible worlds models, that is, if M is a possible worlds model, so is $M_{\chi!}$. This is important because any formula valid in M will still be valid after the operation. Then we have the following rule:

$$!_N \quad \text{From } \vdash \varphi, \text{ infer } \vdash [\hookrightarrow_\sigma] \varphi$$

Now, note that the observation modality has the behavior described in the validities of Table 2.4.

$!_{\mathrm{P}} \vdash \langle \chi! \rangle p \leftrightarrow (\chi \wedge p)$
$!_{\neg} \vdash \langle \chi! \rangle \neg \varphi \leftrightarrow (\chi \wedge \neg \langle \chi! \rangle \varphi)$
$!_{\vee} \vdash \langle \chi! \rangle (\varphi \vee \psi) \leftrightarrow (\langle \chi! \rangle \varphi \vee \langle \chi! \rangle \psi)$
$!_{\Box} \vdash \langle \chi! \rangle \Box \varphi \leftrightarrow (\chi \wedge \Box [\chi!] \varphi)$

Table 2.4: Validities for observation modality.

Let us read some of them. The first indicates that the agent can observe χ and after doing it p will be true if and only if both χ and p are true before the observation. More interestingly, the last one tells us that the agent can observe χ and after doing it she will be (implicitly) informed about φ if and only if χ is true and the agent is (implicitly) informed that after any observation of χ, φ will be true.

These validities give us more than just properties of the observation operation: they provide a way of translating any formula with observation modalities into a *semantically equivalent* one without them. For example, consider the formula $\langle (p \wedge q)! \rangle \Box p$, expressing that $p \wedge q$ can be observed and, after doing it, the agent will be informed about p. By a repeated application of the validities, we can eliminate the observation modalities in the following way:

$$\begin{aligned} \langle (p \wedge q)! \rangle \Box p &\leftrightarrow (p \wedge q) \wedge \Box [(p \wedge q)!] p \\ &\leftrightarrow (p \wedge q) \wedge \Box ((p \wedge q) \rightarrow p) \end{aligned}$$

So $\langle (p \wedge q)! \rangle \Box p$ is semantically equivalent to $(p \wedge q) \wedge \Box ((p \wedge q) \rightarrow p)$. Even if we have a formula with nested occurrences of observation modalities, we can eliminate all of them by following a 'deepest-first' order. For example, eliminating the deepest observation modality from $\langle (p \vee q)! \rangle \langle \neg p! \rangle \Box q$ gives us $\langle (p \vee q)! \rangle (\neg p \wedge \Box (\neg p \rightarrow q))$; then we can eliminate the remaining one. Note how the existence of these validities imply that the 'static' language, the one without observation modalities, is actually expressive enough to encode the way the model will change after the operation.

How are these validities useful when looking for an axiom system? By stating them as axioms, *reduction axioms*, a formula with observation modalities and its translation are not just semantically but also *provably equivalent*. Then, completeness of the language with the observation modality follows from the completeness of the basic 'static' system, since each formula with these modalities can be effectively translated into a provably equivalent one without them. For a more detailed explanation of this technique, we refer to Section 7.4 of van Ditmarsch et al. (2007).

Now we can provide an axiom system for $\mathcal{IE}_{\mathcal{D}}$ with respect to $\mathbf{IE}_K$-models. By Proposition 2.1, this class is closed under the deduction operation, so we can rely on the axiom system IE_K. We provide *reduction axioms*, expressing how deduction operation affect the truth-value of formulas of the language.

Theorem 2.3 (Reduction axioms for the deduction modality) *Table 2.5 provides reduction axioms for the deduction modality. Together with IE_K (Theorem 2.2), they form a sound and complete axiom system for language $\mathcal{IE}_{\mathcal{D}}$ with respect to $\mathbf{IE}_K$-models.*

$\hookrightarrow_p \vdash \langle\hookrightarrow_\sigma\rangle p \leftrightarrow (\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge p)$	$\hookrightarrow_{\mathrm{A}} \vdash \langle\hookrightarrow_\sigma\rangle \mathrm{A}\,\mathsf{cn}(\sigma) \leftrightarrow \mathrm{Pre}_{\hookrightarrow_\sigma}$
$\hookrightarrow_\neg \vdash \langle\hookrightarrow_\sigma\rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge \neg\langle\hookrightarrow_\sigma\rangle\varphi)$	$\hookrightarrow_{\mathrm{A}} \vdash \langle\hookrightarrow_\sigma\rangle \mathrm{A}\,\gamma \leftrightarrow (\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge \mathrm{A}\,\gamma)$ for $\gamma \neq \mathsf{cn}(\sigma)$
$\hookrightarrow_\vee \vdash \langle\hookrightarrow_\sigma\rangle(\varphi \vee \psi) \leftrightarrow (\langle\hookrightarrow_\sigma\rangle\varphi \vee \langle\hookrightarrow_\sigma\rangle\psi)$	$\hookrightarrow_{\mathrm{R}} \vdash \langle\hookrightarrow_\sigma\rangle \mathrm{R}\,\rho \leftrightarrow (\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge \mathrm{R}\,\rho)$
$\hookrightarrow_\Box \vdash \langle\hookrightarrow_\sigma\rangle\Box\varphi \leftrightarrow (\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge \Box[\hookrightarrow_\sigma]\varphi)$	
$\hookrightarrow_N$ From $\vdash \varphi$, infer $\vdash [\hookrightarrow_\sigma]\varphi$	

Table 2.5: Axioms and rule for the deduction modality.

Proof. Soundness follows from the validity of the new axioms and the validity-preserving property of the new rule, just as before. Strong completeness follows from the fact that, by a repeated application of the reduction axioms, any deduction operation formula can be reduced to a formula in $\mathcal{IE}$, for which IE_K is strongly complete with respect to $\mathbf{IE}_K$. ■

The interesting reduction axioms, indicating how access and rule sets are affected by deduction, appear on the right column of Table 2.5. Axioms $\hookrightarrow_{\mathrm{A}}$ indicate that $\mathsf{cn}(\sigma)$ is the unique formula added to access sets, and axiom $\hookrightarrow_{\mathrm{R}}$ indicates that rule sets are not modified.

As an example of what can be derived with the axiom system, consider the formula $\langle\hookrightarrow_\sigma\rangle\top \rightarrow \Box\,\mathsf{cn}(\sigma)$. Its validity was already justified by a semantic argument, but it can also be justified syntactically. [4]

$$
\begin{array}{lll}
\langle\hookrightarrow_\sigma\rangle\top & \leftrightarrow \langle\hookrightarrow_\sigma\rangle(p \vee \neg p) & \\
& \leftrightarrow \langle\hookrightarrow_\sigma\rangle p \vee \langle\hookrightarrow_\sigma\rangle\neg p & \text{by } \hookrightarrow_\vee \\
& \leftrightarrow \Big(\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge p\Big) \vee \Big(\mathrm{Pre}_{\hookrightarrow_\sigma} \wedge \neg p\Big) & \text{by } \hookrightarrow_p, \hookrightarrow_\neg \text{ and Prop. logic} \\
& \leftrightarrow \mathrm{Pre}_{\hookrightarrow_\sigma} & \text{by Prop. logic} \\
& \leftrightarrow \Big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \mathrm{A}\,\gamma\Big) \wedge \mathrm{R}\,\sigma & \text{def. of } \mathrm{Pre}_{\hookrightarrow_\sigma} \\
& \rightarrow \Big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \Box\,\mathrm{A}\,\gamma\Big) \wedge \Box\,\mathrm{R}\,\sigma & \text{by } \mathit{Coh}_{\mathcal{L}_P} \text{ and } \mathit{Coh}_{\mathcal{R}} \\
& \rightarrow \Big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \Box\,\gamma\Big) \wedge \Box\Big(\big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \gamma\big) \rightarrow \mathsf{cn}(\sigma)\Big) & \text{by } \mathit{Tth}_{\mathcal{L}_P} \text{ and } \mathit{Tth}_{\mathcal{R}}. \\
& \leftrightarrow \Box\Big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \gamma\Big) \wedge \Box\Big(\big(\bigwedge_{\gamma\in\mathsf{pm}(\sigma)} \gamma\big) \rightarrow \mathsf{cn}(\sigma)\Big) & \text{dist. of } \Box \text{ over } \wedge \\
& \rightarrow \Box\,\mathsf{cn}(\sigma) & \text{by } K
\end{array}
$$

[4] Through the whole text, this and other syntactic derivations make a slight abuse of notation.

Comparison with previous works

It is illustrative to make a brief comparison between our proposal and the approaches for inference described in Section 2.2.

Dynamic Syntactic Epistemic Logic From Duc's work we have inherited the syntactic representation of the agent's explicit information and the spirit of inference as rule-application. There are small syntactic differences: (1) while Duc uses a single modality $\langle F \rangle$, standing for *"a course of thought"*, our language is more precise about what this 'course of thought' is by explicitly stating the applied rule; (2) Duc's language does not allow us to talk about the real world, something our language can do. There is also the fact that while Duc's framework just focuses on one notion, *explicit* information, our framework represents explicit and also *implicit* information.

But the most important difference is semantic. While Duc represents inference as a relation between worlds, we represent it as an operation over the model. Consequences of this will be discussed below, but first we will look at how our work relates to the other reviewed approach.

Logic for Rule-Based Agents From Jago's work we have inherited the formal definition of the requirements and the consequences of a rule application. There are small language-related differences, like the ones with respect to Duc's approach plus the fact that Jago's system limits the agent's information to literals and rules built from them. Also, his framework represents only one notion of information.

But again, the main difference is the representation of inference as a modal relation, different from our model operation approach.

Modal relation vs. model operation Following the two described approaches, inference was represented in our earlier proposals (Velázquez-Quesada 2008a) as a modal relation between worlds with a relation R_σ for each inference rule σ the agent can apply. But then, consider the following natural requirements for truth-preserving inference.

1. Inference steps should not modify the ontic (factual) situation.

2. In order to apply a rule, the agent needs the premises and the rule.

3. The application of a rule should preserve explicit factual information the agent had before.

4. Explicit information should be increased by the conclusion of the rule.

5. There should be no other difference between explicit information before and after the rule application.

If we represent inference as a modal relation, several restrictions are required in the semantic model in order to satisfy these requirements, making the treatment somehow confusing. But if we represent inference as a modal operation, we do not need to ask for these properties anymore: they are a consequence of the representation. The deduction operation preserves world-valuation, so the ontic situation is not affected. But not only that: we get automatically the four remaining properties, as the validity of the following formulas shows.

2. $\langle \hookrightarrow_\sigma \rangle \top \to \mathrm{Pre}_{\hookrightarrow_\sigma}$
3. $\mathrm{A}\,\gamma \to [\hookrightarrow_\sigma] \mathrm{A}\,\gamma$
4. $[\hookrightarrow_\sigma] \mathrm{A}\,\mathsf{cn}(\sigma)$
5. $\langle \hookrightarrow_\sigma \rangle \mathrm{A}\,\gamma \to \mathrm{A}\,\gamma$ for $\gamma \neq \mathsf{cn}(\sigma)$

There is another important consequence. In our representation, inference is *functional*: the agent can perform an inference step with σ *every time* she has the rule and all its premises. With a relational representation of inference, this property has to be explicitly required (as Jago does) and, more importantly, is not preserved by model operations: adding formulas to the set can make applicable a rule that was not applicable before. This is relevant for us because the second informational action we want to deal with, *observation*, is semantically represented by a model operation.

2.4.2 Dynamics of truth-preserving inference

Just as the agent's explicit knowledge changes, her inferential abilities can also change. This may be because she gets to know a new rule (by means of an observation; Section 2.5), but it may be also because she *builds* new rules from the ones she already has. For example, from the rules $\{p\} \Rightarrow q$ and $\{q\} \Rightarrow r$, it is possible to derive the rule $\{p\} \Rightarrow r$. It takes one step to derive the new rule, but it will save intermediate steps in future inferences.

In fact this situation, a form of transitivity, represents the application of *cut* over the mentioned rules. In general, inference relations can be characterized by *structural rules*, indicating how to derive new rules from the ones already present. In the case of deduction, we have the structural rules of Table 2.6.

In our setting, each application of a structural rule produces a rule that can be added to the rule set. Note that neither *contraction* nor *permutation* yield a *new* rule, since the premises of our rules are given by a set.[5] On the other hand, *reflexivity*, *monotonicity* and *cut* can produce rules that were not present before.

Definition 2.18 (Structural operations) Let $M = \langle W, R, V, \mathsf{A}, \mathsf{R} \rangle$ be an **IE**-model. The *structural operations* defined below return a model that differs from M just in the rule set function, which in each case is defined in the following way.

[5]This is not to say that order or multiplicity of *inference steps* are irrelevant; given our dynamic approach, they definitely matter, as changes in order or number of inference steps can yield different results. We just mean that order and multiplicity of the *premises* are irrelevant because we represent them as a set, and therefore the two mentioned operations will only generate rules that were already considered.

Reflexivity: $\dfrac{}{\varphi \Rightarrow \varphi}$	*Contraction:* $\dfrac{\psi, \chi, \xi, \chi, \phi \Rightarrow \varphi}{\psi, \chi, \xi, \phi \Rightarrow \varphi}$
Permutation: $\dfrac{\psi, \chi, \xi, \phi \Rightarrow \varphi}{\psi, \xi, \chi, \phi \Rightarrow \varphi}$	*Monotonicity:* $\dfrac{\psi, \phi \Rightarrow \varphi}{\psi, \chi, \phi \Rightarrow \varphi}$
Cut: $\dfrac{\chi \Rightarrow \xi \quad \psi, \xi, \phi \Rightarrow \varphi}{\psi, \chi, \phi \Rightarrow \varphi}$	

Table 2.6: Structural rules for deduction.

Reflexivity Let δ be a formula in $\mathcal{L}_P$ and consider the rule

$$\varsigma_\delta := (\,\{\delta\}\,,\ \delta\,)$$

The rule set function R' of the model M_{Ref_δ} is given, for every $w \in W$, by

$$\mathsf{R}'(w) := \mathsf{R}(w) \cup \{\varsigma_\delta\}$$

The operation $(\cdot)_{\mathrm{Ref}_\delta}$ is called the *reflexivity operation* with formula δ.

Monotonicity Let δ be a formula in $\mathcal{L}_P$ and ς a rule over it. Consider the rule

$$\varsigma' := (\,\mathsf{pm}(\varsigma) \cup \{\delta\}\,,\ \mathsf{cn}(\varsigma)\,)$$

extending ς by adding δ to its premises. The rule set function R' of the model $M_{\mathrm{Mon}_{\delta,\varsigma}}$ is given, for every $w \in W$, by

$$\mathsf{R}'(w) := \begin{cases} \mathsf{R}(w) \cup \{\varsigma'\} & \text{if } \varsigma \in \mathsf{R}(w) \\ \mathsf{R}(w) & \text{otherwise} \end{cases}$$

The operation $(\cdot)_{\mathrm{Mon}_{\delta,\varsigma}}$ is the *monotonicity operation* with formula δ and rule ς.

Cut Let ς_1, ς_2 be rules over $\mathcal{L}_P$ such that the conclusion of ς_1 appears in the premises of ς_2. Consider the rule

$$\varsigma' := \Big((\mathsf{pm}(\varsigma_2) \setminus \{\mathsf{cn}(\varsigma_1)\}) \cup \mathsf{pm}(\varsigma_1)\,,\ \mathsf{cn}(\varsigma_2)\Big)$$

combining ς_1 and ς_2. The rule set function R' of the model $M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}$ is given, for every $w \in W$, by

$$\mathsf{R}'(w) := \begin{cases} \mathsf{R}(w) \cup \{\varsigma'\} & \text{if } \{\varsigma_1, \varsigma_2\} \subseteq \mathsf{R}(w) \\ \mathsf{R}(w) & \text{otherwise} \end{cases}$$

The operation $(\cdot)_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}$ is called the *cut operation* with rules ς_1 and ς_2. ◀

Just like the deduction operation, the three structural operations preserve models in the class $\mathbf{IE}_K$.

Proposition 2.2 *If M is an $\mathbf{IE}_K$-model, then so are M_{Ref_δ}, $M_{\mathrm{Mon}_{\delta,\varsigma}}$ and $M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}$.*

Proof. Coherence and truth for formulas as well as equivalence are immediate, since neither access sets nor accessibility relations are modified. For coherence and truth for rules, see Appendix A.4. ■

The language $\mathcal{IE}^{\mathsf{S}}_{\mathcal{D}}$ extends $\mathcal{IE}_{\mathcal{D}}$ by closing it under existential modalities for structural operations: if φ is in $\mathcal{IE}^{\mathsf{S}}_{\mathcal{D}}$, so are $\langle \mathrm{Ref}_\delta\rangle\,\varphi$, $\langle \mathrm{Mon}_{\delta,\varsigma}\rangle\,\varphi$ and $\langle \mathrm{Cut}_{\varsigma_1,\varsigma_2}\rangle\,\varphi$. The formulas are read as *"there is a way of applying the structural operation after which φ is the case"*. In order to formally define their semantic interpretation, we define the following formulas, stating the precondition of each operation. For uniformity, we define a precondition for reflexivity; since this operation can be defined in any situation, we define it simply as $\top$.

$$\begin{aligned}
\mathrm{Pre}_{\mathrm{Ref}_\delta} &:= \top \\
\mathrm{Pre}_{\mathrm{Mon}_{\delta,\varsigma}} &:= \mathrm{R}\,\varsigma \\
\mathrm{Pre}_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}} &:= \mathrm{R}\,\varsigma_1 \wedge \mathrm{R}\,\varsigma_2 \wedge \Big((\textstyle\bigwedge_{\gamma\in\mathsf{pm}(\varsigma_2)} \mathrm{A}\,\gamma) \rightarrow \mathrm{A}\,\mathsf{cn}(\varsigma_1)\Big)
\end{aligned}$$

Definition 2.19 (Semantic interpretation) Let (M, w) be a pointed **IE**-model:

$$\begin{array}{lll}
(M,w) \Vdash \langle \mathrm{Ref}_\delta\rangle\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Ref}_\delta} \text{ and } (M_{\mathrm{Ref}_\delta}, w) \Vdash \varphi \\
(M,w) \Vdash \langle \mathrm{Mon}_{\delta,\varsigma}\rangle\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Mon}_{\delta,\varsigma}} \text{ and } (M_{\mathrm{Mon}_{\delta,\varsigma}}, w) \Vdash \varphi \\
(M,w) \Vdash \langle \mathrm{Cut}_{\varsigma_1,\varsigma_2}\rangle\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}} \text{ and } (M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}, w) \Vdash \varphi
\end{array}$$

Just as before, the universal modalities of the structural operations are defined as the dual of their corresponding existential versions. Just as before, the unfolding yields the following semantic interpretation

$$\begin{array}{lll}
(M,w) \Vdash [\mathrm{Ref}_\delta]\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Ref}_\delta} \text{ implies } (M_{\mathrm{Ref}_\delta}, w) \Vdash \varphi \\
(M,w) \Vdash [\mathrm{Mon}_{\delta,\varsigma}]\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Mon}_{\delta,\varsigma}} \text{ implies } (M_{\mathrm{Mon}_{\delta,\varsigma}}, w) \Vdash \varphi \\
(M,w) \Vdash [\mathrm{Cut}_{\varsigma_1,\varsigma_2}]\,\varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}} \text{ implies } (M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}, w) \Vdash \varphi
\end{array}$$ ◄

Axiom system In order to provide an axiom system for the new formulas, Proposition 2.2 allows us to rely on the axiom system IE_K once again. Table 2.7 provide axioms indicating how the truth value of formulas *after* the structural operations depends on the truth value of formulas *before* them.

Theorem 2.4 (Reduction axioms for structural modalities) *Let* STR *stand for either* Ref_δ, $\mathrm{Mon}_{\delta,\varsigma}$ *or* $\mathrm{Cut}_{\varsigma_1,\varsigma_2}$, *and let ς' stand for the corresponding new rule in each case. Table 2.7 provides reduction axioms for the structural modalities. Together with IE_K (Theorem 2.3), they form a sound and complete axiom system for language $\mathcal{IE}^{\mathsf{S}}_{\mathcal{D}}$ with respect to $\mathbf{IE}_K$-models.*

$\mathrm{STR}_p \vdash \langle \mathrm{STR} \rangle p \leftrightarrow (\mathrm{Pre}_{\mathrm{STR}} \wedge p)$	$\mathrm{STR}_{\mathrm{A}} \vdash \langle \mathrm{STR} \rangle \mathrm{A}\,\gamma \leftrightarrow (\mathrm{Pre}_{\mathrm{STR}} \wedge \mathrm{A}\,\gamma)$
$\mathrm{STR}_\neg \vdash \langle \mathrm{STR} \rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{\mathrm{STR}} \wedge \neg\langle \mathrm{STR} \rangle \varphi)$	$\mathrm{STR}_{\mathrm{R}} \vdash \langle \mathrm{STR} \rangle \mathrm{R}\,\varsigma' \leftrightarrow \mathrm{Pre}_{\mathrm{STR}}$
$\mathrm{STR}_\vee \vdash \langle \mathrm{STR} \rangle (\varphi \vee \psi) \leftrightarrow (\langle \mathrm{STR} \rangle \varphi \vee \langle \mathrm{STR} \rangle \psi)$	$\mathrm{STR}_{\mathrm{R}} \vdash \langle \mathrm{STR} \rangle \mathrm{R}\,\sigma \leftrightarrow (\mathrm{Pre}_{\mathrm{STR}} \wedge \mathrm{R}\,\sigma)$ for $\sigma \neq \varsigma'$
$\mathrm{STR}_\Box \vdash \langle \mathrm{STR} \rangle \Box\varphi \leftrightarrow (\mathrm{Pre}_{\mathrm{STR}} \wedge \Box[\mathrm{STR}]\varphi)$	
STR_N From $\vdash \varphi$, infer $\vdash [\mathrm{STR}]\varphi$	

Table 2.7: Axioms and rules for the reflexivity, monotonicity and cut modalities.

Proof. Just like the reduction axioms for the deduction modality, soundness follows from the validity of the new axioms and the validity-preserving property of the new rules. Strong completeness follows from the fact that, by a repeated application of such axioms, any structural operation formula can be reduced to a formula in $\mathcal{IE}_{\mathcal{D}}$, for which we already have a sound and strongly complete axiom system with respect to $\mathbf{IE}_K$-models. ■

The three structural operations have similar reduction axioms. The difference between them is the precondition for each one to take place, and the new rule each one introduces. While the reflexivity operation with δ can be performed in any case, adding the rule $\{\delta\} \Rightarrow \delta$, the monotonicity operation with δ and ς can be performed only if the agent has already the rule ς, producing a rule that extends ς's premises with δ. Finally, the cut operation with ς_1 and ς_2 can be performed only if ς_2's premises include ς_1's conclusion and the agent has already these both rules, producing a rule whose premises are those of ς_2 minus ς_1's conclusion plus those of ς_1, and whose conclusion is that of ς_2.

The relevant axioms of Table 2.7 are those expressing how rule sets are affected by structural operations; from them we can derive validities analogous to those given at the end of Section 2.4.1 for the case of access sets and deduction.

2.4.3 Combining dynamics

Strictly speaking, we do not need axioms relating deduction and structural operations. We can focus on the deepest occurrence of them, apply the corresponding reduction axioms to eliminate it and then proceed with the next until we remove all the operation modalities. Nevertheless, it is interesting to see how the operations interact between them; in particular, it is interesting to see how deduction is affected by structural operations.

We finish this section presenting the validities of Table 2.8, expressing how deduction after structural operations is related to deduction before them. For each structural operation, the first formula indicates that the operation does not affect deduction with a rule different from the new one, and the second indicates how deduction with the new rule changes. For this last case, the

formula presents a disjunction of two possibilities: the new rule was already in the original rule set (so just deduction is needed) or it was not (so we ask for some requisites). As an example, the second formula for monotonicity indicates that a sequence of this operation and then deduction with the generated rule ς' is equivalent to a single deduction with ς' (if ς' was already present) or to a sequence of deduction with ς and then monotonicity with the agent having explicitly knowledge about the added premise δ and the original rule ς. See Appendix A.5 for details about the validity proofs.

Reflexivity with ς_δ the rule $\{\delta\} \Rightarrow \delta$
$\langle \mathrm{Ref}_\delta \rangle \langle \hookrightarrow_\sigma \rangle \varphi \;\leftrightarrow\; \langle \hookrightarrow_\sigma \rangle \langle \mathrm{Ref}_\delta \rangle \varphi$ for $\sigma \neq \varsigma_\delta$
$\langle \mathrm{Ref}_\delta \rangle \langle \hookrightarrow_{\varsigma_\delta} \rangle \varphi \;\leftrightarrow\; \big(\langle \hookrightarrow_{\varsigma_\delta} \rangle \varphi \vee (\mathrm{A}\,\delta \wedge \langle \mathrm{Ref}_\delta \rangle \varphi)\big)$
Monotonicity with ς' the rule $\mathsf{pm}(\varsigma) \cup \{\delta\} \Rightarrow \mathsf{cn}(\varsigma)$
$\langle \mathrm{Mon}_{\delta,\varsigma} \rangle \langle \hookrightarrow_\sigma \rangle \varphi \;\leftrightarrow\; \langle \hookrightarrow_\sigma \rangle \langle \mathrm{Mon}_{\delta,\varsigma} \rangle \varphi$ for $\sigma \neq \varsigma'$
$\langle \mathrm{Mon}_{\delta,\varsigma} \rangle \langle \hookrightarrow_{\varsigma'} \rangle \varphi \;\leftrightarrow\; \big(\langle \hookrightarrow_{\varsigma'} \rangle \varphi \vee (\mathrm{A}\,\delta \wedge \mathrm{R}\,\varsigma \wedge \langle \hookrightarrow_{\varsigma} \rangle \langle \mathrm{Mon}_{\delta,\varsigma} \rangle \varphi)\big)$
Cut with ς' the rule $(\mathsf{pm}(\varsigma_2) \setminus \{\mathsf{cn}(\varsigma_1)\}) \cup \mathsf{pm}(\varsigma_1) \Rightarrow \mathsf{cn}(\varsigma_2)$
$\langle \mathrm{Cut}_{\varsigma_1,\varsigma_2} \rangle \langle \hookrightarrow_\sigma \rangle \varphi \;\leftrightarrow\; \langle \hookrightarrow_\sigma \rangle \langle \mathrm{Cut}_{\varsigma_1,\varsigma_2} \rangle \varphi$ for $\sigma \neq \varsigma'$
$\langle \mathrm{Cut}_{\varsigma_1,\varsigma_2} \rangle \langle \hookrightarrow_{\varsigma'} \rangle \varphi \;\leftrightarrow\; \big(\langle \hookrightarrow_{\varsigma'} \rangle \varphi \vee (\mathrm{A}\,\mathsf{pm}(\varsigma_1) \wedge \mathrm{R}\,\varsigma_1 \wedge (\mathrm{A}\,\mathsf{cn}(\varsigma_1) \rightarrow \langle \hookrightarrow_{\varsigma_2} \rangle \langle \mathrm{Cut}_{\varsigma_1,\varsigma_2} \rangle \varphi))\big)$

Table 2.8: Formulas relating structural operations and deduction.

2.5 Observation

So far, our language can express just *internal* dynamics. We can express how deductive steps modify explicit knowledge, and even how structural operations extend the available rules, but we cannot express how knowledge is affected by *external* interaction. We now add the other fundamental source of information; we extend our framework to express the effect of *observations*.

This action has been already studied in a *DEL* setting: an observation is interpreted as an operation that removes those worlds where the observed fact does not hold (Definition 1.5). In our framework we have a finer representation of the agent's information: we distinguish between an implicit form, given by the accessibility relation, and an explicit one, given by the access sets. Then, even after fixing the effect of an observation over the agent's implicit information, there are several possibilities for how the operation will affect the explicit part, each one of them representing a different way in which the

agent processes external information. Here, we present one of the possible definitions, what we call an *explicit observation*.

2.5.1 Explicit observation

Different kinds of observations may affect the agent's explicit information in different ways. For example, if the observation is a formula, one option is to keep the A-sets as before (an *implicit* observation); another possibility is to add a rule without premises that will allow the agent to derive the observation one inferential step later (a *semi-explicit* observation). An *explicit* observation has the most intuitive effect: it adds the observed formula to its corresponding set.

Definition 2.20 (Explicit observation operation) Let $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$ be an **IE**-model, and let χ be a formula of (a rule based on) $\mathcal{L}_P$. The model $M_{\chi!^+} = \langle W', R', V', \mathsf{A}', \mathsf{R}'\rangle$ is given by

- $W' := \{w \in W \mid (M, w) \Vdash \chi\}$ $\quad \big(W' := \{w \in W \mid (M, w) \Vdash \mathsf{tr}(\chi)\}\big)$,
- $R' := R \cap (W' \times W')$

and, for every $w \in W'$,

- $V'(w) := V(w)$,
- $\mathsf{A}'(w) := \mathsf{A}(w) \cup \{\chi\}$ $\quad \big(\mathsf{A}'(w) := \mathsf{A}(w)\big)$,
- $\mathsf{R}'(w) := \mathsf{R}(w)$ $\quad \big(\mathsf{R}'(w) := \mathsf{R}(w) \cup \{\chi\}\big)$.

The operation $(\cdot)_{\chi!^+}$ is called the *explicit observation operation* with χ. ◂

The explicit observation operation behaves just like the observation operation with respect to worlds, accessibility relation and valuation. It removes worlds where the observation (its translation, in case the observation is a rule) does not hold, restricting the accessibility relation to the new domain and leaving unmodified the atomic valuation of the preserved worlds. With respect to access and rule sets, explicitly observing χ adds χ itself to the corresponding set of every world, so the agent will have the observation explicitly, as expected.

The explicit observation operation also preserves $\mathbf{IE}_K$-models.

Proposition 2.3 *Let M be an $\mathbf{IE}_K$-model and let χ be a formula in (a rule based on) $\mathcal{L}_P$. If M is in $\mathbf{IE}_K$, so is $M_{\chi!^+}$.*

Proof. Equivalence is immediate since we go to a sub-model, and the coherence properties are also simple because access (rule) sets are extended uniformly. The interesting property is truth for formulas (rules), and it follows from the fact that χ ($\mathsf{tr}(\chi)$) is propositional, and that atomic valuations of the preserved worlds are not modified. See Appendix A.6 for details. ■

The language $\mathcal{IE}_{\mathcal{D}}^{\mathbf{S}!^+}$ extends $\mathcal{IE}_{\mathcal{D}}^{\mathbf{S}}$ with existential modalities for explicit observations. The new formulas $\langle\chi!^+\rangle\varphi$ are read as *"there is a way of explicitly observing χ after which φ is the case"*. For the agent to observe the formula χ we need for χ to be true; for the agent to observe the rule χ we need for χ to be truth-preserving:

$$\mathrm{Pre}_{\chi!^+} := \begin{cases} \chi & \text{if } \chi \text{ is a formula} \\ \mathsf{tr}(\chi) & \text{if } \chi \text{ is a rule} \end{cases}$$

The semantics of explicit observation formulas is given as follows.

Definition 2.21 (Semantic interpretation) Let (M, w) be a pointed **IE**-model.

$$(M,w) \Vdash \langle\chi!^+\rangle\varphi \quad \text{iff} \quad (M,w) \Vdash \mathrm{Pre}_{\chi!^+} \text{ and } (M_{\chi!^+}, w) \Vdash \varphi$$

The case of its universal counterpart, defined as usual, is given by

$$(M,w) \Vdash [\chi!^+]\varphi \quad \text{iff} \quad (M,w) \Vdash \mathrm{Pre}_{\chi!^+} \text{ implies } (M_{\chi!^+}, w) \Vdash \varphi$$

In words, $\langle\chi!^+\rangle\varphi$ holds at w in M if and only if at w, the agent can observe χ (i.e., χ is true/truth-preserving) and, after explicitly doing it, φ holds. ◂

Axiom system A sound and complete axiom system for the new language with respect to $\mathbf{IE}_K$-models can be given based on those already provided and reduction axioms for the new modality.

Theorem 2.5 (Reduction axioms for the explicit observation modality) *Table 2.9 provides reduction axioms for the explicit observation modality. Together with* IE_K *(Theorem 2.4), they form a sound and complete axiom system for language* $\mathcal{IE}_{\mathcal{D}}^{\mathbf{S}!^+}$ *with respect to* $\mathbf{IE}_K$*-models.* ■

$!^+_p$	$\vdash \langle\chi!^+\rangle p \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge p)$	If χ is a formula:
$!^+_\neg$	$\vdash \langle\chi!^+\rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \neg\langle\chi!^+\rangle\varphi)$	$!^+_{\mathrm{A}}$ $\vdash \langle\chi!^+\rangle \mathrm{A}\,\chi \leftrightarrow \mathrm{Pre}_{\chi!^+}$
$!^+_\vee$	$\vdash \langle\chi!^+\rangle(\varphi \vee \psi) \leftrightarrow (\langle\chi!^+\rangle\varphi \vee \langle\chi!^+\rangle\psi)$	$!^+_{\mathrm{A}}$ $\vdash \langle\chi!^+\rangle \mathrm{A}\,\gamma \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \mathrm{A}\,\gamma)$ for $\gamma \neq \chi$
$!^+_\Diamond$	$\vdash \langle\chi!^+\rangle \Diamond\varphi \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \Diamond\langle\chi!^+\rangle\varphi)$	$!^+_{\mathrm{R}}$ $\vdash \langle\chi!^+\rangle \mathrm{R}\,\rho \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \mathrm{R}\,\rho)$
$!^+_N$	From $\vdash \varphi$, infer $\vdash [\chi!^+]\varphi$	If χ is a rule:
		$!^+_{\mathrm{A}}$ $\vdash \langle\chi!^+\rangle \mathrm{A}\,\gamma \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \mathrm{A}\,\gamma)$
		$!^+_{\mathrm{R}}$ $\vdash \langle\chi!^+\rangle \mathrm{R}\,\chi \leftrightarrow \mathrm{Pre}_{\chi!^+}$
		$!^+_{\mathrm{R}}$ $\vdash \langle\chi!^+\rangle \mathrm{R}\,\rho \leftrightarrow (\mathrm{Pre}_{\chi!^+} \wedge \mathrm{R}\,\rho)$ for $\rho \neq \chi$

Table 2.9: Axioms and rules for explicit observation modality.

The relevant axioms are those indicating how explicit information about formulas and rules is affected, and appear on the right column of the table. The agent is always informed about the observation explicitly after observing it, and any other explicit information was already present before the observation. The axioms look similar to those for deduction and structural operations, but again the important difference is the precondition. While in the case of deduction and structural operations the agent needs to have enough explicit information to extract the new piece, an observation is a more radical informational process: it just need for the observation to be *true* (*truth-preserving*).

We finish this section like the previous one, by presenting some validities expressing how the two informational processes considered in this chapter, truth-preserving inference and observation, interact with each other (details about the proof of their validity can be found in Appendix A.7). Table 2.10 presents two cases, according to whether the observed χ is a formula or a rule. Then we make a further difference, this time according to whether the observation enables the application of a rule or not. The first formula indicates that an observation does not affect deduction when the observation is not part of what the agent needs to perform the inference; the second formula presents the disjunction of two possibilities: the observation was already explicit information or it was not. These principles, together with those of Table 2.8, indicate how external and internal dynamics intertwine when we process information, as it will be shown when reviewing the restaurant example (Section 2.6).

If χ is a formula:		
$\langle \chi!^{+} \rangle \langle \hookrightarrow_{\sigma} \rangle \varphi$	$\leftrightarrow \quad \langle \hookrightarrow_{\sigma} \rangle \langle \chi!^{+} \rangle \varphi$	for $\chi \notin \mathsf{pm}(\sigma)$
$\langle \chi!^{+} \rangle \langle \hookrightarrow_{\sigma} \rangle \varphi$	$\leftrightarrow \quad \Big(\langle \hookrightarrow_{\sigma} \rangle \langle \chi!^{+} \rangle \varphi \vee (\mathrm{A}\,\chi \wedge \langle \hookrightarrow_{\sigma} \rangle \langle \chi!^{+} \rangle \varphi) \Big)$	for $\chi \in \mathsf{pm}(\sigma)$
If χ is a rule:		
$\langle \chi!^{+} \rangle \langle \hookrightarrow_{\sigma} \rangle \varphi$	$\leftrightarrow \quad \langle \hookrightarrow_{\sigma} \rangle \langle \chi!^{+} \rangle \varphi$	for $\chi \neq \sigma$
$\langle \chi!^{+} \rangle \langle \hookrightarrow_{\chi} \rangle \varphi$	$\leftrightarrow \quad \Big(\langle \hookrightarrow_{\chi} \rangle \langle \chi!^{+} \rangle \varphi \vee (\mathrm{R}\,\chi \wedge \langle \hookrightarrow_{\sigma} \rangle \langle \chi!^{+} \rangle \varphi) \Big)$	for $\chi = \sigma$

Table 2.10: Formulas relating explicit observation and deduction.

2.6 Back to the Restaurant

Let us represent the restaurant example with our framework. The new waiter's initial information can be given by a model M with six possible worlds, each one of them indicating a possible distribution of the dishes, and all of them

indistinguishable from each other. For the language, consider atomic propositions of the form p_d where p stands for a person (**f**ather, **m**other or **y**ou) and d stands for some dish (**m**eat, **f**ish or **v**egetarian). The waiter explicitly knows each person will get only one dish, so we can put the rules

$$\rho_1 : \{\mathsf{y}_\mathsf{f}\} \Rightarrow \neg\mathsf{y}_\mathsf{v} \qquad\qquad \rho_2 : \{\mathsf{f}_\mathsf{m}\} \Rightarrow \neg\mathsf{f}_\mathsf{v}$$

and similar ones in each world. Moreover, he explicitly knows that each dish corresponds to one person, so we can add the following rule, among any others

$$\sigma : \{\neg\mathsf{y}_\mathsf{v}, \neg\mathsf{f}_\mathsf{v}\} \Rightarrow \mathsf{m}_\mathsf{v}$$

Let w be the real world, where y_f, f_m and m_v are true. In this initial situation, the waiter does not know neither explicitly nor implicitly that your mother has the vegetarian dish:

$$(M, w) \Vdash \neg\Box\,\mathsf{m}_\mathsf{v} \wedge \neg\,\mathrm{A}\,\mathsf{m}_\mathsf{v}$$

While approaching to the table, the waiter can increase the rules he knows. This does not give him new explicit facts, but allows him to reduce the number of inference steps he will need later. He has ρ_1 and σ, and the conclusion of the first is in the premises of the second, so he can apply *cut* over them, getting

$$\varsigma_1 : \{\mathsf{y}_\mathsf{f}, \neg\mathsf{f}_\mathsf{v}\} \Rightarrow \mathsf{m}_\mathsf{v}$$

Then, we have

$$(M, w) \Vdash \langle \mathrm{Cut}_{\rho_1,\sigma} \rangle \left(\neg\Box\,\mathsf{m}_\mathsf{v} \wedge \neg\,\mathrm{A}\,\mathsf{m}_\mathsf{v} \wedge \mathrm{R}\,\varsigma_1 \right)$$

Moreover, he can apply *cut* again, this time with ρ_2 and ς_1, obtaining the rule

$$\varsigma_2 : \{\mathsf{y}_\mathsf{f}, \mathsf{f}_\mathsf{m}\} \Rightarrow \mathsf{m}_\mathsf{v}$$

Now we have

$$(M, w) \Vdash \langle \mathrm{Cut}_{\rho_1,\sigma} \rangle \langle \mathrm{Cut}_{\rho_2,\varsigma_1} \rangle \left(\neg\Box\,\mathsf{m}_\mathsf{v} \wedge \neg\,\mathrm{A}\,\mathsf{m}_\mathsf{v} \wedge \mathrm{R}\,\varsigma_2 \right)$$

After the answer to the first question, *"Who has the fish?"*, the waiter explicitly knows that you have the fish. Four possible worlds are removed, but he still does not know (neither explicitly nor implicitly) that your mother has the vegetarian dish. Then,

$$(M, w) \Vdash \langle \mathrm{Cut}_{\rho_1,\sigma} \rangle \langle \mathrm{Cut}_{\rho_2,\varsigma_1} \rangle \langle \mathsf{y}_\mathsf{f}!^+ \rangle \left(\neg\Box\,\mathsf{m}_\mathsf{v} \wedge \neg\,\mathrm{A}\,\mathsf{m}_\mathsf{v} \wedge \mathrm{R}\,\varsigma_2 \wedge \mathrm{A}\,\mathsf{y}_\mathsf{f} \right)$$

Then he asks *"Who has the meat?"*, and the answer not only gives him *explicit* knowledge about the fact that your father has the meat, but also gives him *implicit knowledge* about the fact that your mother has the vegetarian dish.

$$(M, w) \Vdash \langle \mathrm{Cut}_{\rho_1,\sigma} \rangle \langle \mathrm{Cut}_{\rho_2,\varsigma_1} \rangle \langle \mathsf{y}_\mathsf{f}!^+ \rangle \langle \mathsf{f}_\mathsf{m}!^+ \rangle \left(\Box\,\mathsf{m}_\mathsf{v} \wedge \neg\,\mathrm{A}\,\mathsf{m}_\mathsf{v} \wedge \mathrm{R}\,\varsigma_2 \wedge \mathrm{A}\,\mathsf{y}_\mathsf{f} \wedge \mathrm{A}\,\mathsf{f}_\mathsf{m} \right)$$

Now he can perform the final inference step:

$$(M, w) \Vdash \langle \mathrm{Cut}_{\rho_1,\sigma} \rangle \langle \mathrm{Cut}_{\rho_2,\varsigma_1} \rangle \langle \mathsf{y}_\mathsf{f}!^+ \rangle \langle \mathsf{f}_\mathsf{m}!^+ \rangle \left(\Box\,\mathsf{m}_\mathsf{v} \wedge \mathrm{R}\,\varsigma_2 \wedge \mathrm{A}\,\mathsf{y}_\mathsf{f} \wedge \mathrm{A}\,\mathsf{f}_\mathsf{m} \wedge \langle \hookrightarrow_{\varsigma_2} \rangle \mathrm{A}\,\mathsf{m}_\mathsf{v} \right)$$

Two structural operations, two explicit observations and one truth-preserving inference are all that is needed.

2.7 Remarks

By extending the possible worlds model with a set of formulas and a set of rules at each possible world, the framework presented in this chapter allows us to make a finer distinction in an agent's information. We can represent not only 'epistemic' implicit information, but also explicit information. With this semantic model, explicit information can be defined in several ways, and the present chapter has explored the option in which the agent's explicit information is given by the set of formulas she has in the evaluation point ($\mathrm{A}\,\varphi$). Then we have asked for extra requirements that produce true implicit and explicit information, that is, implicit and explicit *knowledge*. This merging of syntax and semantics provides us a fine grained structure that allows us to represent the information of non-ideal (i.e., non-omniscient) agents. A list of the static notions introduced in this chapter, including their definition and the relevant properties the model should satisfy, is presented in Table 2.11.

Notion	**Definition**	**Relevant model requirements**
Implicit information	$\Box\,\varphi$	——
Explicit information	$\mathrm{A}\,\gamma$	Coherence (Definition 2.13).
Implicit knowledge	$\Box\,\varphi$	Equivalence accessibility relation.
Explicit knowledge	$\mathrm{A}\,\gamma$	Coherence and truth (Definition 2.15).

Table 2.11: Static notions of information.

On the dynamic side, we have provided a notion of *explicit observation*, similar in its effects to the *observation* act in *DEL*. But our focus is not on explicit versions of acts already defined and studied; providing a finer representation of an agent's information highlights informational acts hidden before. In particular the notion of truth-preserving inference, an act that is irrelevant in standard *DEL* due to the omniscient nature of the represented agents, becomes now significant and, moreover, gets an intuitive and clear representation. But there is more. Once our act of rule-based inference has been defined, we have shown how structural operations allow the agent to extend the rules she can apply. In all the cases we have provided the model operation, the corresponding modalities for the language, and reduction axioms that express how the operations modify the truth-value of formulas in the language, therefore allowing us to derive how the notions of information are affected. We have also presented validities describing how deduction, structural operations and observations interact with each other. A list of the reviewed actions and a brief description of their effect is presented in Table 2.12.

Action	Description
Truth-preserving inference.	Turns implicit knowledge into explicit knowledge.
Structural operations.	Add truth-preserving rules the agent can apply.
Explicit observation.	Changes the agent's implicit and explicit knowledge.

Table 2.12: Actions and their effects.

Still, there are some not completely satisfactory points in our proposal. Among them, the most important is the one that limits the agent's explicit information to only propositional formulas, leaving out high-order information (information about her own and, eventually, other agent's information) and information about how actions affect her and other agent's information. The reason for restricting access sets to purely propositional formulas is that, in general, the truth-value of formulas of the full implicit/explicit language is not preserved by the explicit observation operation, and then the operation does not preserve the *truth* property for formulas. Under our definition of explicit information, this property is needed to deal with the case of true information, that is, knowledge.

Let us look at the problem in more detail. In general, a true observed formula of the full implicit/explicit language cannot be simply added to an access set because it may become false after being observed. Classical examples of such cases are Moore sentences of the form $p \wedge \neg\Box p$. Intuitively, an observation of *"p is the case and the agent does not know it (implicitly)"* will make the agent to know (implicitly) p, and therefore the observation is not true anymore. Technically, an explicit observation of $p \wedge \neg\Box p$ keeps only those worlds in which the observation is true *in the original model*, but the operation affects the accessibility relation so $\neg\Box p$ will not be true in the model that results from the operation.

A first attempt to solve this limitation would be to change the definition of the new access set function in order to keep only those formulas that are still true in the new model. Nevertheless, such definition faces circularity. The new access set should contain only those formulas of the original one that are still true in the new one; but, in particular, in order to decide whether an explicit information formula $\mathrm{A}\,\gamma$ is true or not in the new model, we need the new access set, precisely the one we are just defining.

Yang (2009) suggests another possibility. Though it is reasonable to ask for our non-omniscient agent to have true propositional information, maybe it is too much to ask for her to have also true *high-order* information. He suggests to allow arbitrary formulas of the full language in access sets, but restrict the truth property to purely propositional ones. As he mentions, our non-omniscient agent does not need to realize automatically all the high-order

consequences of the observation. Nevertheless, she should be able to realize eventually that some explicit (high-order) information she held correctly before has been 'outdated' by an informational act, and therefore she should be able to correct herself.

There is another option. As we mentioned, the definition of explicit information that we have used is not the only possibility. We will discuss some other alternatives in the following chapter, when we will focus on another action logical omniscience hides: changes in awareness.

Chapter 3

The dynamics of awareness

Logical omniscience is not the only idealization Epistemic Logic agents have. In possible worlds models it is taken for granted not only that the agent can recognize as true all the formulas that are so in the worlds she considers possible; it is also assumed that she can talk about any formula. In other words, a possible worlds model assumes that an agent is *aware of* all formulas of the language. And once again, the idealization leaves out important and interesting actions: this time, changes in awareness.

This chapter focuses on dynamics of the *awareness of* notion. Such changes are an every-day issue in our life; we can easily imagine situations where new possibilities are introduced (you have lost your keys and someone suggest that you may have left them in the kitchen), and others in which some possibilities are dropped (while watching a soccer match we do not usually think about the finite model property of modal logic).

In order to deal with dynamics of a system, we need the system first. We will start by providing a brief summary of a famous framework for dealing with the *awareness of* notion: Fagin and Halpern (1988)'s *Awareness Logic*.

3.1 Awareness Logic

The *awareness logic* of Fagin and Halpern (1988) is based on two observations. First, the modal operator $\Box$ should not be understood as the information the agent actually has, but as the information the agent can eventually get: her implicit information. Second, in order to make explicit her implicit information, the agent should be *aware of* it.

The awareness logic language extends the base language of *EL* with an operator A that allows us to build formulas of the form $\mathrm{A}\varphi$.

Definition 3.1 (Language $\mathcal{L}$) Let $\mathtt{P}$ be a set of atomic propositions. Formulas φ, ψ of the *awareness language* $\mathcal{L}$ are given by

$$\varphi ::= p \mid \mathrm{A}\,\varphi \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi$$

with $p \in \mathtt{P}$. Other Boolean connectives ($\vee, \rightarrow, \leftrightarrow$) as well the existential modal operator ($\Diamond$) are defined as usual. ◀

In this chapter, formulas of the form $\mathrm{A}\,\varphi$ are read as *"the agent is aware of φ"*, and formulas $\Box\varphi$ as *"the agent is informed about φ implicitly"*. The language is interpreted in possible worlds models that assign a set of formulas to the agent in each world, representing in this way the information she is aware of.

Definition 3.2 (Awareness model) Let $\mathtt{P}$ be a set of atomic propositions. An *awareness model* is a tuple $M = \langle W, R, \mathsf{A}, V\rangle$ where $\langle W, R, V\rangle$ is a standard possible worlds model (Definition 1.1), and

- $\mathsf{A} : W \rightarrow \wp(\mathcal{L})$ is the *awareness function*, returning the formulas that the agent 'has in mind'. $\mathsf{A}(w)$ is the agent's *awareness set* at w.

As usual, a *pointed awareness model* (M, w) also has a distinguished world w. ◀

The semantic interpretation of formulas in $\mathcal{L}$ is entirely as expected.

Definition 3.3 (Semantic interpretation) Let (M, w) be a pointed awareness model with $M = \langle W, R, \mathsf{A}, V\rangle$. Atomic propositions and boolean connectives are interpreted as usual; for $\mathrm{A}\,\varphi$ and $\Box\varphi$ we have:

$$\begin{array}{lll} (M,w) \Vdash \mathrm{A}\,\varphi & \text{iff} & \varphi \in \mathsf{A}(w) \\ (M,w) \Vdash \Box\varphi & \text{iff} & \text{for all } u \in W,\ Rwu \text{ implies } (M,u) \Vdash \varphi. \end{array}$$ ◀

Note how, though the syntax and the semantic representation is the same, the *awareness of* notion is conceptually different from the *access* notion we used in the previous chapter. While access sets are understood as 'what the agent has acknowledged as true', being *aware of* is a matter of attention; by saying *"the agent is aware of φ"* we simply indicate that *"the agent entertains φ"*. The concept does not imply any attitude pro or con: the agent may believe φ, but also reject it. Stated in other, but related terms, "awareness of" does not imply "awareness that".

On these models we can impose standard epistemic assumptions about the accessibility relation, such as reflexivity, transitivity, and symmetry. Moreover, further conditions can be imposed on the awareness sets, like closure under commutation for conjunction and disjunction, or being generated by some subset of atomic propositions, according to the specific notion of awareness

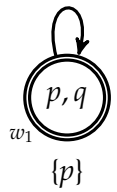
one has in mind.[1] Nevertheless, these requirements are orthogonal to the main concern in this chapter, and we will not assume any of them.

The axiom system for awareness logic is exactly that for the minimal epistemic logic (Table 2.1). Since no special properties about the accessibility relation or the awareness sets are considered, no particular axioms are needed.

Let us turn now to the definition of explicit information. In order for the agent to have explicit information about some formula, besides having it as implicit information, the agent should be *aware* of it. In other words, the agent needs *"to be aware of a concept before [she] can have beliefs about it"* (Fagin and Halpern 1988). This yields the following definition for explicit information:

$$\Box\varphi \wedge \mathrm{A}\,\varphi$$

Example 3.1 In the one-world model below, the agent is implicitly informed that p and also that q. But while she is aware of p, she is not aware of q, so her explicit information about p and q differs.



$$(M, w_1) \Vdash \Box p \wedge \Box q$$
$$(M, w_1) \Vdash \mathrm{A}\,p \wedge \neg\,\mathrm{A}\,q$$
$$(M, w_1) \Vdash (\Box p \wedge \mathrm{A}\,p) \wedge \neg(\Box q \wedge \mathrm{A}\,q)$$

◀

Leaving the rule set function and rule formulas aside, there are three main differences between awareness logic and the framework we presented in the previous chapter. First, A-sets are now interpreted as what the agent is aware of, different from the former "agent's explicit information". Second, these sets are allowed to have *any* formula of the awareness language, without restricting them to the propositional ones like we did. Third, and maybe more interestingly, explicit information is defined now as implicit information plus awareness, $\Box\varphi \wedge \mathrm{A}\,\varphi$, different from the $\mathrm{A}\,\varphi$ we used before.

3.2 Other options for explicit information

As we have mentioned before, several authors coincide in that the $\Box$ operator should not be understood as 'full-blooded information' representing what the agent actually has, but as a notion of *implicit* information, representing what

[1] Such conditions are studied in depth in Fagin and Halpern (1988) and, more recently, in Halpern (2001).

she can eventually get. But when it comes to defining the finer notion of *explicit information* there are different opinions. Even in frameworks similar to the one we have presented in the previous chapter there are variations. Let us leave aside the interpretation of the A-sets for a moment, and review the options.

Explicit information as a primitive notion In the previous chapter we assumed a *primitive* notion of explicit information given by the function A assigning a set of formulas to each possible world. For a proper representation of the knowledge notion, we needed to assume that all formulas in such sets were not only preserved by the accessibility relation but also *true* in the corresponding world. Since the last property is not preserved by standard model operations, we had to restrict formulas in A-sets to those whose truth-value is not affected by such changes: purely propositional formulas.

Explicit information as a defined notion The notion of explicit information can also be defined as a combination of $\Box$ and A. Fagin and Halpern (1988) already provides us one candidate, $\Box\varphi \wedge \mathrm{A}\,\varphi$, which says that φ is explicit information whenever it is implicit information and belongs to the A-set of the evaluation point.

Another interesting option arises when we take a closer look to the consequences of our requirements for dealing with knowledge in the previous chapter. We asked for propositional formulas γ in A-sets to be true (truth: $\mathrm{A}\,\gamma \to \gamma$), for these formulas to be preserved by the accessibility relation (coherence: $\mathrm{A}\,\gamma \to \Box\,\mathrm{A}\,\gamma$) and for this relation to be reflexive ($\Box\varphi \to \varphi$) (also transitive and symmetric). Note how these properties gives us the following equivalence.

$$\begin{array}{lll} \mathrm{A}\,\gamma & \leftrightarrow \Box\,\mathrm{A}\,\gamma & \text{by coherence } (\to) \text{ and reflexivity } (\leftarrow) \\ & \leftrightarrow \Box(\gamma \wedge \mathrm{A}\,\gamma) & \text{by truth } (\to) \text{ and propositional logic } (\leftarrow) \end{array}$$

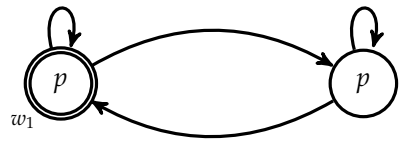
In $\mathbf{IE}_K$-models, our definition of explicit information is equivalent to $\Box(\gamma \wedge \mathrm{A}\,\gamma)$.

What is interesting here is how $\Box(\gamma \wedge \mathrm{A}\,\gamma)$ encodes the coherence and truth requirements. The formula asks directly for γ to be present in the A-set of all R-accessible worlds and for it to be true in each one of them; hence these two properties are not needed anymore. Now, while coherence was a requirement for our most general class of models, truth and reflexivity were required for dealing with the particular case of true implicit and explicit information, that is, implicit and explicit knowledge. But if we define explicit *information* as $\Box(\varphi \wedge \mathrm{A}\,\varphi)$, then just reflexivity is needed in order to have true implicit *and* explicit information, analogous to what happen in classical Epistemic Logic.

This alternative definition has several advantages. The most important is that since the truth property is no longer necessary, we can lift the restriction of A-sets, allowing them to have *any* formula of the language, and therefore allowing the agent to have explicit information not only about propositional

facts, but also about her own (and eventually) other agent's information, and about how this information will change after actions are performed. Having true formulas in A-sets is not needed anymore since, after an action takes place, our new definition automatically 'recomputes' what is explicit information. Besides that, we have the validity $\Box(\varphi \wedge A\varphi) \rightarrow \Box\varphi$, so explicit information is implicit information in the general case.

These two advantages are also shared by Fagin and Halpern (1988)'s definition of explicit information as $\Box\varphi \wedge A\varphi$, but there is another reason that makes our $\Box(\varphi \wedge A\varphi)$ more appealing. Consider the property of *explicit information having implicit positive introspection*: if the agent is *explicitly* informed about φ, then she has *implicit* information about this. Under Fagin and Halpern (1988)'s definition of explicit information, this property is expressed by the formula $(\Box\varphi \wedge A\varphi) \rightarrow \Box(\Box\varphi \wedge A\varphi)$. But this formula is not valid in awareness models, even when we restrict ourselves to those with transitive accessibility relations, a property that characterizes positive introspection in *EL*. The following model proves it, since at w_1 it satisfies $\Box p \wedge A p$ but not $\Box(\Box p \wedge A p)$:



Why is this undesirable? Implicit information is understood as "the best the agent can do". Then, if the agent does not have implicit information about her explicit information, intuitively she will not be able to make explicit this, that is, she will not be able to achieve explicit positive introspection by herself.

With our alternative definition, the notion of implicit positive introspection is expressed by $\Box(\varphi \wedge A\varphi) \rightarrow \Box\Box(\varphi \wedge A\varphi)$. The formula is not valid in the general class of models, but it is in the class of transitive models. In other words, with our definition, implicit positive introspection depends on the properties of the accessibility relation, just like in classical *EL*. In the same way, considering an euclidean accessibility relation gives us implicit *negative* introspection, witness the validity of $\neg\Box(\varphi \wedge A\varphi) \rightarrow \Box\neg\Box(\varphi \wedge A\varphi)$.

Finally, recall that in classical *EL* we have not only the notion of information, $\Box\varphi$, but also the notion of possibility: $\Diamond\varphi$ says that the agent considers φ possible. Defining explicit information as $\Box(\varphi \wedge A\varphi)$ gives us also a very natural notion: $\Diamond(\varphi \wedge A\varphi)$ says that the agent considers φ *explicitly possible*.

For the mentioned reasons, we will define explicit information as follows:

$$\mathrm{Ex}\,\varphi := \Box(\varphi \wedge A\varphi)$$

Once we have fixed a definition of explicit information, it is time to concentrate on our main issue: *dynamics of awareness*.

3.3 Operations on awareness models

Awareness models suggest a natural and simple dynamics. Though the agent is not logically omniscient, she can get new information by various possibly complex acts. But we want to dig deeper. In line with our definition for explicit information, it also makes sense to look for simple actions transforming models that can be put together to analyze more complex informational acts. We will see later on how these transform explicit information.

Defining the basic actions Our models have two separate components for representing information: the accessibility relation and the awareness sets. The following operations modify these components in a simple way, allowing us to define complex epistemic actions later on.

The *consider* operation represents an "awareness raising" action:

Definition 3.4 (The *consider* operation) Let $M = \langle W, R, \mathsf{A}, V\rangle$ be a model and χ any formula in $\mathcal{L}$. The model $M_{+\chi} = \langle W, R, \mathsf{A}', V\rangle$ is M with its awareness sets extended with χ, that is,

$$\mathsf{A}'(w) := \mathsf{A}(w) \cup \{\chi\} \qquad \text{for every } w \in W \qquad \blacktriangleleft$$

'Considering' extends the formulas that an agent is aware of, but we can also define a *drop* operation with the opposite effect: reducing awareness sets. This fits with the operational idea that agents can expand and shrink the set of issues having their current attention.

Definition 3.5 (The *drop* operation) Let $M = \langle W, R, \mathsf{A}, V\rangle$ be a model and χ a formula in $\mathcal{L}$. The model $M_{-\chi} = \langle W, R, \mathsf{A}', V\rangle$ reduces M's awareness sets by removing χ, that is,

$$\mathsf{A}'(w) := \mathsf{A}(w) \setminus \{\chi\} \qquad \text{for every } w \in W \qquad \blacktriangleleft$$

This operation can be seen as a form of 'forgetting', an action usually disregarded in Dynamic Epistemic Logic (but see van Ditmarsch et al. (2009) and van Ditmarsch and French (2009) for proposals).

The preceding actions affect what an agent is aware of. The next one, known from *DEL*, modifies her implicit information by discarding those worlds where some *observed* formula χ fails:

Definition 3.6 (The *implicit observation* operation) Let $M = \langle W, R, \mathsf{A}, V\rangle$ be a model and χ a formula in $\mathcal{L}$. The model $M_{\chi!} = \langle W', R', \mathsf{A}', V'\rangle$ is given by

- $W' := \{w \in W \mid (M, w) \Vdash \chi\}$
- $R' := R \cap (W' \times W')$

and, for every $w \in W'$,

- $\mathsf{A}'(w) := \mathsf{A}(w)$
- $V'(w) := V(w)$. $\blacktriangleleft$

The observation is implicit because, although it removes worlds, it does not affect what the agent is aware of in the preserved ones.

Building complex actions Complex actions can now be built by combining basic ones. As an example, it seems natural to think that a public observation of some fact is in fact done consciously, generating awareness. The corresponding operation of "explicit observation" can be defined in the following way.

Definition 3.7 The *explicit observation* operation, analogous in its effect to a *public announcement* in *PAL* (Plaza 1989; Gerbrandy 1999), can be defined by means of an implicit observation followed by an act of consideration:

$$M_{\mathrm{EO}(\chi)} := (M_{\chi!})_{+\chi}$$ ◀

The definition also works if we interchange the order of the operations because we are transforming two independent components of our models.[2]

Preserving static constraints Though we have not imposed constraints on the static awareness models, it is interesting to note that some reasonable requirements, like our previous coherence (also called *weak introspection on* A*-sets*) or equivalence relations for accessibility, are preserved by our operations.

Proposition 3.1 Considering *preserves coherence and equivalence relations.*

Proof. The equivalence property of R is obviously preserved, since R is not modified. For coherence, take a world w in $M_{+\chi}$ and any $\varphi \in \mathsf{A}'(w)$. Suppose Rwu. If φ is already in $\mathsf{A}(w)$, then $\varphi \in \mathsf{A}(u)$ because M satisfies the principle, and then $\varphi \in \mathsf{A}'(u)$ by the definition of A'. If φ is not in $\mathsf{A}(w)$, then it should be χ itself, which by definition is also in $\mathsf{A}'(u)$. ■

By a similar argument, the *drop* operation, too, preserves the two mentioned properties.

Proposition 3.2 Dropping*, too, preserves coherence and equivalence relations.* ■

Finally, our actions of implicit observation have the same effect:

Proposition 3.3 *Implicit observation preserves coherence and equivalence relations.*

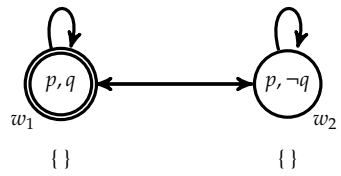
Proof. Equivalence relations are preserved automatically since we go to a sub-model. Next, for coherence, use the fact that the sub-model $M_{\chi!}$ has the same awareness sets at its worlds as M, while its epistemic accessibility is a sub-relation of that for M. ■

It is also worthwhile to notice how some properties one might impose on the A-sets (the truth property, the already mentioned closure under commutation for conjunction and disjunction, or being generated by some subset of atomic propositions, as in Fagin and Halpern (1988)) are not preserved by the operations *consider* and *drop*.

[2]Still, one might argue that implicit observation and considering take place *simultaneously*. While this makes sense, we will not pursue it here.

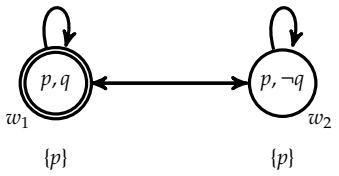
3.4 The actions in action

Consider the following model:

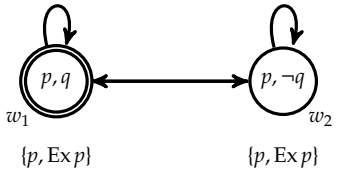


In the leftmost world, w_1, the agent does not have implicit information about q, but she is implicitly informed about p, though not explicitly.

After the agent *considers* p, we get the model on the right: in both worlds, the agent is now explicitly informed about p.

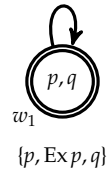


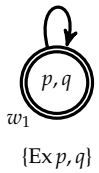
We do not have the truth requirement of the previous chapter anymore, so our agent can also get explicit information about her own awareness, or implicit and explicit information. Here is how this can happen:



When she *considers* $\mathrm{Ex}\, p$, we get the model on the left. The agent has explicit information about her having explicit information of p. By acting, she has achieved positive introspection.

Next, consider the above *explicit observation* of q: an *implicit observation* followed by *consideration* of q. This yields the model on the right where q is now part of the agent's explicit information.





Finally, *dropping* p makes the agent lose earlier explicit information about it (that is, we get $\neg\mathrm{Ex}\, p$). Moreover, by our definition of explicit information, she no longer has explicit information that $\mathrm{Ex}\, p$, since the latter formula is no longer true, and therefore, it is no longer implicit information.[3]

There are many further scenarios with complex many-world patterns, but the above will suffice to show the interest of our setting.

[3]This may seem strange since the formula $\mathrm{Ex}\, p$ is still in the awareness set of the world, but this only means that the agent is aware of it, not that she still endorses it.

3.5 A complete dynamic logic

In order to express how our dynamic operations affect awareness, implicit and explicit information, we extend the static awareness language with modalities representing each basic operation. If χ and φ are formulas in the resulting extended language (still called $\mathcal{L}$ in this section), then so are

$\langle +\chi \rangle \varphi$	*there is a way of considering χ after which φ is the case.*
$\langle -\chi \rangle \varphi$	*there is a way of dropping χ after which φ is the case.*
$\langle \chi! \rangle \varphi$	*there is a way of observing χ implicitly after which φ is the case.*

Definition 3.8 (Semantic interpretation) Let (M, w) be a pointed awareness model and let χ, φ be formulas in the extended language $\mathcal{L}$. Then,

$$
\begin{array}{lll}
(M,w) \Vdash \langle +\chi \rangle \varphi & \text{iff} & (M_{+\chi}, w) \Vdash \varphi \\
(M,w) \Vdash \langle -\chi \rangle \varphi & \text{iff} & (M_{-\chi}, w) \Vdash \varphi \\
(M,w) \Vdash \langle \chi! \rangle \varphi & \text{iff} & (M,w) \Vdash \chi \text{ and } (M_{\chi!}, w) \Vdash \varphi
\end{array}
$$

The universal versions of the modalities are defined as the dual of their respective existential, as usual. ◂

The main difference among the new modalities is the precondition. The agent can consider or drop a formula χ without any further requirement, but for her to implicitly observe χ, χ needs to be *true*. In particular, the lack of precondition and the fact that the operations are functional make the semantic interpretation of the existential and the universal modalities for the consider and the drop operation coincide:

$$
\begin{array}{lll}
(M,w) \Vdash [+\chi] \varphi & \text{iff} & (M_{+\chi}, w) \Vdash \varphi \\
(M,w) \Vdash [-\chi] \varphi & \text{iff} & (M_{-\chi}, w) \Vdash \varphi
\end{array}
$$

3.5.1 Dynamic completeness theorem

We now formulate a sound and complete logic for the semantic validities in the extended language $\mathcal{L}$:

Theorem 3.1 (Reduction axioms for the action modalities) *The valid formulas of the extended awareness language $\mathcal{L}$ in awareness models are those provable by the axioms and rules for the static language (Table 2.1; see Section 3.1) plus the reduction axioms and modal inference rules listed in Table 3.1.* ■

These axioms express the syntactic basics of the considering and dropping operations, merged with the axioms of Observation Logic (Section 1.4). For instance, how do the propositions that the agent is aware of change when the agent *considers* χ? Our axioms show the two possibilities. After considering

$+_p \vdash \langle +\chi\rangle p \;\leftrightarrow\; p$	$+_{\mathrm{A}} \vdash \langle +\chi\rangle \mathrm{A}\,\chi \;\leftrightarrow\; \top$
$+_\neg \vdash \langle +\chi\rangle \neg\varphi \;\leftrightarrow\; \neg\langle +\chi\rangle \varphi$	$+_{\mathrm{A}} \vdash \langle +\chi\rangle \mathrm{A}\,\varphi \;\leftrightarrow\; \mathrm{A}\,\varphi$ for $\varphi \neq \chi$
$+_\vee \vdash \langle +\chi\rangle (\varphi \vee \psi) \;\leftrightarrow\; (\langle +\chi\rangle \varphi \vee \langle +\chi\rangle \psi)$	
$+_\Diamond \vdash \langle +\chi\rangle \Diamond \varphi \;\leftrightarrow\; \Diamond \langle +\chi\rangle \varphi$	
$+_N$ From $\vdash \varphi$, infer $\vdash [+\chi]\varphi$	
$-_p \vdash \langle -\chi\rangle p \;\leftrightarrow\; p$	$-_{\mathrm{A}} \vdash \langle -\chi\rangle \mathrm{A}\,\chi \;\leftrightarrow\; \bot$
$-_\neg \vdash \langle -\chi\rangle \neg\varphi \;\leftrightarrow\; \neg\langle -\chi\rangle \varphi$	$-_{\mathrm{A}} \vdash \langle -\chi\rangle \mathrm{A}\,\varphi \;\leftrightarrow\; \mathrm{A}\,\varphi$ for $\varphi \neq \chi$
$-_\vee \vdash \langle -\chi\rangle (\varphi \vee \psi) \;\leftrightarrow\; (\langle -\chi\rangle \varphi \vee \langle -\chi\rangle \psi)$	
$-_\Diamond \vdash \langle -\chi\rangle \Diamond \varphi \;\leftrightarrow\; \Diamond \langle -\chi\rangle \varphi$	
$-_N$ From $\vdash \varphi$, infer $\vdash [-\chi]\varphi$	
$!_p \;\vdash \langle \chi!\rangle p \;\leftrightarrow\; (\chi \wedge p)$	$!_{\mathrm{A}} \;\vdash \langle \chi!\rangle \mathrm{A}\,\varphi \;\leftrightarrow\; (\chi \wedge \mathrm{A}\,\varphi)$
$!_\neg \;\vdash \langle \chi!\rangle \neg\varphi \;\leftrightarrow\; (\chi \wedge \neg\langle \chi!\rangle \varphi)$	
$!_\vee \;\vdash \langle \chi!\rangle (\varphi \vee \psi) \;\leftrightarrow\; (\langle \chi!\rangle \varphi \vee \langle \chi!\rangle \psi)$	
$!_\Diamond \;\vdash \langle \chi!\rangle \Diamond \varphi \;\leftrightarrow\; (\chi \wedge \Diamond \langle \chi!\rangle \varphi)$	
$!_N$ From $\vdash \varphi$, infer $\vdash [\chi!]\varphi$	

Table 3.1: Axioms and rules for the action modalities.

χ, the agent is aware of a $\varphi \neq \chi$ if and only if she was aware of φ before; but also, considering χ always makes the agent aware of χ. The *drop* operation has an analogous effect in the opposite direction. The rest of the axioms are simple commutation clauses, indicating the independence of modifying the domain of worlds and the awareness sets.

3.5.2 How the logic describes our major issues

Our logic states how each basic operator of the language is affected by our three actions. By combining these effects and unfolding the definitions, the logic also explains how the derived notion of *explicit information* changes under these actions. We discuss a few cases, using our earlier definition $\Box(\varphi \wedge \mathrm{A}\,\varphi)$, and suppressing detailed calculations:

Explicit information For the action of *considering* χ and explicit information about a different formula φ, an application of the reduction axioms gives us the following valid principle

$$[+\chi]\,\mathrm{Ex}\,\varphi \;\leftrightarrow\; \Box\Big([+\chi]\varphi \wedge \mathrm{A}\,\varphi\Big) \qquad (\text{for } \varphi \neq \chi)$$

The principle states that after considering χ the agent will be explicitly informed about φ if and only if she is already implicitly informed that the considering act will make φ true and that she is aware of φ. One might have expected a simpler direct reduction principle $[+\chi]\,\mathrm{Ex}\,\varphi \;\leftrightarrow\; \mathrm{Ex}\,\varphi$, but this formula is not valid in the general case, since the *consider* action may have changed truth values for sub-formulas of φ. Nevertheless, for propositional formulas γ we do have the validity $[+\chi]\,\mathrm{Ex}\,\gamma \;\leftrightarrow\; \mathrm{Ex}\,\gamma$: considering χ does not affect explicit information about propositional facts different from χ.

In the particular case of explicit information about χ itself, however, we get the following.

Fact 3.1 *The formula* $[+\chi]\,\mathrm{Ex}\,\chi \;\leftrightarrow\; \Box\,\chi$ *is valid.*

Proof. Using our reduction axioms, we get

$$\begin{aligned}
[+\chi]\,\mathrm{Ex}\,\chi &\leftrightarrow [+\chi]\,\Box\,(\chi \wedge \mathrm{A}\,\chi)\\
&\leftrightarrow \Box\,[+\chi]\,\chi \wedge \Box\,[+\chi]\,\mathrm{A}\,\chi\\
&\leftrightarrow \Box\,[+\chi]\,\chi\\
&\leftrightarrow \Box\,\chi
\end{aligned}$$

■

The last step is justified by the following proposition.

Proposition 3.4 *The formula* $\chi \leftrightarrow [+\chi]\,\chi$ *is valid.*

Proof. The reason is that, given our semantics, an act of considering χ can only change truth values for $\mathrm{A}\,\chi$ and formulas containing it. But then, χ itself cannot be affected by the operation, since it cannot contain $\mathrm{A}\,\chi$. ■

This shows how a *consider* action makes implicit information explicit.

Now consider the *K* axiom, the one to blame for logical omniscience in *EL*:

$$\Box\,(\varphi \to \psi) \;\to\; (\Box\,\varphi \to \Box\,\psi)$$

This formula is still valid in awareness models, and that is reasonable since implicit information is expected to be closed under logical consequence. But the following formula is also valid

$$\mathrm{Ex}\,(\varphi \to \psi) \;\to\; (\mathrm{Ex}\,\varphi \to \Box\,\psi)$$

The reason is that, since explicit information is also implicit, $\mathrm{Ex}\,(\varphi \to \psi)$ and $\mathrm{Ex}\,\varphi$ already imply $\Box\,(\varphi \to \psi)$ and $\Box\,\varphi$. Then, *considering* is the action that 'fills the gap', turning explicit the formerly implicit information:

$$\mathrm{Ex}\,(\varphi \to \psi) \;\to\; (\mathrm{Ex}\,\varphi \to [+\psi]\,\mathrm{Ex}\,\psi) \quad \text{is valid}$$

One might think that the real act here is a richer one of *drawing the inference*, but in our analysis it is the explicit consideration of the conclusion what 'gives the last little push' toward explicit information.[4]

[4] In Chapter 4 we ask for awareness and acknowledgement of formulas as true in order to get explicit information. Then, the needed acts are those of awareness raising *and* inference.

But our proposal can describe more, including the behaviour of explicit information under the *drop* operation. Here is what happens with formulas φ that differ from the dropped χ:

$$[-\chi]\,\mathrm{Ex}\,\varphi \;\leftrightarrow\; \Box\big([-\chi]\,\varphi \wedge \mathrm{A}\,\varphi\big) \qquad (\text{for } \varphi \neq \chi)$$

For explicit information about χ itself, we get the following.

Fact 3.2 *The formula* $[-\chi]\,\mathrm{Ex}\,\chi \leftrightarrow \Box\,\bot$ *is valid.*

Proof. Using our reduction axioms as above,

$$\begin{aligned}
[-\chi]\,\mathrm{Ex}\,\chi &\leftrightarrow [-\chi]\,\Box\,(\chi \wedge \mathrm{A}\,\chi)\\
&\leftrightarrow \Box\,([-\chi]\,\chi \wedge [-\chi]\,\mathrm{A}\,\chi)\\
&\leftrightarrow \Box\,([-\chi]\,\chi \wedge \bot)\\
&\leftrightarrow \Box\,\bot
\end{aligned}$$

■

The validity states that after dropping χ the agent has explicit information about it if and only if she is implicitly informed about contradictions. In the particular cases of consistent information (technically, seriality for the accessibility relation) or true information (reflexivity), this validity becomes

$$\neg[-\chi]\,\mathrm{Ex}\,\chi$$

read as *"one never has explicit information about χ after dropping it"*.

Still, even after dropping it, the agent does keep χ as implicit information, witness the following valid law:

Fact 3.3 *The formula* $\mathrm{Ex}\,\chi \rightarrow [-\chi]\,\Box\,\chi$ *is valid.*

Proof. Again, using the axioms and unfolding the definitions,

$$\begin{aligned}
\mathrm{Ex}\,\chi &\rightarrow \Box\,\chi \wedge \Box\,\mathrm{A}\,\chi\\
&\rightarrow \Box\,\chi\\
&\rightarrow \Box\,[-\chi]\,\chi\\
&\rightarrow [-\chi]\,\Box\,\chi
\end{aligned}$$

■

Our proof uses the following proposition, whose justification is analogous to the one for $\chi \leftrightarrow [+\chi]\,\chi$ (Proposition 3.4).

Proposition 3.5 *The formula* $\chi \;\leftrightarrow\; [-\chi]\,\chi$ *is valid.* ■

Finally, we analyze the effect of an implicit observation over explicit information. For any φ and χ, unfolding the definition of explicit information via our axioms (we suppress intermediate steps here) gives

$$[\chi!]\,\mathrm{Ex}\,\varphi \leftrightarrow \Big(\chi \to \Box\Big([\chi!]\,\varphi \wedge (\chi \to \mathrm{A}\,\varphi)\Big)\Big)$$

The principle states that after an implicit observation of χ the agent will be explicitly informed about φ if and only if, *conditional to the truth of the observation*, she is already implicitly informed that this observing act will make φ true and that she is aware of φ. This outcome is our solution to the earlier-mentioned problem of update making explicit information 'out of synch' with reality. (Recall that this was the reason for the restriction to purely factual assertions in the previous chapter.) Explicit information is now a defined notion, so it automatically re-adjusts to whatever happens to the modalities $\Box$ and A, and our logic tells us precisely how.

We have extracted the effect of our basic epistemic actions over explicit information defined as $\Box(\varphi \wedge \mathrm{A}\,\varphi)$. Thus, we replace discussion whether the agents's information *is* closed under logical consequence by a much richer picture of what they can *do* to change their information.

Moreover, this style of analysis works not only for the stated notion of explicit information; it can also provide us with validities expressing the way different definitions of explicit information are affected by dynamic actions, like Fagin and Halpern (1988)'s $\Box\varphi \wedge \mathrm{A}\,\varphi$ or others, like $\Box\varphi \wedge \mathrm{A}\,\Box\varphi$.

3.5.3 Schematic validities and algebra of actions

While all this seems a straightforward dynamic epistemic technique, there is a catch. In deriving the principles of the previous section, we have used more than the reduction axioms of our logic per se. Several important 'schematic' principles did not follow from our reduction axioms. In particular, we have used the two principles

$$[+\chi]\,\chi \leftrightarrow \chi \qquad \text{and} \qquad [-\chi]\,\chi \leftrightarrow \chi$$

whose validity involved additional considerations. Of course, each specific instance of such a formula can be derived, given our completeness theorem. But that does not mean there is any illuminating *uniform derivation* of an "algebraic" sort. Indeed, an explicit characterization of the schematic validities in dynamic-epistemic logics (valid for all substitutions of formulas for proposition letters) is a well-known open problem (cf. van Benthem (2010)), even in the case of Public Announcement Logic. Given the importance of such general principles here, that problem becomes even more urgent.

Algebra of actions We end this section with one particular source of schematic validities. As important as it is to understand how actions affect our information, their general algebraic structure is of interest too. We briefly discuss some validities, to show that this *"algebra of actions"* raises some interesting issues:

- In general, *drop* does not neutralize *consider*: $[+\chi]\,[-\chi]\,\varphi \leftrightarrow \varphi$ is not valid. If the agent is initially aware of χ, *consider* makes no change, but *drop* does, yielding a model where χ is not in the awareness set. The actual validity is the qualified

$$\neg\, \mathrm{A}\,\chi \rightarrow ([+\chi]\,[-\chi]\,\varphi \leftrightarrow \varphi)$$

- The dual case behaves in the same way: *consider* does not neutralize *drop* in general, but we do have:

$$\mathrm{A}\,\chi \rightarrow ([-\chi]\,[+\chi]\,\varphi \leftrightarrow \varphi)$$

As for unqualified algebraic laws, we do have idempotence:

- A sequence of *consider* actions for the same formula has the same effect as a single one, and the *drop* operation behaves similarly:

$$[+\chi]\,\varphi \leftrightarrow [+\chi]\,[+\chi]\,\varphi \qquad \text{and} \qquad [-\chi]\,\varphi \leftrightarrow [-\chi]\,[-\chi]\,\varphi$$

Next, given the dynamics of the system, we do not expect strong commutation laws between considering and dropping (the fact that they do not cancel each other gives us a clue). Nevertheless, we do expect commutation of these operations with implicit observation, since the latter modifies an independent component of our models. For example, the following formulas

$$[\chi!]\,[+\chi]\,\varphi \leftrightarrow [+\chi]\,[\chi!]\,\varphi \qquad \text{and} \qquad [\chi!]\,[-\chi]\,\varphi \leftrightarrow [-\chi]\,[\chi!]\,\varphi,$$

are valid even for formulas χ using the modality A. The reason is, once again, that the operations $+\chi$ and $-\chi$ can only change the truth value of $\mathrm{A}\,\chi$, and hence that of χ cannot be affected.

This action algebra, yielding more validities when we restrict our attention to just factual assertions, clearly involves uniform schematic validities that once more are not immediately obvious from our earlier completeness theorem. In fact, *PAL* itself (what we have called observation logic) has an algebra of actions describing the behaviour of successive announcements, but it tends to go unnoticed since two successive announcements can be compressed into a single one, as the following validity indicates:

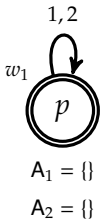
$$[\chi_1!]\,[\chi_2!]\,\varphi \leftrightarrow [(\chi_1 \wedge [\chi_1!]\,\chi_2)!]\,\varphi$$

This compression disappears when the operation changes the accessibility relation, as it is done in dynamic epistemic logics for changes in preferences or beliefs: two successive *upgrades* cannot be compressed into a single one (van Benthem and Liu 2007; van Benthem 2007).

3.6 From single to multi-agent scenarios

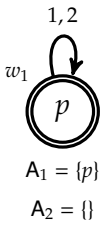
So far, we have considered activities of single agents, including not only their observations, but also their acts of awareness raising. Now, the latter are typically private, and hence it makes sense to look at scenarios with privacy. But a bit paradoxically, privacy only becomes visible in a multi-agent setting. Here is a first simple illustration with two agents:

Example 3.2 Consider the following model M, generalizing the single-agent framework to a multi-agent setting in a straightforward way:



In the unique world of the model, w_1, each agent is implicitly informed about p, but no agent is aware of p ($\neg \mathrm{A}_1 p \wedge \neg \mathrm{A}_2 p$). Moreover, agents have implicit information about each other's lack of awareness about p. E.g., agent 2 is implicitly informed that agent 1 is not aware of p ($\Box_2 \neg \mathrm{A}_1 p$).

Now let an event take place: agent 1 considers p. The model M_{+p^1} is given by



In the new situation, agent 1 is aware of p ($\mathrm{A}_1 p$), and now has explicit information about it. But there is more: though agent 2 does not have any new explicit information, she is now implicitly informed that agent 1 is aware of p ($\Box_2 \mathrm{A}_1 p$).

Is this a realistic scenario? Independently of the modelling, it seems strange that an internal action that takes place only in agent 1's mind can affect immediately the information of agent 2. This shows the need of a more detailed analysis of how awareness models should change in a setting that allows not only public but also private actions. ◄

3.6.1 Multi-agent static framework

The extension of the static awareness framework to a setting with many agents in a group $\mathtt{Ag}$ is straightforward. In the language of multi-agent $\mathcal{L}$, we just add agent indexes to the A and the $\Box$ modalities (A_i and $\Box_i$, respectively). In the semantic models, R becomes a *function* from $\mathtt{Ag}$ to $\wp(W \times W)$ returning an *accessibility relation* R_i for each agent $i \in \mathtt{Ag}$, and A becomes a function from $\mathtt{Ag} \times W$ to $\wp(\mathcal{L})$ returning the *awareness* set $\mathsf{A}_i(w)$ of each agent i at each possible world w. The semantic interpretation of formulas is then as before, using A_i and R_i to interpret formulas of the form $\mathrm{A}_i \varphi$ and $\Box_i \varphi$, respectively.

Again, in this multi-agent case we will not impose special semantic constraints. But it is interesting to notice that if we had been dealing with the notion of *knowledge*, going from public to private actions would have made us to change the notion to one in which information is not required to be true (just as in *DEL* in general). This is because, being unable to observe some actions, an agent's information can go out of synch with reality.

3.6.2 Multi-agent actions: the general case

To make our actions *private*, we need a mechanism in which we can represent actions that affect different agents in different ways. The *action models* of Baltag et al. (1999) allow us to do that. The key observation behind them is that, just as the agent can be uncertain about which one is the real world, she can also be uncertain about which particular event has taken place. In such situations, the uncertainty of the agent about the action can be represented with a model similar to that used for representing her uncertainty about the static situation.

More precisely, an action model consists of a collection of possible *events* connected by means of an accessibility relation, indicating the events each agent considers possible. Different from the worlds of a possible worlds model, and as their name indicate, each event is not understood as a possible state of affairs, but as an event that might have taken place. Instead of associating them with an atomic valuation, each event is associated with a *precondition*, indicating what that particular event requires to take place.

In order to make action models suitable for our purposes, we will extend them in essentially the manner of van Benthem et al. (2006) where, besides affecting the agent's uncertainty, action models can also affect the real world. In our case, besides affecting the agent's uncertainty, our action models will also be able to affect the agent's awareness.

Definition 3.9 (Multi-agent action model) With $\mathtt{P}$ the set of atomic propositions and $\mathtt{Ag}$ the finite set of agents, a *multi-agent action model* is a tuple $C = \langle E, T, \mathsf{Pre}, \mathsf{Pos}_{\mathsf{A}} \rangle$ where

- $\langle E, T, \mathsf{Pre} \rangle$ is an action model (Baltag et al. 1999) with E a finite non-empty set of *events*, $T : \mathtt{Ag} \to \wp(W \times W)$ a function returning an *accessibility relation* T_i for each agent $i \in \mathtt{Ag}$ and $\mathsf{Pre} : E \to \mathcal{L}$ the *precondition* function indicating the requirement for each event to be executed;

- $\mathsf{Pos}_{\mathsf{A}} : (\mathtt{Ag} \times E \times \wp(\mathcal{L})) \to \wp(\mathcal{L})$ is the *postcondition* function, assigning a new set of formulas in $\mathcal{L}$ to every tuple of an agent, event, and (old) set of formulas in $\mathcal{L}$.

A *pointed action model* (C, e) has a distinguished event e. ◄

Observe how, in our action model, each event e comes not only with a precondition $\mathsf{Pre}(e)$, a formula expressing what e needs to take place), but also with a *postcondition* $\mathsf{Pos}_{\mathsf{A}i}(e, X)$, the set of formulas agent i would be aware of if e takes place. This function $\mathsf{Pos}_{\mathsf{A}}$ is a generalization of the *substitution function* in van Benthem et al. (2006) for representing factual change.

We have defined the way we will represent our actions. It is now time to define how they will affect the agent's information.

Definition 3.10 (Product update) Let $M = \langle W, R, \mathsf{A}, V\rangle$ be a multi-agent awareness model and let $C = \langle E, T, \mathsf{Pre}, \mathsf{Pos}_{\mathsf{A}}\rangle$ be a multi-agent action model. The *product update* operation $\otimes$ yields the model $M \otimes C = \langle W', R', \mathsf{A}', V'\rangle$, given by

- $W' := \{(w, e) \mid (M, w) \Vdash \mathsf{Pre}(e)\}$
- $R'_i(w_1, e_1)(w_2, e_2)$ iff $R_i w_1 w_2$ and $T_i e_1 e_2$

and, for every $(w, e) \in W'$,

- $V'(w, e) := V(w)$
- $\mathsf{A}'_i(w, e) := \mathsf{Pos}_{\mathsf{A}i}(e, \mathsf{A}_i(w))$ ◀

The set of worlds of the model $M \otimes C$ is given by the restricted cartesian product of W and E: a pair (w, e) will be a world in the new model if and only if event e can be executed at world w. For each agent i, her uncertainty about the situation *after an action*, R'_i, is a combination of her uncertainty about the situation *before the action*, R_i, and her uncertainty *about the action*, T_i. The agent will not distinguish (w_2, e_2) from (w_1, e_1) if and only if she does not distinguish w_2 from w_1 and e_2 from e_1. Or, from the opposite perspective, the agent *will distinguish* (w_2, e_2) from (w_1, e_1) if and only if she already distinguishes w_2 from w_1, or e_2 from e_1. For the new atomic valuation, each world (w, e) inherits that of its static component w: an atom p holds at (w, e) if and only if p holds at w.

Now observe how the function $\mathsf{Pos}_{\mathsf{A}}$ works: for each agent i and each event e, $\mathsf{Pos}_{\mathsf{A}}$ takes agent i's awareness set at w in M, and returns her awareness set at (w, e) in $M \otimes C$. Note how $\mathsf{Pos}_{\mathsf{A}}$ does not have restrictions on the format of the definition; in fact, it can even return a new awareness set that is completely unrelated to the original one. The cases of interest in this chapter have a simple definition, and more uniform expressions will be explored in Chapter 5.

In order to express how product updates affect the agents' information, the *extended* multi-agent language $\mathcal{L}$ has extra modalities: if (C, e) is a pointed action model and φ is a formula in the extended multi-agent $\mathcal{L}$, then so is $\langle C, e\rangle \varphi$. The semantic interpretation of these new formulas is as follows:

Definition 3.11 (Semantic interpretation) Let (M, w) be a pointed multi-agent model and let (C, e) be a pointed action model with $C = \langle E, T, \mathsf{Pre}, \mathsf{Pos}_{\mathsf{A}}\rangle$.

$$(M, w) \Vdash \langle C, e\rangle \varphi \quad \text{iff} \quad (M, w) \Vdash \mathsf{Pre}(e) \text{ and } (M \otimes C, (w, e)) \Vdash \varphi$$ ◀

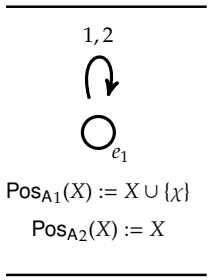
It is time to look at concrete cases illustrating the mechanism.

3.6.3 Public *consider* and *drop*

To begin with, our multi-agent setting generalizes the single agent case, since we can define action models for our earlier (now public) *consider* and *drop* operations.

Definition 3.12 (Public *consider* action) Let χ be a formula in multi-agent $\mathcal{L}$. The action of *agent j publicly considering χ* is given by the pointed action model $(\mathrm{Pub}^j_{+\chi}, e_1)$ with $\mathrm{Pub}^j_{+\chi} = \langle E, T, \mathsf{Pre}, \mathsf{Pos}_\mathsf{A}\rangle$ defined as

- $E := \{e_1\}$
- $\mathsf{Pre}(e_1) := \top$
- $T_i := \{(e_1, e_1)\}$ for every agent i
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}j}(e_1, X) := X \cup \{\chi\} \\ \mathsf{Pos}_{\mathsf{A}i}(e_1, X) := X \qquad \text{for } i \neq j \end{cases}$



The diagram on the left shows the action model $\mathrm{Pub}^1_{+\chi}$ in the 2-agent case (with the precondition omitted).

◀

Definition 3.13 (Public *drop* action) Let χ be a formula in multi-agent $\mathcal{L}$. The action of *agent j publicly dropping χ* is given by the pointed action model $(\mathrm{Pub}^j_{-\chi}, e_1)$, which differs from a *public considering* only in its postcondition function for j in e_1:

$$\mathsf{Pos}_{\mathsf{A}j}(e_1, X) := X \setminus \{\chi\}$$

◀

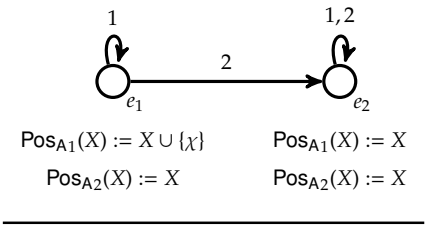
The public versions of the actions have just one event, and their accessibility relations T_i indicate that all involved agents recognize this. Moreover, the precondition in the unique world is simply $\top$. Then, the application of $(\mathrm{Pub}^j_{+\chi}, e_1)$ ($(\mathrm{Pub}^j_{-\chi}, e_1)$, respectively) on a multi-agent static model M yields a copy of M in which χ has been added to (removed from, respectively) the awareness set of agent j in all worlds.

3.6.4 Private *consider* and *drop*

But our mechanism can also define *private* actions. Here are simple versions of the earlier *consider* and *drop*. As usual, these encode what takes place, but also how different agents 'view' this.

Definition 3.14 (Private *consider* action) Let χ be a formula in multi-agent $\mathcal{L}$. The action of *agent j privately considering χ* is given by the pointed action model $(\mathrm{Pri}^j_{+\chi}, e_1)$ with $\mathrm{Pri}^j_{+\chi} = \langle E, T, \mathsf{Pre}, \mathsf{Pos}_\mathsf{A}\rangle$ defined as

- $E := \{e_1, e_2\}$
- $T_i := \begin{cases} \{(e_1,e_1),(e_2,e_2)\} & \text{if } i = j \\ \{(e_1,e_2),(e_2,e_2)\} & \text{otherwise} \end{cases}$
- $\mathsf{Pre}(e_1) = \mathsf{Pre}(e_2) := \top$
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}j}(e_1, X) := X \cup \{\chi\}, & \mathsf{Pos}_{\mathsf{A}j}(e_2, X) := X \\ \mathsf{Pos}_{\mathsf{A}i}(e_1, X) := X, & \mathsf{Pos}_{\mathsf{A}i}(e_2, X) := X \quad \text{for } i \neq j \end{cases}$



The diagram on the left shows the model $\mathrm{Pri}^1_{+\chi}$ for 2 agents (preconditions again omitted).

◀

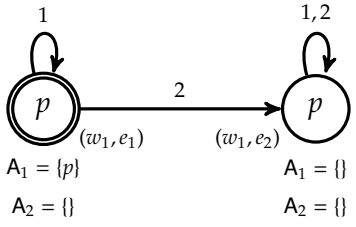
Definition 3.15 (Private *drop* action) Let χ be a formula in multi-agent $\mathcal{L}$. The action of *agent j privately dropping χ* is given by the pointed action model $(\mathrm{Pri}^j_{-\chi}, e_1)$, which differs from a *private considering* only in its postcondition function for j in e_1:

$$\mathsf{Pos}_{\mathsf{A}j}(e_1, X) := X \setminus \{\chi\}$$

◀

The difference between the public and the private version of the actions is that the private actions involve two events: one in which χ is added to (removed from) agent j's awareness set (the event e_1), and another in which there is no change (the event e_2). Moreover, the accessibility relations T_i indicate that, while j recognizes which event is the real one (our e_1), the other agents do not consider that event possible, sticking to the 'no change' option. Then, the application of $(\mathrm{Pri}^j_{+\chi}, e_1)$ ($(\mathrm{Pri}^j_{-\chi}, e_1)$, respectively) on a multi-agent static model M yields a model containing two copies of M: one, recognized as the real one only by j, in which j's awareness set has changed, and another, viewed by the other agents as the only possibility, in which nothing has happened.

Example 3.3 Recall the model M from Example 3.2. After agent 1 considers p privately (i.e., after applying $(\mathrm{Pri}^1_{+p}, e_1)$), we get a better version of the initial situation that started the thread of this section:



In the evaluation point, (w_1, e_1), agent 1 is aware of p ($A_1\, p$), just like she does after publicly considering p. But this time, agent 2's implicit information does not change: she is still implicitly informed that agent 1 is not aware of p ($\Box_2 \neg A_1\, p$).

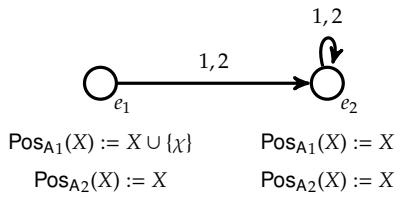
◀

3.6.5 Unconscious versions

The flexibility of the postcondition mechanism is great. We can represent many further scenarios, even *unconscious* actions, hidden from all agents, including the one that 'performs' it! We just give an illustration:

Definition 3.16 (Unconscious *drop* action) Let χ be a formula in the multi-agent $\mathcal{L}$. The action of *agent j unconsciously dropping χ* is given by the pointed action model $(\mathrm{Unc}^j_{-\chi}, e_1)$, differing from its *private* counterpart only in the definition of the accessibility relation:

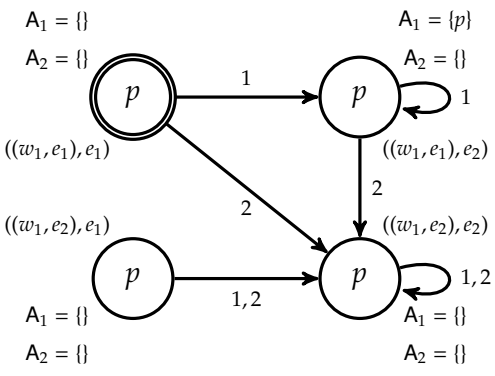
- $T_i := \{(e_1, e_2), (e_2, e_2)\}$ for all agents i



The diagram on the left depicts $\mathrm{Unc}^1_{-\chi}$ in a 2-agent scenario.

◀

Example 3.4 Consider the model $(M \otimes \mathrm{Pri}^1_{+p}, (p, e_1))$ of Example 3.3. If agent 1 unconsciously drops p, we get the following updated model:



In $\big((w_1, e_1), e_1\big)$, agent 1 is not aware of p ($\neg A_1\, p$), but she is implicitly informed that she is aware of it ($\Box_1 A_1\, p$).

The updated model contains two copies of the original one. Here, agent 1 considers only the rightmost world of the upper copy, and agent 2 only the rightmost world of the lower one. ◄

Much more can be said about this scenario, and we feel that we have a promising take here on unconscious actions such as *forgetting*. But our purpose here was just to demonstrate the flexibility of the framework.

3.6.6 Completeness of the multi-agent system

In principle, there is a complete axiom system for our product update mechanism for awareness, and it looks like our earlier single-agent logic, with indices attached. Its principles for atomic formulas, boolean operations, and implicit knowledge are the usual ones from Dynamic Epistemic Logic *DEL*. As an illustration, we have the valid equivalence

$$\vdash \langle C, e\rangle \Diamond_i \varphi \quad \leftrightarrow \quad \mathsf{Pre}(e) \rightarrow \bigvee_{T_i e f} \Diamond_i \langle C, f\rangle \varphi$$

But to formulate a precise result, the crucial issue is stating the right reduction axiom for awareness given the *postconditions*. Consider the earlier axioms that we gave for our two basic syntactic operations of *consider* and *drop*. These described the postconditions (the effect of the operations on the A-sets) inside the language, exploiting the simple format of their effects. For instance, we have φ in our A-set after an act $+\chi$ if we had φ before, or φ is actually the just added formula χ. This case distinction in the reduction axiom reflects directly the simple disjunctive *definition of the postcondition* for the action $\chi!$: $\mathsf{A}'(w) := \mathsf{A}(w) \cup \{\chi\}$.

The same is true in our more general setting with action models: *simple definitions of postconditions in our action models will allow us to provide matching reduction axioms*. Consider, as an illustration, the system in van Benthem et al. (2006); it allows factual change by modifying the set of worlds in which each atomic proposition is true. But this new set is not an arbitrary one; it is defined syntactically by means of a formula of the language. This suggest that, in order to get a proper completeness theorem, we should look for some uniform syntactic expressions from which reduction axioms can be derived. We will deal with this issue in Chapter 5.

Though we have merely made some proposals, the defined actions show the power of a simple syntactic extension of the well-known *DEL* action models and its product update.

3.7 Remarks

In this chapter we have noticed that, besides the lack of acknowledgement of a given formula as true, there is another reason for which an agent may have just implicit information about it: she may not be *aware of* it. This notion has been the main protagonist of the present chapter. Based in the *Awareness Logic* of Fagin and Halpern (1988), we have discussed its role in the definition of the notion of *explicit information*, presenting several possibilities that make also use of the *implicit information* notion. Table 3.2 summarizes the notions of information discussed in this chapter and the definition we have worked with.

Notion	**Definition**	**Relevant model requirements**
Awareness of	$\mathrm{A}\,\varphi$	—
Implicit information	$\Box\,\varphi$	—
Explicit information	$\Box\,(\varphi \wedge \mathrm{A}\,\varphi)$	—

Table 3.2: Static notions of information.

Here it is important to emphasize again the difference in the interpretation of the A-sets in this and the previous chapter. In Chapter 2 we discuss an agent that does not need to be omniscient because she does not need to recognize every true formula as such; in that case the A-sets are interpreted directly as the agent's explicit information. In the present chapter we have discussed an agent that does not need to be omniscient because she does not need to have full attention; in this case the A-sets are interpreted as those formulas the agent *is aware of*, but this does not imply any attitude, positive or negative, about them. Note that awareness by itself is not enough to provide the agent explicit information (we are aware of many possibilities, but that definitely does not imply that we assume that all of them are true); this is another reason why we have changed our definition of explicit information from the $\mathrm{A}\,\varphi$ of the previous chapter, to the current $\Box\,(\varphi \wedge \mathrm{A}\,\varphi)$.

On the dynamic side, just like in the previous chapter, the introduction of finer notions of information has allowed us to describe acts that, though important for real human agents, have been neglected in the classical *DEL* literature due to the strong idealization of the represented agents. In this case, the highlighted acts are those that change the information the agent is aware of (therefore modifying her explicit information), and we have provided a formal representation for two of them: *consider*, an act of awareness raising, and *drop*, an act of awareness reduction. For the third static notion discussed in this chapter, *implicit information*, we have not only recalled an action that affects it, *implicit observation* (the *PAL* public announcement), but also shown how an

explicit version, *explicit observation*, can be defined with the help of the consider action. Moreover, we have made the jump from a single-agent case to a multi-agent scenario, and we have observed that, in public, privacy becomes important. Accordingly, we have provided proper multi-agent representation for private and even unconscious versions of the awareness-changing acts already proposed for the single-agent case. These defined actions sketch the power of a syntactic extension of well-known action models and product update, and further examples of their application will be presented in Chapter 5. Table 3.3 summarizes the actions that have been defined in this chapter.

Action	**Description**
Consider (in its public and private versions).	The agent becomes aware of a formula.
Drop (in its public and private versions).	The agent loses awareness of a formula.
Implicit observation.	Changes the agent's implicit information.
Explicit observation.	Changes the agent's implicit and explicit information.

Table 3.3: Actions and their effects.

The notion of awareness (and its dual, unawareness) has been an interesting research topic not only in Logic (Ågotnes and Alechina 2007; van Ditmarsch et al. 2009) but also in Computer Science Halpern (2001); Halpern and Rêgo (2005, 2009) and particularly in Economics (Modica and Rustichini 1994; Dekel et al. 1998; Modica and Rustichini 1999; Heifetz et al. 2003; Board and Chung 2006; Sillari 2006; Samet 2007). Dynamics of the notion have been recently explored also in van Ditmarsch and French (2009) and Hill (2010). Our particular approach shows how a significant informational dynamics can take place over existing awareness models, generalizing acts of observation and awareness change. It also shows how this leads to useful technical systems, and we have provided results about them in the spirit of Dynamic Epistemic Logic. Thus, we have shown that the 'reductionist approach' to explicit knowledge in terms of implicit semantic knowledge and syntactic awareness is feasible and interesting in its own right.

During the present chapter some new issues have arisen. Our multi-agent setting can describe many more agent activities than what we have shown, and we have only scratched the surface. Also, many technical issues remain open, like the issue of schematic validities and action algebra. In the case of the first, there is already one interesting result: the set of schematic validities in *Public Announcement Logic* is decidable (Holliday et al. 2010).

But beyond this, there is a more important question: how the two notions of explicit information that we have worked with so far are related? In Chapter 2 we worked under the intuition that the only reason why implicit information may not be explicit is because the agent could fail to recognize true formulas as such, just like we may fail to recognize that the conclusion of a theorem is true. But it is now clear that acknowledging a formula as true could not be enough because we could still need the adequate 'attention' to the subject.

In the present chapter we have reduced explicit information to implicit information plus awareness, that is, the only reason why implicit information may not be explicit is because the agent could not have full attention, just like we may fail to recognize that our lost keys are in the kitchen because we do not even consider that possibility. But it is also clear that full attention could not be enough to make explicit our implicit information because we could still need to recognize the information as true. Think of a conclusion that I am pondering, and that in fact follows from some premises whose truth I explicitly have. I could still fail to see explicitly the conclusion. In fact, this shows how our acts of awareness raising are not acts of inference, since under this chapter's definition, merely becoming aware of φ was enough to upgrade information from implicit to explicit.

A more satisfying notion of explicit information is one that combines the two ideas: the agent needs to be aware of the subject, but also recognize it as true. This idea, and the focus on the particular case of true information (knowledge) will be the topic of our next chapter.

Chapter 4

Awareness, implicit and explicit knowledge

The two previous chapters studied the notion of explicit information by looking at two different requirements. In Chapter 2, we asked for explicit information to be implicit information that the agent has acknowledged as true, highlighting the fact that a 'real' agent, even with full attention about the current possibilities, may fail to recognize that some facts are indeed the case. In Chapter 3, we asked for explicit information to be implicit information that the agent is *aware of*, highlighting the fact that a 'real' agent, even with full reasoning abilities, may not pay full attention to all relevant possibilities.

As we mentioned in the closing remarks of Chapter 3, we can obtain a more satisfying notion of explicit information by putting these two ingredients together: explicit information needs attention and acknowledgement of formulas as true. This gives us a broader range of attitudes and allows us to represent situations that are not possible within the frameworks of the two previous chapters, like a situation in which, though the agent has accepted something as true, she is not paying attention to (aware of) it right now and therefore she does not have it explicitly, or situations in which, though aware of and implicitly informed about a fact, the agent still fails to recognize it as true.

This chapter starts by looking at a definition of explicit information that involves the two mentioned requirements, putting particular attention on the case of *true* information, that is, *knowledge*. Then we review how some of the already defined actions, namely changes in awareness, inference and explicit observation (announcement in this case), work in this richer setting.

4.1 Twelve Angry Men

Consider the following quote, taken from the script of Sydney Lumet's 1957 movie "12 Angry Men".

> *"You've listened to the testimony [. . .] It's now your duty to sit down and try and separate the facts from the fancy. One man is dead. Another man's life is at stake. If there's a reasonable doubt [. . .] as to the guilt of the accused [. . .], then you must bring me a verdict of not guilty. If there's no reasonable doubt, then you must [. . .] find the accused guilty. However you decide, your verdict must be unanimous."*

The quote illustrates a very common collective decision-making situation: a group of agents should put their particular information together in order to establish whether a given state-of-affairs holds or not (Kornhauser and Sager 1986). But before the very act of voting, these scenarios typically include a deliberation phase, and it is precisely in this phase in which new issues are introduced and explicit information is exchanged, allowing the agents to perform further reasoning steps and therefore reach a better 'merging' of their individual information.

Take, as a more concrete example, the following excerpt from the Jury's deliberation in the mentioned movie.

Example 4.1 (12 Angry Men)

A: *Now, why were you rubbing your nose like that?*

H: *If it's any of your business, I was rubbing it because it bothers me a little.*

A: *Your eyeglasses made those two deep impressions on the sides of your nose.*

A: *I hadn't noticed that before.*

A: *The woman who testified that she saw the killing had those same marks on the sides of her nose.*

. . .

G: *Hey, listen. Listen, he's right. I saw them too. I was the closest one to her. She had these things on the side of her nose.*

. . .

D: *What point are you makin'?*

D: *She had [. . .] marks on her nose. What does that mean?*

A: *Could those marks be made by anything other than eyeglasses?*

. . .

D: *[. . .] How do you know what kind of glasses she wore? Maybe they were sunglasses! Maybe she was far-sighted! What do you know about it?*

C: *I only know the woman's eyesight is in question now.*

. . .

C: *Don't you think the woman may have made a mistake?*

B: *Not guilty.*

◀

The excerpt shows the dynamics of the two discussed ingredients: awareness and acknowledgement of facts as true. Juror A supports the idea that the defendant cannot be proven guilty beyond reasonable doubt, and juror H's action of rubbing his nose makes A aware of an issue that has not been considered before: marks on the nose. When he considers the issue, he remembers that the witness of the killing had such marks, and he announces it. Now everyone knows (in particular, G remembers) that the woman had marks on the side on her nose. Then, A draws an inference and announces what he has concluded: the marks are due to the use of glasses. After this announcement takes place, it is now C who performs an inference, concluding that the woman's eyesight can be questioned. Finally, B makes the last reasoning step, announcing then to everybody that the defendant is not guilty beyond any reasonable doubt.

During the deliberation we can see the interplay of at least three different notions of information: what each juror is aware of, and his implicit and explicit information. Moreover, we can also see two of the main informational actions we have studied, inference and changes in awareness, as well as a small variant of a third, broadcasted explicit observations, that is, announcements.

How can we represent formally such deliberative situations? We already have frameworks for dealing with agents that may fail to recognize formulas as true (Chapter 2) and with agents that may lack full attention (Chapter 3), so the natural idea is combine them. But before going into discussions of the dynamics, we must settle down what will be the static notions of information we will work with, and what is the relation between them.

4.2 Awareness, implicit and explicit information

The deliberation shows at least three different notions of information. The strongest of them, that of *explicit information*, is what is directly available to the agent without any further reasoning step. In our running example, all members of the Jury are explicitly informed (in this particular case, they *explicitly know*) that a killing has taken place, that a boy is being accused of the killing, and that a woman has testified affirming that she saw the killing.

There is also information that is not directly available to the agents; information that follows from what they explicitly know but should be 'put in the light'. In the example, at some stage agent D has recognized (that is, knows explicitly) that the witness had marks on her nose. From that information it follows that she wears glasses, but D is just *implicitly* informed about it; he needs to perform an inference step to reach that conclusion.

But even if at that point D does not have explicit information about the witness using glasses, he considers it as a possibility, just like he considers possible for the accused to be innocent or guilty. Such possibilities are part of the current discussion or, more syntactically, they are part of the agent's

current language. On the other hand, before H rubs his nose, the possibility of the woman having or not marks in her nose is not considered by the agents: they are not *aware of* that possibility. Again, just like in the previous chapter, being *aware of* a possibility just means that the agent *entertains* it, and does not imply by itself any attitude positive or negative towards the possibility. But also note that *not* being currently aware of a possibility does not imply that an agent does not have information about it. In our example, while H was completely uninformed about the witness having marks on her nose or not, A knew that the witness had such marks, but he just disregarded that information.

Here is a more mathematical example relating the three mentioned notions. Consider an agent trying to prove that if $p \to q$ and p are the case, then so is q. She is *explicitly* informed that $p \to q$ and p are the case, but she is informed about q just *implicitly*: she needs to perform an inference step in order to make it explicit. While trying to find the proof, the agent is *aware of* p and q, but not of r, s and other atomic propositions. Again, this does not say that she has or does not have relevant information about r, s and so; it just says that these atoms are not part of the information the agent entertains right now.

Relation between the notions The relation we assume between *implicit* and *explicit* information is as before: explicit information is implicit information that has been 'put in the light' by some reasoning mechanism. Therefore, explicit information is always part of the implicit one.

The relation between implicit information and information we are aware of can be seen from two different perspectives. We could assume that the agent's implicit information is everything that the agent can get to know, including what she would get if she became aware of every possibility. Then, the information the agent is *aware of* would be part of her *implicit* information. From our discussion before, it can be seen that we will adopt another perspective: the information the agent is aware of actually defines her *language*, and neither implicit nor explicit information can go beyond it. Therefore, implicit information is part of the information the agent is aware of. Figure 4.1 shows the hierarchy that will be used in this chapter.

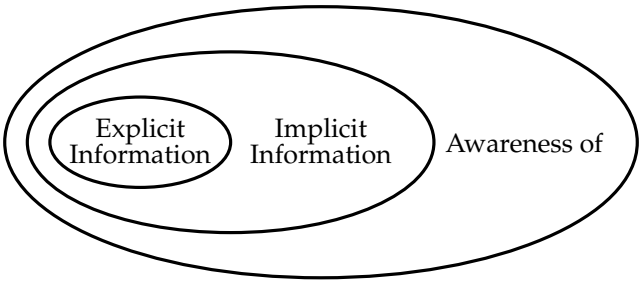


Figure 4.1: Awareness of, implicit and explicit information.

It is important to make the following remark. The notion of *awareness of* we used in the previous chapter was understood as a set of formulas. And though Fagin and Halpern (1988) study several closure properties of it, in Chapter 3 we did not assumed any of them. In this chapter we will work with a more restricted version of awareness: the one that is generated from a set of atomic propositions and therefore defines the agent's language. The intuition is that if an agent is aware of a formula φ, then she should be aware of all its subformulas and, in particular, she should be aware of all the atomic propositions that appear φ. On the other hand, if the agent is aware of a given set of atomic propositions, then she should be aware of any formula that can be built from it.[1] Accordingly, the awareness of an agent will not be defined by a set of arbitrary formulas, but by a set of *atomic propositions*.

4.3 The static framework

We start by defining the formal language that allows us to describe situations like Example 4.1, together with its semantic model and semantic interpretation.

4.3.1 Basic language, models and interpretation

Definition 4.1 (Language $\mathcal{L}$) Let $\mathtt{P}$ be a set of atomic propositions and let $\mathtt{Ag}$ be a set of agents. Formulas φ, ψ and rules ρ of the language $\mathcal{L}$ are given by

$$\varphi ::= p \mid {}^{[i]}p \mid \mathrm{A}_i\,\varphi \mid \mathrm{R}_i\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \Box_i\varphi$$
$$\rho ::= (\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi)$$

where $p \in \mathtt{P}$ and $i \in \mathtt{Ag}$. We denote by $\mathcal{L}_f$ the set of formulas of $\mathcal{L}$, and by $\mathcal{L}_r$ its set of rules. Other boolean connectives ($\wedge, \rightarrow, \leftrightarrow$) as well as the existential modalities $\Diamond_i$ are defined as usual ($\Diamond_i\varphi := \neg\Box_i\neg\varphi$, for the latter). ◄

The language $\mathcal{L}$ extends that of *EL* with three new basic components: ${}^{[i]}p$, $\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$. Formulas of the form ${}^{[i]}p$ indicate that *agent i has proposition p available (at her disposal)* for expressing her information, and will be used to define the notion of *awareness of*. Formulas of the form $\mathrm{A}_i\,\varphi$ (*access* formulas) and $\mathrm{R}_i\,\rho$ (*rule-access* formulas) indicate that *agent i can access formula* φ and *rule* ρ, respectively. While the first will be used to define the agent's explicit information, the second will be used to express the *processes* the agent can use to extend this explicit information. These processes, in our case syntactic rules, deserve a brief discussion once again.

[1] These intuitions reflect the assumption that the agent can process negation, disjunction and other boolean connectives without any problem. It also assumes that, in principle, the agent does not have problem with reasoning about herself and other agents (but see the discussion on awareness of agents).

The rules Let us go back to our example for a moment, and consider how the three notions of information change. The possibilities an agent considers, the *awareness of* notion, change as a consequence of the appearance of a possibility not currently considered, and the *implicit information* notion changes as a consequence of announcements, that is, communication. But changes in explicit information do not need to be the result of external influences; they can also be the result of the agent's own reasoning steps. And in order to perform such reasoning steps the agent needs certain extra 'procedural' information, just like the Pythagoras theorem is needed to get the length of the hypothenuse from the length of the legs, or, in a simpler setting, just like *modus ponens* is needed to get q from p and $p \to q$. This is precisely the role of our syntactic rules, the most natural way of representing this 'procedural' information in our logical setting. Rules are precisely what allow the agent to infer further consequences of her explicit information.

Let us recall the following rule-related definitions.

Definition 4.2 (Premises, conclusion and translation) Let ρ be a rule in $\mathcal{L}_r$ of the form $(\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi)$. We define

$$\begin{array}{rl} \mathsf{pm}(\rho) := \{\psi_1, \ldots, \psi_{n_\rho}\} & \text{the } \textit{set of premises} \text{ of } \rho \\ \mathsf{cn}(\rho) := \varphi & \text{the } \textit{conclusion} \text{ of } \rho \end{array}$$

Moreover, we define a rule's *translation*, $\mathsf{tr}(\rho) \in \mathcal{L}_f$, as an implication whose antecedent is the (finite) conjunction of the rule's premises and whose consequent is the rule's conclusion:

$$\mathsf{tr}(\rho) := \Big(\bigwedge_{\psi \in \mathsf{pm}(\rho)} \psi \Big) \to \mathsf{cn}(\rho)$$

◀

Besides the already discussed formulas $\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$, we also have now formulas expressing availability of atoms: ${}^{[i]}p$. Let us discuss them.

Availability of formulas Formulas of the form ${}^{[i]}p$ allow us to express local availability of atomic propositions, and they will allow us to define what an agent is *aware of*. The notion can be extended to express local availability of formulas of the whole language in the following way.

Definition 4.3 Let i, j be agents in $\mathtt{Ag}$. Define

$$\begin{array}{rclcrcl} {}^{[i]}({}^{[j]}\varphi) & := & {}^{[i]}\varphi & \qquad & {}^{[i]}(\neg\varphi) & := & {}^{[i]}\varphi \\ {}^{[i]}(\mathrm{A}_j\,\varphi) & := & {}^{[i]}\varphi & & {}^{[i]}(\varphi \vee \psi) & := & {}^{[i]}\varphi \wedge {}^{[i]}\psi \\ {}^{[i]}(\mathrm{R}_j\,\rho) & := & {}^{[i]}\rho & & {}^{[i]}(\Box_j\varphi) & := & {}^{[i]}\varphi \end{array}$$

and

$${}^{[i]}\rho \;:=\; {}^{[i]}\mathsf{tr}(\rho)$$

◀

Intuitively, formulas of the form $[i]\varphi$ express that φ is available to agent i, and this happens exactly when all the atoms in φ are available to her. For example, $[i](\neg p)$ is defined as $[i]p$, that is, the formula $\neg p$ is available to agent i whenever p is available to her. On the other hand, $[i](p \vee q)$ is given by $[i]p \wedge [i]q$, that is, $p \vee q$ is available to agent i whenever *both* p and q are available to her.

Note how the definition of availability for agent i in the case of formulas involving an agent j ($[j]\varphi$, $\mathrm{A}_j\,\varphi$, $\mathrm{R}_j\,\rho$ and $\Box_j\varphi$) simply discard any reference to j. With this definition, we are implicitly assuming that all agents are 'available' to each other, that is, all agents can talk about any other agent. Some other approaches, like van Ditmarsch and French (2009), consider also the possibility of agents that are not necessarily aware of all other agents. We will not pursue such generalization here, but we emphasize that this idea has interesting consequences, as we will mention once we provide our definitions for the *awareness of*, *implicit* and *explicit information* notions in Section 4.3.2.

Having defined the language $\mathcal{L}$, we now define the semantic model in which the formulas will be interpreted.

Definition 4.4 (Semantic model) Let $\mathtt{P}$ be the set of atomic propositions and $\mathtt{Ag}$ the set of agents. A *semantic model* for the language $\mathcal{L}$ is a tuple $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}_i, \mathsf{R}_i\rangle$ where:

- $\langle W, R_i, V\rangle$ is a standard multi-agent possible worlds model with W the non-empty *set of worlds*, $R_i \subseteq W \times W$ an *accessibility relation* for each agent i and $V : W \to \wp(\mathtt{P})$ the *atomic valuation*;
- $\mathsf{PA}_i : W \to \wp(\mathtt{P})$ is the *propositional availability function*, indicating the set of atomic propositions agent i has at her disposal at each possible world;
- $\mathsf{A}_i : W \to \wp(\mathcal{L}_f)$ is the *access set function*, indicating the set of formulas agent i can access (i.e., has acknowledge as true) at each possible world;
- $\mathsf{R}_i : W \to \wp(\mathcal{L}_r)$ is the *rule set function*, indicating the set of rules agent i can access (i.e., has acknowledged as truth-preserving) at each possible world.

The pair (M, w) with M a semantic model and w a world in it is called a *pointed semantic model*. We denote by $\mathbf{M}$ the class of all semantic models. ◀

Our semantic model extends possible worlds models with three functions, PA_i, A_i and R_i, that allow us to give semantic interpretation to the new formulas.

Definition 4.5 (Semantic interpretation) Let the pair (M, w) be a pointed semantic model with $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}_i, \mathsf{R}_i\rangle$. The *satisfaction* relation $\Vdash$ between formulas of $\mathcal{L}$ and (M, w) is given by

$$
\begin{array}{lll}
(M,w) \Vdash p & \text{iff} & p \in V(w) \\
(M,w) \Vdash {}^{[i]}p & \text{iff} & p \in \mathsf{PA}_i(w) \\
(M,w) \Vdash \mathrm{A}_i\,\varphi & \text{iff} & \varphi \in \mathsf{A}_i(w) \\
(M,w) \Vdash \mathrm{R}_i\,\rho & \text{iff} & \rho \in \mathsf{R}_i(w) \\
(M,w) \Vdash \neg\varphi & \text{iff} & \text{it is not the case that } (M,w) \Vdash \varphi \\
(M,w) \Vdash \varphi \vee \psi & \text{iff} & (M,w) \Vdash \varphi \text{ or } (M,w) \Vdash \psi \\
(M,w) \Vdash \Box_i\varphi & \text{iff} & \text{for all } u \in W,\ R_iwu \text{ implies } (M,u) \Vdash \varphi
\end{array}
$$ ◀

The multi-agent version of the basic epistemic axiom system is sound and complete for this framework.

Theorem 4.1 (Sound and complete axiom system for $\mathcal{L}$ w.r.t. M) *The axiom system of Table 4.1 is sound and strongly complete for formulas of $\mathcal{L}$ w.r.t.* **M***-models.*

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$
K	$\vdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$	*Nec*	If $\vdash \varphi$, then $\vdash \Box_i\varphi$
Dual	$\vdash \Diamond_i\varphi \leftrightarrow \neg\Box_i\neg\varphi$		

Table 4.1: Axiom system for $\mathcal{L}$ w.r.t. **M**.

Proof. (Sketch of proof) The proof is similar to that of Theorem 2.1. The axioms are valid and the rules preserve validity, so we get soundness. Completeness is proved by building the standard modal canonical model with the adequate definitions for the propositional availability, access set and rule set functions:

$$
\mathsf{PA}_i(w) := \{p \in \mathtt{P} \mid {}^{[i]}p \in w\} \qquad \mathsf{R}_i(w) := \{\rho \in \mathcal{L}_r \mid \mathrm{R}_i\,\rho \in w\}
$$
$$
\mathsf{A}_i(w) := \{\varphi \in \mathcal{L}_f \mid \mathrm{A}_i\,\varphi \in w\}
$$

With these definitions, it is easy to show that the new formulas also satisfy the *Truth Lemma*, that is,

$$
(M,w) \Vdash {}^{[i]}p \quad \text{iff} \quad {}^{[i]}p \in w \qquad (M,w) \Vdash \mathrm{R}_i\,\rho \quad \text{iff} \quad \mathrm{R}_i\,\rho \in w
$$
$$
(M,w) \Vdash \mathrm{A}_i\,\varphi \quad \text{iff} \quad \mathrm{A}_i\,\varphi \in w
$$

This gives us completeness. ■

Once again, note how there are no axioms for formulas of the form ${}^{[i]}p$, $\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$. Such formulas can be seen as particular atomic propositions that correspond to the particular *valuation functions* PA_i, A_i and R_i. Since these functions do not have any special property and there is no restriction in the way they interact with each other, we do not need special axioms for them (but see Subsection 4.3.2 for some interaction properties).

Nevertheless, Definition 4.3 gives us validities expressing the behaviour of ${}^{[i]}\varphi$. The formulas of Table 4.2 are valid in **M**-models.

$[i](\neg\varphi) \leftrightarrow [i]\varphi$	$[i]([j]\varphi) \leftrightarrow [i]\varphi$
$[i](\varphi \vee \psi) \leftrightarrow [i]\varphi \wedge [i]\psi$	$[i](\mathrm{A}_j\,\varphi) \leftrightarrow [i]\varphi$
$[i](\Box_j\varphi) \leftrightarrow [i]\varphi$	$[i](\mathrm{R}_j\,\rho) \leftrightarrow [i]\mathsf{tr}(\rho)$

Table 4.2: Validities derived from Definition 4.3.

4.3.2 The relevant notions and basic properties

With the language, semantic model and semantic interpretation defined, it is now time to formalize the notions informally introduced in Section 4.2.

Definition 4.6 The notions of *awareness*, *implicit information* and *explicit information* are defined as in Table 4.3. ◄

Agent i is *aware of* formula φ	$\mathrm{Aw}_i\,\varphi := \Box_i\,[i]\varphi$
Agent i is *aware of* rule ρ	$\mathrm{Aw}_i\,\rho := \Box_i\,[i]\mathsf{tr}(\rho)$
Agent i is *implicitly informed* about formula φ	$\mathrm{Im}_i\,\varphi := \Box_i\big([i]\varphi \wedge \varphi\big)$
Agent i is *implicitly informed* about rule ρ	$\mathrm{Im}_i\,\rho := \Box_i\big([i]\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho)\big)$
Agent i is *explicitly informed* about formula φ	$\mathrm{Ex}_i\,\varphi := \Box_i\big([i]\varphi \wedge \varphi \wedge \mathrm{A}_i\,\varphi\big)$
Agent i is *explicitly informed* about rule ρ	$\mathrm{Ex}_i\,\rho := \Box_i\big([i]\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho) \wedge \mathrm{R}_i\,\rho\big)$

Table 4.3: Formal definitions of awareness, implicit and explicit information.

First, let us review the new definition for the notion of awareness. In Chapter 3, this notion is given directly by the so-called awareness set: an agent i is aware of φ at a world w if and only if $\varphi \in \mathsf{A}_i(w)$. Now the notion is not defined from a set of formulas, but from a set of atomic propositions. But not only that. This new notion of awareness needs more than just availability at the evaluation point: the agent is aware of φ at a world w if and only if $[i]\varphi$ holds *in all the worlds she considers possible*. By Propositions 4.1 and 4.2 below, $[i]\varphi$ holds if and only if $[i]p$ holds for every atom p of φ, so in fact the agent is aware of φ if and only if she has available every atom of φ in all the worlds she considers possible. Our notion of awareness corresponds now to a language based on those atomic propositions that appear in the PA-set of all worlds reachable through the agent's accessibility relation. We emphasize that this form of syntactic representation of awareness based on atomic propositions is not given by a set of formulas with a special closure property (like in Fagin and Halpern (1988)), but rather by a property on atomic propositions (the PA-sets) lifted to a property on formulas via a recursive definition (Definition 4.3).

The second notion, implicit information, is not independent from awareness anymore: even if φ holds in all the worlds agent i considers possible, she will not have implicit information about it unless she is aware of it.

Finally the strongest notion, that of explicit information. It asks not only for awareness and implicit information, but also for access in all the R_i-accessible worlds. In other words, in order to have explicit information about a given φ, the agent not only should be aware of and have implicit information about it: she should also acknowledge it as true in all the worlds she considers possible. Here it is also important to notice how this definition follows the spirit of the definition of explicit information of Chapter 3 in which all the needed ingredients fall under the scope of the modal operator $\Box_i$. Then, in order to deal with true explicit information, that is, in order to deal with *explicit knowledge*, we just need to ask for equivalence accessibility relations (see Subsection 4.3.3) and no extra special properties (like *truth* of Chapter 2) are required. Hence, no restrictions for formulas in the A_i-sets are needed.

The rest of this subsection will be devoted to the study of basic properties of these notions and the way they interact with each other. We will focus on *awareness of*, *implicit* and *explicit information* about formulas, but the cases for rules can be obtained in a similar way.

The *awareness of* notion The notion of *awareness of* for agent i is now defined in terms of the formulas the agent has available in all the worlds she can access. We say that agent i is aware of φ, $\mathrm{Aw}_i\,\varphi$, if and only if she has φ at her disposal in all the worlds she considers possible, $\Box_i\,{}^{[i]}\varphi$. As we have mentioned, this notion of awareness defines the language of the agent. First, if the agent is aware of a formula φ, then she is aware of all the atoms in the formula. But not only that; if the agent is aware of a set of atomic propositions, then she is aware of every formula built from such atoms. These statements are made formal and proved in Proposition 4.1 and Proposition 4.2 below.

In order to prove Proposition 4.1, we first need the following lemma.

Lemma 4.1 *Let the pair (M, w) be a pointed semantic model and i an agent. Let φ be a formula in $\mathcal{L}$, and denote by $\mathsf{atm}(\varphi)$ the set of atomic propositions occurring in φ.*

If i has φ at her disposal, that is, if $(M, w) \Vdash {}^{[i]}\varphi$, then she has at her disposal all atoms in it, that is, $(M, w) \Vdash {}^{[i]}p$ for every $p \in \mathsf{atm}(\varphi)$. In other words, the formula

$$ {}^{[i]}\varphi \rightarrow {}^{[i]}p $$

is valid for every $p \in \mathsf{atm}(\varphi)$.

Proof. We prove the following equivalent (and more semantic) statement:

$$ (M, w) \Vdash {}^{[i]}\varphi \quad \text{implies} \quad \mathsf{atm}(\varphi) \subseteq \mathsf{PA}_i(w) $$

The two statements are indeed equivalent, given the semantic interpretation of formulas of the form ${}^{[i]}p$ (Definition 4.5).

The proof is by induction on φ. The base case is immediate, and the inductive ones follow from the inductive hypothesis and the validities of Table 4.2 derived from Definition 4.3. Details can be found in Appendix A.8. ■

Proposition 4.1 *Let the pair (M, w) be a pointed semantic model and i an agent. Let φ be a formula in $\mathcal{L}$.*

If i is aware of φ, that is, if $(M, w) \Vdash \mathrm{Aw}_i\, \varphi$, then she is aware of all its atoms, that is, $(M, w) \Vdash \mathrm{Aw}_i\, p$ for every $p \in \mathsf{atm}(\varphi)$. In other words, the formula

$$\mathrm{Aw}_i\, \varphi \to \mathrm{Aw}_i\, p$$

is valid for every $p \in \mathsf{atm}(\varphi)$.

Proof. Suppose $(M, w) \Vdash \mathrm{Aw}_i\, \varphi$. Then, $(M, w) \Vdash \Box_i\, {}^{[i]}\varphi$, that is, $(M, u) \Vdash {}^{[i]}\varphi$ for every u such that $R_i w u$. Pick any such u; by Lemma 4.1, $(M, u) \Vdash {}^{[i]}p$ for every $p \in \mathsf{atm}(\varphi)$. Hence, $(M, w) \Vdash \Box_i\, {}^{[i]}p$, that is $(M, w) \Vdash \mathrm{Aw}_i\, p$, for every $p \in \mathsf{atm}(\varphi)$.■

In order to prove Proposition 4.2, we first need the following lemma.

Lemma 4.2 *Let the pair (M, w) be a pointed semantic model and i an agent. Let $\{p_1, \ldots, p_n\} \subseteq \mathsf{P}$ be a set of atomic propositions.*

If i has all atoms in $\{p_1, \ldots, p_n\}$ at her disposal, that is, if $(M, w) \Vdash {}^{[i]}p_k$ for every $k \in \{1, \ldots, n\}$, then she has at her disposal any formula built from such atoms, that is, $(M, w) \Vdash {}^{[i]}\varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. In other words, the formula

$$\Big(\bigwedge_{k \in \{1,\ldots,n\}} {}^{[i]}p_k \Big) \to {}^{[i]}\varphi$$

is valid for every φ built from $\{p_1, \ldots, p_n\}$.

Proof. Again, we will prove an equivalent (and more semantic) statement:

$$\{p_1, \ldots, p_n\} \subseteq \mathsf{PA}_i(w) \quad \text{implies} \quad (M, w) \Vdash {}^{[i]}\varphi$$

for every φ built from $\{p_1, \ldots, p_n\}$. Again, the two statements are indeed equivalent because of the semantic interpretation of formulas of the form ${}^{[i]}p$.

By assuming $\{p_1, \ldots, p_n\} \subseteq \mathsf{PA}_i(w)$, the proof proceeds by induction on φ. The base case is again immediate, and the inductive ones follow from the inductive hypothesis and the validities of Table 4.2 derived from Definition 4.3. Details can be found in Appendix A.8. ■

Proposition 4.2 *Let the pair (M, w) be a pointed semantic model and i an agent. Let $\{p_1, \ldots, p_n\} \subseteq \mathsf{P}$ be a subset of atomic propositions.*

If i is aware of all atoms in $\{p_1, \ldots, p_n\}$, that is, if $(M, w) \Vdash \mathrm{Aw}_i\, p_k$ for every $k \in \{1, \ldots, n\}$, then she is aware of any formula built from such atoms, that is, $(M, w) \Vdash \mathrm{Aw}_i\, \varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. In other words, the formula

$$\Big(\bigwedge_{k \in \{1,\ldots,n\}} \mathrm{Aw}_i\, p_k \Big) \to \mathrm{Aw}_i\, \varphi$$

is valid for every φ built from $\{p_1, \ldots, p_n\}$.

Proof. Suppose $(M, w) \Vdash \mathrm{Aw}_i\, p_k$ for every $k \in \{1, \ldots, n\}$. Then, $(M, w) \Vdash \Box_i\, {}^{[i]} p_k$, that is, $(M, u) \Vdash {}^{[i]} p_k$ for every $k \in \{1, \ldots, n\}$ and every u such that $R_i w u$. Pick any such u; by Lemma 4.2, $(M, u) \Vdash {}^{[i]}\varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. Hence, $(M, w) \Vdash \Box_i\, {}^{[i]}\varphi$, that is $(M, w) \Vdash \mathrm{Aw}_i\, \varphi$. ■

As mentioned before, our *awareness of* notion assumes that all agents are aware of each other. We could drop this assumption and, following van Ditmarsch and French (2009), extend PA_i-sets to provide not only the atoms but also the agents agent i has at her disposal in each possible world. Formulas of the form ${}^{[i]}\varphi$ can be redefined accordingly: for example, ${}^{[i]}(\Box_j \varphi)$ becomes ${}^{[i]}\varphi \wedge {}^{[i]} j$, with ${}^{[i]} j$ true at (M, w) if and only if $j \in \mathsf{PA}_i(w)$. This gives us a more fine-grained *awareness of* notion, and has interesting consequences.

First, we can represent agents that are not aware of *themselves* by simply not including i in the PA_i-sets. Moreover, if an agent i is not aware of any agent, then we have an agent whose explicit information can only be *propositional*: though she may have non-propositional formulas in her A_i-sets, she will not have explicit information about them because she will not be aware of them. In particular, the agent's explicit information will be completely non-introspective since *she will not be aware of herself*. Finally, consider again the notion of introspection. In classical *EL*, *knowledge* of φ is defined as $\Box\varphi$ in models with *equivalence* accessibility relations; this gives the agent positive and negative introspection. In the approaches of the previous and the present chapter this is not the case for the corresponding notion of explicit knowledge even with equivalence relations (as we will see), but the agent can reach introspection by performing the adequate inference steps.[2] With the mentioned extension of agent awareness, explicit introspection becomes a matter not only of the adequate inference, but also a privilege of *self-aware* agents.

The *implicit information* notion This notion defines everything the agent can get to know without changing her current awareness and provided she has the tools (that is, the rules) to perform the necessary inferences. It is defined as everything that is true in all the worlds the agent considers possible, modulo her current awareness.

[2] In the approach of Chapter 2 the agent's explicit information is, by design, limited to propositional formulas, so no action can give explicit knowledge positive or negative introspection.

Note how this notion has a weak form of omniscience: the agent has implicit information about validities built from atomic propositions she is aware of. Moreover, implicit information is closed under logical consequence.

Proposition 4.3 *Let* (M, w) *be any pointed semantic model and let* i *be an agent.*

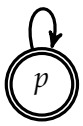
- *Suppose* φ *is a validity and, for every* $p \in \mathsf{atm}(\varphi)$*, we have* $(M, w) \Vdash \mathrm{Aw}_i\, p$*. Then* $(M, w) \Vdash \mathrm{Im}_i\, \varphi$*.*
- *If* $(M, w) \Vdash \mathrm{Im}_i\, (\varphi \to \psi) \wedge \mathrm{Im}_i\, \varphi$*, then* $(M, w) \Vdash \mathrm{Im}_i\, \psi$*.*

Proof. For the first property, φ is a validity, so it holds in any world of any semantic model, in particular, $(M, w) \Vdash \Box_i \varphi$. Since $(M, w) \Vdash \mathrm{Aw}_i\, p$ for every $p \in \mathsf{atm}(\varphi)$, Proposition 4.2 gives us $(M, w) \Vdash \mathrm{Aw}_i\, \varphi$, that is, $(M, w) \Vdash \Box_i\, ^{[i]} \varphi$. Hence $(M, w) \Vdash \Box_i \varphi \wedge \Box_i\, ^{[i]} \varphi$, that is, $(M, w) \Vdash \mathrm{Im}_i\, \varphi$.

For the second property, suppose $(M, w) \Vdash \mathrm{Im}_i\, (\varphi \to \psi) \wedge \mathrm{Im}_i\, \varphi$. Then we have $(M, w) \Vdash \Box_i (\varphi \to \psi) \wedge \Box_i \varphi$, hence we have $(M, w) \Vdash \Box_i \psi$. But we also have $(M, w) \Vdash \Box_i\, ^{[i]} (\varphi \to \psi)$, hence $(M, w) \Vdash \Box_i\, ^{[i]} \psi$ and therefore $(M, w) \Vdash \mathrm{Im}_i\, \psi$. ■

The *explicit information* notion This is the strongest of the three notions: explicit information implies awareness and implicit information.

Since A_i-sets do not have any special requirement, nothing needs to be explicitly known, and therefore the notion does not have any closure property. This suits us well, since the explicit information of an agent i does not *need* to have any strong closure requirement. We can easily imagine a situation in which she is not explicitly informed about some validity she implicitly has, like the one represented in the leftmost model of the following diagram (with PA_i- and A_i-sets presented in that order), or another in which her explicit information is not closed under logical consequence, like the one represented on the rightmost model.



- $\mathrm{Im}\, (p \vee \neg p)$
- $\neg \mathrm{Ex}\, (p \vee \neg p)$

$\{p\}, \{\,\}$

- $\mathrm{Ex}\, p \wedge \mathrm{Ex}\, (p \to q)$
- $\neg \mathrm{Ex}\, q$

$\{p, q\}, \{p, p \to q\}$

But the fact that *explicit information* does not *need* to have any special property does not mean that it cannot. From our *dynamic* perspective, explicit information does not need built-in properties that guarantee the agent has certain amount of minimal information; what it needs is the appropriate set of actions that explains how the agent gets that information.

Hierarchy of the notions By simply unfolding their definitions, it follows that our three notions behave exactly like in Figure 4.1 (page 82).

Proposition 4.4 (The hierarchy of the notions) *In* **M***-models, our relevant notions of information have the following properties:*

- *explicit information* $\subseteq$ *implicit information;*
- *implicit information* $\subseteq$ *awareness of.*

This hierarchy is reflected in the following validities:

$$\mathrm{Ex}_i\,\varphi \to \mathrm{Im}_i\,\varphi \qquad \text{and} \qquad \mathrm{Im}_i\,\varphi \to \mathrm{Aw}_i\,\varphi$$ ■

Interaction between the model components In our general class of models there is no relation between propositional availability, access set and rule set functions and accessibility relations. But by asking for particular requirements we obtain particular kinds of agents.

Consider the following properties, relating accessibility relations with propositional availability and access, respectively.

- If available atoms are preserved by the accessibility relation, that is, if $p \in \mathsf{PA}_i(w)$ implies $p \in \mathsf{PA}_i(u)$ for all worlds u such that R_iwu, then agent i's information satisfies what we call *weak introspection on available atoms*, a property characterized by the formula

$$^{[i]}p \to \Box_i\, ^{[i]}p$$

- In a similar way, if accessible formulas are preserved by the accessibility relation, that is, if $\varphi \in \mathsf{A}_i(w)$ implies $\varphi \in \mathsf{A}_i(u)$ for all worlds u for which R_iwu, then agent i's information satisfies *weak introspection on accessible formulas*, characterized by

$$\mathrm{A}_i\,\varphi \to \Box_i \mathrm{A}_i\,\varphi$$

Note the effect of these properties (similar in spirit to the *coherence* of Chapter 2) in combination with properties of R. With preorders, PA_i and A_i become persistent; with equivalence relations, PA_i and A_i become a function from *equivalence classes* to sets of atoms and formulas, respectively. Moreover, reflexive models with these two properties have the following validities:

- $\Box_i\, ^{[i]}\varphi \;\leftrightarrow\; ^{[i]}\varphi$,
- $\Box_i(\,^{[i]}\varphi \wedge \varphi) \;\leftrightarrow\; (\,^{[i]}\varphi \wedge \Box_i\varphi)$,
- $\Box_i(\,^{[i]}\varphi \wedge \varphi \wedge \mathrm{A}_i\,\varphi) \;\leftrightarrow\; (\,^{[i]}\varphi \wedge \Box_i\varphi \wedge \mathrm{A}_i\,\varphi)$.

This shows how, under the mentioned properties, our definitions for the three notions coincide in spirit with the definition of explicit information of Fagin and Halpern (1988) where access, the A-part of the definition, falls outside the scope of the modal operator.[3]

[3] A more detailed comparison between the works is provided in Subsection 4.3.5.

More interestingly, and as we mentioned before, our semantic models do not impose any restriction for formulas in access sets. In particular, they can contain formulas involving atomic propositions that are not in the corresponding propositional availability set, that is, $\mathrm{A}_i\varphi \wedge \neg({}^{[i]}\varphi)$ is satisfiable. Models in which formulas in access sets are built only from available atoms (semantically, $\varphi \in \mathsf{A}_i(w)$ implies $\mathsf{atm}(\varphi) \subseteq \mathsf{PA}_i(w)$; syntactically, $\mathrm{A}_i\varphi \rightarrow {}^{[i]}\varphi$) forces what we call *strong unawareness*: if the agent is unaware of φ, then becoming aware of it does not give her any explicit information about φ, simply because φ (or any formula involving it) cannot be in her access set.

On the other hand, our unrestricted setting allows us to additionally represent what we call *weak unawareness*: becoming aware of φ can give the agent explicit information about it because φ can be already in her access set. This allows us to model a *remembering* notion: I am looking for the keys in the bedroom, and then when someone introduces the possibility for them to be in the kitchen, I remember that I actually left them next to the oven.

Other definable notions What about the reading of other combinations of access to worlds with access to formulas and propositional availability? Though we will not pursue a systematic study of all the technically definable notions and their interpretation, a good intuition about them can be obtained by reading them in terms of what they miss in order to become explicit information. For example the *EL* definition of information, $\Box_i\varphi$, characterizes now information that will become explicit as soon as the agent becomes aware of φ and acknowledges it as true.[4] In the same way, $\Box_i(\varphi \wedge \mathrm{A}_i\varphi)$ expresses that φ is a piece of information that only needs for the agent to become aware of it in order to become explicit information; in other words, φ is information the agent is not currently aware of (some form of forgotten information), as we will see when we use the framework to represent our running example (Subsection 4.3.4).

4.3.3 Working with *knowledge*

Our current definitions do not guarantee that the agent's information is *true*, simply because the real world does not need to be among the ones she considers possible. Different from Chapter 2, in order to work with true information, that is, with the notion of *knowledge*, this time we only need to work in models where the accessibility relations are reflexive: the *truth* requirement is not needed anymore. Following the standard *EL* approach, we will assume that the relations are full equivalence relations.

Definition 4.7 (Class $\mathbf{M}_K$) A semantic model $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}_i, \mathsf{R}_i\rangle$ is in the class $\mathbf{M}_K$ if and only if R_i is an *equivalence* relation for every agent i. ◄

[4] This emphasizes that, in classical *EL*, understanding $\Box_i\varphi$ as explicit information assumes, precisely, that the agent is aware of all relevant formulas, and has acknowledged as true those that are so in each possible world.

Proposition 4.5 *In $\mathbf{M}_K$-models, every piece of implicit and explicit information is true (in the case of formulas) and truth-preserving (in the case of rules). In other words,* $\mathrm{Im}_i\,\varphi \to \varphi$ *and* $\mathrm{Ex}_i\,\varphi \to \varphi$ *are valid in the case of formulas, and* $\mathrm{Im}_i\,\rho \to \mathsf{tr}(\rho)$ *and* $\mathrm{Ex}_i\,\rho \to \mathsf{tr}(\rho)$ *in the case of rules.*

Proof. For the case of formulas, we prove the first validity. Suppose $(M, w) \Vdash \mathrm{Im}_i\,\varphi$; then $(M, w) \Vdash \Box_i({}^{[i]}\varphi \wedge \varphi)$, that is, ${}^{[i]}\varphi \wedge \varphi$ holds in all worlds R_i-accessible from w. But R_i is reflexive, so ${}^{[i]}\varphi \wedge \varphi$ holds at w and then so does φ; hence, $\mathrm{Im}_i\,\varphi \to \varphi$ is valid. The second validity follows from this and the hierarchy proved in Proposition 4.4. The case of rules can be proved in a similar way. ■

When working with models in $\mathbf{M}_K$, we will use the term *knowledge* instead of the term *information*, that is, we will talk about implicit and explicit *knowledge*. A sound and complete axiom system for validities of $\mathcal{L}$ in $\mathbf{M}_K$-models is given by the standard multi-agent $S5$ system.

Theorem 4.2 (Axiom system for $\mathcal{L}$ w.r.t. $\mathbf{M}_K$) *The axiom system of Table 4.1 plus the axioms of Table 4.4 is sound and strongly complete for formulas of $\mathcal{L}$ with respect to models in $\mathbf{M}_K$.* ■

T	$\vdash \Box_i\varphi \to \varphi$	for every agent i
4	$\vdash \Box_i\varphi \to \Box_i\Box_i\varphi$	for every agent i
5	$\vdash \neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$	for every agent i

Table 4.4: Extra axioms for $\mathcal{L}$ w.r.t. $\mathbf{M}_K$

Introspection Let us review what introspection properties our three notions of information have.

Our notion of awareness has positive introspection, regardless of the properties of the accessibility relation. It defines a language and does not take into account awareness about agents, so every time an agent i is aware of a formula φ, she is also aware of her being aware.

Proposition 4.6 *In our general class of models* $\mathbf{M}$*, agents are always aware of her own awareness. In other words, we have the following validity:*

$$\mathrm{Aw}_i\,\varphi \to \mathrm{Aw}_i\,\mathrm{Aw}_i\,\varphi$$

Proof. Suppose $\mathrm{Aw}_i\,\varphi$ holds at some world in a given model. From Proposition 4.1 we know that the agent is aware of all atoms in φ; from Proposition 4.2 we know that she is also aware of every formula built from such atoms, in particular, she is aware of $\mathrm{Aw}_i\,\varphi$ itself. ■

Note that if we consider also awareness about agents, for the awareness notion to be positively introspective, the agent needs to be aware of herself.

On the other hand, awareness does not have negative introspection: if the agent is not aware of φ, that is, if $\neg \mathrm{Aw}_i\, \varphi$ holds, this does not imply that she is aware of not being aware of φ, that is, $\mathrm{Aw}_i \neg \mathrm{Aw}_i\, \varphi$ does not need to hold. In fact, what we have is the following.

Proposition 4.7 *In our general class of models* $\mathbf{M}$*, agents are not aware of her lack of awareness. In other words, we have the following validity:*

$$\neg \mathrm{Aw}_i\, \varphi \rightarrow \neg \mathrm{Aw}_i \neg \mathrm{Aw}_i\, \varphi$$

Proof. Suppose $\neg \mathrm{Aw}_i\, \varphi$ holds at some world in a given model. From Proposition 4.2 we know that the agent is not aware of at least one atom of φ; then from Proposition 4.1 we know she cannot be aware of any formula involving such atom, in particular, she cannot be aware of $\neg \mathrm{Aw}_i\, \varphi$ itself. ■

Now consider the notion of implicit information. In the general case we have neither positive nor negative introspection, just like in *EL*. Things change if we focus on $\mathbf{M}_K$-models, that is, if we focus on implicit *knowledge*.

In standard *EL*, assuming equivalence accessibility relations (in particular, assuming transitive relations) gives $\Box\, \varphi$ positive introspection, that is, $\Box_i \varphi \rightarrow \Box_i \Box_i \varphi$. In our case, the assumption gives $\mathrm{Im}_i\, \varphi$ positive introspection, that is,

Proposition 4.8 *In* $\mathbf{M}_K$*-models, implicit knowledge has the positive introspection property, that is, the following formula is valid:*

$$\mathrm{Im}_i\, \varphi \rightarrow \mathrm{Im}_i\, \mathrm{Im}_i\, \varphi$$

Proof. Unfolding the definitions produces the following formula

$$\Box_i\, {}^{[i]}\varphi \wedge \Box_i \varphi \;\; \rightarrow \;\; \Box_i\, {}^{[i]}\Big(\Box_i({}^{[i]}\varphi \wedge \varphi)\Big) \wedge \Box_i \Box_i\, {}^{[i]}\varphi \wedge \Box_i \Box_i \varphi$$

By Propositions 4.1 and 4.2, the first conjunct of the antecedent implies the first one of the consequent; by transitivity (axiom *4* of Table 4.4), the first and second conjunct of the antecedent imply the second and third ones of the consequent, respectively. ■

But implicit knowledge is not negatively introspective, that is, the formula $\neg \mathrm{Im}_i\, \varphi \rightarrow \mathrm{Im}_i \neg \mathrm{Im}_i\, \varphi$ is not valid. The reason is that, though the accessibility relation is euclidean (that is, we have axiom *5*), $\mathrm{Im}_i\, \varphi$ may fail because i is not aware of φ; then she would not be aware of $\neg \mathrm{Im}_i\, \varphi$ either and therefore she would not know it implicitly (recall that awareness is a requisite for implicit information, and therefore for implicit knowledge). Nevertheless, negative introspection holds if the agent is aware of φ.

Proposition 4.9 *In $\mathbf{M}_K$-models, if the agent does not know φ implicitly but still she is aware of it, then she knows implicitly that she does not know φ implicitly. That is,*

$$\Big(\neg \mathrm{Im}_i\, \varphi \wedge \mathrm{Aw}_i\, \varphi\Big) \to \mathrm{Im}_i\, \neg \mathrm{Im}_i\, \varphi$$

Proof. Suppose $\neg \mathrm{Im}_i\, \varphi \wedge \mathrm{Aw}_i\, \varphi$ holds. Then we have $\Box_i\, {}^{[i]}\varphi$ and $\neg \Box_i \varphi$. From the first and Propositions 4.1 and 4.2, we get $\Box_i\, {}^{[i]}(\neg \mathrm{Im}_i\, \varphi)$. From both and axioms *4* and *5*, respectively, we get $\Box_i \Box_i\, {}^{[i]}\varphi$ and $\Box_i \neg \Box_i \varphi$, that is, $\Box_i \neg (\Box_i\, {}^{[i]}\varphi \wedge \Box_i \varphi)$ or, shortening, $\Box_i \neg \mathrm{Im}\, \varphi$. These two pieces gives us $\mathrm{Im}_i\, \neg \mathrm{Im}_i\, \varphi$. ■

For explicit information, the notion lacks positive introspection. Even if we move to explicit knowledge, the A_i-sets do not need to satisfy any closure property and, in particular, $\varphi \in \mathsf{A}_i(w)$ does not imply $\mathrm{Ex}_i\, \varphi \in \mathsf{A}_i(w)$: having recognized φ as true does not make the agent automatically recognize her explicit knowledge about it. Positive introspection is a property of $\mathbf{M}_K$-models that additionally satisfy the just described requirement, syntactically characterized by the formula $\mathrm{A}_i\, \varphi \to \mathrm{A}_i\, \mathrm{Ex}_i\, \varphi$.

Proposition 4.10 *In $\mathbf{M}_K$-models in which $\mathrm{A}_i\, \varphi \to \mathrm{A}_i\, \mathrm{Ex}_i\, \varphi$ is valid, explicit information has the positive introspection property, that is, the following formula is valid:*

$$\mathrm{Ex}_i\, \varphi \to \mathrm{Ex}_i\, \mathrm{Ex}_i\, \varphi$$

Proof. Unfolding the definitions produces the following formula

$$\begin{aligned} \Box_i\, {}^{[i]}\varphi \wedge \Box_i \varphi \wedge \Box_i \mathrm{A}_i\, \varphi \;\to\; & \Box_i\, {}^{[i]}\Big(\Box_i({}^{[i]}\varphi \wedge \varphi \wedge \mathrm{A}_i\, \varphi)\Big) \;\wedge\; \Box_i \Box_i\, {}^{[i]}\varphi \\ & \wedge\, \Box_i \Box_i \varphi \;\wedge\; \Box_i \Box_i \mathrm{A}_i\, \varphi \;\wedge\; \Box_i \mathrm{A}_i\, \mathrm{Ex}_i\, \varphi \end{aligned}$$

By Propositions 4.1 and 4.2, the first conjunct of the antecedent implies the first of the consequent. By transitivity, the first, second and third conjunct of the antecedent imply the second, third and fourth of the consequent, respectively. Finally, the third conjunct of the antecedent and the assumed $\mathrm{A}_i\, \varphi \to \mathrm{A}_i\, \mathrm{Ex}_i\, \varphi$ imply the fifth conjunct of the consequent. ■

The notion of explicit information also lacks negative introspection. This time, even in the knowledge case ($\mathbf{M}_K$-models) , awareness or even the stronger notion of implicit knowledge of φ are not enough; we also need to assume also the validity of $\neg \Box_i \mathrm{A}_i\, \varphi \to \Box_i \mathrm{A}_i\, \neg \mathrm{Ex}_i\, \varphi$, a formula requiring that, if there are epistemic alternatives where the agent has not acknowledged φ, then in all epistemic alternatives she has acknowledged that she does not know φ explicitly. Only then explicit knowledge gets negative introspection.

Proposition 4.11 *In $\mathbf{M}_K$-models in which $\neg\Box_i \mathrm{A}_i \varphi \rightarrow \Box_i \mathrm{A}_i \neg \mathrm{Ex}_i \varphi$ is valid, if the agent does not know φ explicitly but still she has implicit knowledge about it, then she knows explicitly that she does not know φ explicitly. In a formula,*

$$\left(\neg \mathrm{Ex}_i \varphi \wedge \mathrm{Im}_i \varphi\right) \rightarrow \mathrm{Ex}_i \neg \mathrm{Ex}_i \varphi$$

Proof. The antecedent of the implication gives us $\Box_i {}^{[i]}\varphi$, $\Box_i \varphi$ and $\neg\Box_i \mathrm{A}_i \varphi$. From the first we get $\Box_i {}^{[i]}(\neg \mathrm{Ex}_i \varphi)$ as before. From the first, second and third together with *4* and *5* we get $\Box_i\Box_i {}^{[i]}\varphi$, $\Box_i\Box_i\varphi$ and $\Box_i\neg\Box_i \mathrm{A}_i \varphi$, that is, $\Box_i\neg(\Box_i {}^{[i]}\varphi \wedge \Box_i\varphi \wedge \Box_i \mathrm{A}_i \varphi)$ or, shortening, $\Box_i\neg \mathrm{Ex}_i \varphi$. The third and the further assumption gives us $\Box_i \mathrm{A}_i \neg \mathrm{Ex}_i \varphi$. These three pieces give us $\mathrm{Ex}_i \neg \mathrm{Ex}_i \varphi$. ■

Though explicit *knowledge* does not have neither positive nor negative introspection in the general case, it does have a weak form of them.

Proposition 4.12 *In $\mathbf{M}_K$-models, explicit knowledge has the weak positive and negative introspection property, that is, the following formulas are both valid.*

$$\mathrm{Ex}_i \varphi \rightarrow \mathrm{Im}_i \mathrm{Ex}_i \varphi \quad \textit{and} \quad (\neg \mathrm{Ex}_i \varphi \wedge \mathrm{Aw}_i \varphi) \rightarrow \mathrm{Im}_i \neg \mathrm{Ex}_i \varphi$$

Proof. The proof is similar to those of Propositions 4.8 and 4.10, by unfolding the definitions and applying standard modal principles. ■

The first statement of Proposition 4.12 says that if i has explicit knowledge that φ, then she implicitly knows (that is, she should be in principle able to infer) that she has explicit knowledge that φ. The second one says that if she does not have explicit knowledge that φ but still she is aware of it, then she implicitly knows that she does not have explicit knowledge.

Note how, from our dynamic perspective, the additionally properties we have asked for to reach introspection can be understood not as static requirements, but as actions that, after performed, will yield the indicated results.

4.3.4 The information state of the Jury

We can now provide a formal analysis of Example 4.1.

Example 4.2 (The juror's information) Define the following atoms:

`gls`: *the woman wears glasses*	`mkns`: *she has marks in the nose*
`esq`: *her eyesight is in question*	`glt`: *the accused is guilty beyond any reasonable doubt*

A	$\Box_A(\mathsf{tr}(\sigma_1) \wedge \mathrm{R}_A\,\sigma_1)$ $\Box_A(\mathsf{tr}(\sigma_2) \wedge \mathrm{R}_A\,\sigma_2)$ $\Box_A(\mathsf{tr}(\sigma_3) \wedge \mathrm{R}_A\,\sigma_3)$	$\Box_A(\mathtt{mkns} \wedge \mathrm{A}_A\,\mathtt{mkns})$	$\mathrm{Aw}_A\,\mathtt{glt}$ $\mathrm{Aw}_A\,\mathtt{esq}$
B	$\Box_B(\mathsf{tr}(\sigma_1) \wedge \mathrm{R}_B\,\sigma_1)$ $\Box_B(\mathsf{tr}(\sigma_2) \wedge \mathrm{R}_B\,\sigma_2)$ $\Box_B(\mathsf{tr}(\sigma_3) \wedge \mathrm{R}_B\,\sigma_3)$		$\mathrm{Aw}_B\,\mathtt{glt}$
C	$\Box_C(\mathsf{tr}(\sigma_1) \wedge \mathrm{R}_C\,\sigma_1)$ $\Box_C(\mathsf{tr}(\sigma_2) \wedge \mathrm{R}_C\,\sigma_2)$ $\Box_C(\mathsf{tr}(\sigma_3) \wedge \mathrm{R}_C\,\sigma_3)$		$\mathrm{Aw}_C\,\mathtt{glt}$
G	$\Box_G(\mathsf{tr}(\sigma_1) \wedge \mathrm{R}_G\,\sigma_1)$ $\Box_G(\mathsf{tr}(\sigma_2) \wedge \mathrm{R}_G\,\sigma_2)$ $\Box_G(\mathsf{tr}(\sigma_3) \wedge \mathrm{R}_G\,\sigma_3)$	$\Box_G(\mathtt{mkns} \wedge \mathrm{A}_G\,\mathtt{mkns})$	$\mathrm{Aw}_G\,\mathtt{glt}$

Table 4.5: Information state of the agents in Example 4.1.

Let the relevant rules, abbreviated as $\varphi \Rightarrow \psi$ with φ the (conjunction of the) premise(s) and ψ the conclusion, be the following:

$$\sigma_1 : \mathtt{mkns} \Rightarrow \mathtt{gls} \qquad \sigma_2 : \mathtt{gls} \Rightarrow \mathtt{esq} \qquad \sigma_3 : \mathtt{esq} \Rightarrow \neg\mathtt{glt}$$

Table 4.5 indicates the information state of the relevant members of the Jury at the beginning of the conversation. In words, not only are the three rules truth-preserving in all worlds every agent considers possible, but also each agent has acknowledged that (first column). In other words, each one of them accepts that if somebody has some marks on her nose then she wears glasses, that if she wears glasses then we can question her eyesight, and that someone with questioned eyesight cannot be a credible eye-witness. However, only A and G have access to the bit of information which is needed to trigger the inference, namely, that the witness had those peculiar marks on her nose (second column). Still, this is not enough since they are not considering the atoms `mkns` and `gls` in their 'working languages', that is, they are not currently aware of these possibilities. The only bit of language they are considering concerns the defendant being guilty or not and, in A's case, the concern about the witness eyesight (third column). ◀

All in all, the key aspect in Example 4.2 here is that the pieces of information that can possibly generate explicit knowledge are spread across the group. The effect of the deliberation is to share these bits through communication, which is the topic of Section 4.4. Before getting there, however, it is worthwhile to put the developed framework in perspective with other options for the notions of awareness, implicit and explicit information.

4.3.5 Other approaches

We have defined our main notions of information, studying some of their properties in the general case (Subsection 4.3.2) as well under the assumption of models built on equivalence classes (Subsection 4.3.3). We will now make a brief recapitulation about alternative definitions of these notions, comparing them with the approach of the present chapter.

Syntactic awareness vs. semantic awareness The proposed formalization of the *awareness of* notion is based on the intuition that, at each state, each agent has only a particular subset of the language at her disposal for expressing her information. This intuition is modeled via formulas of the form $^{[i]}p$ for an atom p, and their inductive extension to any formula (Definition 4.3).

This is a syntactic way to look at the atomic propositions available to agents and, thus, to look at awareness generated by a set of atoms. An alternative semantic approach can be obtained by means of a relation that holds between two worlds whenever they coincide in the truth-value of the atomic propositions in a given $\mathtt{Q} \subseteq \mathtt{P}$, as presented in Grossi (2009).

Definition 4.8 (Propositional equivalence up to a signature) Let $\mathtt{P}$ be a set of atomic propositions, W a set of possible worlds and V an atomic valuation function $V : W \to \wp(\mathtt{P})$ as before. Two worlds $w, u \in W$ are *propositionally equivalent up to a signature* $\mathtt{Q} \subseteq \mathtt{P}$ (*propositionally* $\mathtt{Q}$*-equivalent*) if and only if, for every $p \in \mathtt{Q}$, we have $p \in V(w)$ if and only if $p \in V(u)$, that is, if w and u coincide in the atomic valuation of all atoms in $\mathtt{Q}$. When $w, u \in W$ are $\mathtt{Q}$-equivalent, we write $w \sim_{\mathtt{Q}} u$. Note that $\sim_{\mathtt{Q}}$ is indeed an equivalence relation.[5] ◄

We can use the relation $\sim_{\mathtt{Q}}$ to define a semantic notion of awareness based on atomic propositions. Suppose we are working with a language based on the atoms p and q, but our agent is only aware of p. Then, intuitively, she cannot make propositional difference between worlds that make p and q true, and worlds that make p true but q false, simply because she cannot perceive the only propositional difference between these worlds: the truth-value of q.

More generally, an agent cannot make propositional difference between worlds that coincide in the truth-value of all the atoms she is aware of. In the mentioned case the agent is only aware of p, and therefore she cannot make a propositional distinction between two worlds w and u if $V(w) = \{p, q\}$ and $V(u) = \{p\}$, but also if $V(w) = \{q\}$ and $V(u) = \{\,\}$. These propositionally indistinguishable worlds are precisely the worlds related by $\sim_{\{p\}}$.

The equivalence relation $\sim_{\mathtt{Q}}$ creates equivalence classes by grouping worlds that have the same atomic valuation for atoms in $\mathtt{Q}$. Then, any propositional formula built only from such atoms has an uniform truth-value, true or false,

[5] This notion of states *propositionally* equivalent up to a signature can be extended to a notion of states *modally* equivalence up to a signature (van Ditmarsch and French 2009).

in all the worlds of each equivalence class. So take an agent i whose available atoms at world w are $\mathsf{PA}_i(w)$. If she can make use of a propositional formula γ to express her information (that is, if γ is built from atoms in $\mathsf{PA}_i(w)$, what our $[i]\gamma$ expresses), then the formula $[\sim_{\mathsf{PA}_i(w)}]\,\gamma \vee [\sim_{\mathsf{PA}_i(w)}]\,\neg\gamma$ is true at w, with the 'box' modality $[\sim_{\mathsf{PA}_i(w)}]$ interpreted via the relation $\sim_{\mathsf{PA}_i(w)}$ in the standard way.

Nevertheless, the other direction does not hold: the fact that $[\sim_{\mathsf{PA}_i(w)}]\,\gamma \vee [\sim_{\mathsf{PA}_i(w)}]\,\neg\gamma$ holds at w does not imply that γ is built only from atoms in $\mathsf{PA}_i(w)$. The reason is that, ultimately, the truth-value of formulas of the form $[\sim_{\mathsf{PA}_i(w)}]\,\gamma$ depends only on the *truth-value* of γ's atoms, regardless of which ones they actually are. For example, in a model in which p and q are true in all the worlds, both $[\sim_{\mathsf{PA}_i(w)}]\,p \vee [\sim_{\mathsf{PA}_i(w)}]\,\neg p$ and $[\sim_{\mathsf{PA}_i(w)}]\,q \vee [\sim_{\mathsf{PA}_i(w)}]\,\neg q$ are true, even if agent i can use p ($p \in \mathsf{PA}_i(w)$) but not q ($q \notin \mathsf{PA}_i(w)$). So different from our approach, this semantic alternative does not define a language by itself.

Other approaches to awareness Let us make a brief comparison between the approach to awareness of this chapter and other syntactic proposals.

The notion of awareness in Fagin and Halpern (1988), and hence our dynamization of it in Chapter 3, is modelled by assigning to each agent a set of formulas in each possible world. Such sets, in principle, lack any particular property, but several possibilities are mentioned in that paper, including awareness based on a set of atomic propositions like the one have discussed in this chapter. From this perspective we can say that we look at one particular form of awareness, but we emphasize again that our notion is not defined from a set of formulas with some particular closure property, as Fagin and Halpern (1988) proposes, but from a set of atoms, our PA_i-sets, and then a recursive definition of formulas build from them, our Definition 4.3.

There is a more important difference. Their notion is defined as a set of formulas *at the evaluation point*: an agent is aware of φ at world w if and only if φ is in the corresponding set *of world w*. This differs from our definition, where we look not at the atomic propositions the agent has at her disposal in the evaluation point, but at those she has *in every world she considers possible*. As we have mentioned, the two definitions coincide under the assumption that the accessibility relations are reflexive and preserve PA_i-sets.

Putting aside the notion of *awareness of an agent* discussed before, in van Ditmarsch and French (2009) the authors present an approach similar to ours: each possible world assigns to each agent a set of atomic propositions, and the notion of *awareness of* is defined in terms of such set. Nevertheless, they follow Fagin and Halpern's idea, defining the notion relative only to the evaluation point, $[i]p$ in our syntax. Their notion indeed defines a *language*, like ours, but it does it relative to a single point and not to the worlds the agent considers possible. Such definition allows situations like *"the agent is aware of p ($[i]p$ in our notation), but she is not aware of it in all the worlds she considers possible ($\neg\Box_i\,[i]p$)"*.

Other definitions of explicit information The formal definition of explicit information/knowledge has several variants in the literature. Among them, we can mention the $\Box_i\varphi \wedge A_i\,\varphi$ (i.e., standard epistemic modality plus awareness) of Fagin and Halpern (1988); van Ditmarsch and French (2009), and the $A_i\,\varphi$ of Duc (1997); Jago (2009); van Benthem (2008c) and our Chapter 2. The definition we have used in this chapter follows the spirit of the one we argued for in Section 3.2, in which all the ingredients of explicit information fall under the scope of the modal universal modality $\Box$.

We have already argued about the reasons for going from the $A_i\,\varphi$ of Chapter 2 to the $\Box_i(\varphi \wedge A_i\,\varphi)$ of Chapter 3 and for choosing it over the $\Box_i\varphi \wedge A_i\,\varphi$ of Fagin and Halpern (1988) (Section 3.2). Let us now emphasize again the reasons for having two components $^{[i]}\varphi$ and $A_i\,\varphi$, that is, for distinguishing between the formulas the agent is aware of and those she has acknowledged as true.

First, having two components accounts for cases well-known, for instance, in mathematical practice: while trying to prove a statement we are (hopefully) aware of all the relevant notions. But even being aware does not guarantee that we can recognize as true what it is so. In such cases, formulas of the form $^{[i]}\varphi$ allow us to express what the agent can talk about, and formulas of the form $A_i\,\varphi$ allow us to express what she has acknowledged as true (with inference and observation the most common acts that result in such acknowledgment).

Now consider what we would get by using only one component. Having a notion of explicit information that uses only formulas the agent has acknowledged as true would fall short in capturing situations in which the agent does not consider all relevant possibilities, like our running Jury example. On the other hand, a definition in which only the *awareness of* notion is taken as an extra component of explicit information gives us two possibilities: either this awareness notion defines a language based on atomic propositions, or it does not. If it does not, like in the original general awareness from Fagin and Halpern (1988), then the agent can be aware of φ and ψ without being aware of $\varphi \wedge \psi$; this is undesirable because, from this chapter's perspective, becoming aware of a possibility should also make the agents aware of boolean combinations of it. On the other hand, if this notion is defined as a full language, like in van Ditmarsch and French (2009), then the agent's explicit information would be closed under logical consequence. Being explicitly informed about φ and $\varphi \to \psi$ would mean that the agent has implicit information about both formulas and is also aware of them. Since implicit information is closed under logical consequence, the agent will be implicitly informed about ψ; since the agent is aware of $\varphi \to \psi$, she is also aware of ψ. Therefore, the agent will be explicitly informed about ψ, which is also clearly undesirable in our setting.

The present chapter works on the idea that two requirements are needed in order to make explicit a piece of information. First, the agent should be aware of that information, in the sense that such information should be a notion

the agent can express with her current language. Second, the agent should have also acknowledged somehow that this information is in fact true. Based on these two extra requirements we have identified three different notions of information, awareness of, implicit and explicit information, and we have reviewed some of their properties. It is now time to turn our attention to the actions that modify them.

4.4 Dynamics of information

Our framework allows us to describe the information a set of agents have at some given stage. It is time to provide the tools that allow us to describe how this information changes. Three are the informational actions relevant for our example and our discussion: becoming aware, inference and public announcement.

The first action, *becoming aware*, makes the agent aware of a given atomic proposition q. This is the processes through which the agent extends her current language, and it can be interpreted as the introduction of a topic in a conversation. The second one, *inference*, allows the agent to extend the information she can access by means of a rule application. This is the process through which the agent acknowledges as true certain information that up to this point has been just implicit, therefore making it part of her explicit information. The third one, *announcement*, represents the agent's interaction with the external world: she announces to the others something that she explicitly knows.

For each one of these actions, we define a model operation representing it.

Definition 4.9 (Dynamic operations) Let $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}_i, \mathsf{R}_i \rangle$ be a semantic model in **M**.

- Take $q \in \mathtt{P}$ and $j \in \mathtt{Ag}$. The *atomic awareness* operation produces the model $M_{\rightsquigarrow^j_q} = \langle W, R_i, V, \mathsf{PA}'_i, \mathsf{A}_i, \mathsf{R}_i \rangle$, differing from M just in the propositional availability function of agent j, which is given for every world $w \in W$ by

$$\mathsf{PA}'_j(w) := \mathsf{PA}_j(w) \cup \{q\}$$

  In words, the operation $\rightsquigarrow^j_q$ adds the atomic proposition q to the propositional availability set of agent j in all worlds of the model.

- Take $\sigma \in \mathcal{L}_r$ and $j \in \mathtt{Ag}$. The *agent inference* operation produces the model $M_{\hookrightarrow^j_\sigma} = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}'_i, \mathsf{R}_i \rangle$, differing from M just in the access set function of agent j, which is given for every world $w \in W$ by

$$\mathsf{A}'_j(w) := \begin{cases} \mathsf{A}_j(w) \cup \{\mathsf{cn}(\sigma)\} & \text{if } \sigma \in \mathsf{R}_j(w) \text{ and } \mathsf{pm}(\sigma) \subseteq \mathsf{A}_j(w) \\ \mathsf{A}_j(w) & \text{otherwise} \end{cases}$$

  In words, the operation $\hookrightarrow^j_\sigma$ adds the conclusion of σ to the access set of agent j in those worlds in which the agent has already σ and its premises.

- Take $\chi \in \mathcal{L}_f$ and $j \in \mathtt{Ag}$, and recall that $\mathsf{atm}(\chi)$ denotes the set of atomic propositions occurring in χ. The *announcement* operation produces the model $M_{j:\chi!} = \langle W', R_i', V', \mathsf{PA}_i', \mathsf{A}_i', \mathsf{R}_i' \rangle$ where, for every agent $i \in \mathtt{Ag}$,

  - $W' := \left\{ w \in W \mid (M, w) \Vdash \mathrm{Ex}_j \chi \right\}$
  - $R_i' := R_i \cap (W' \times W')$

  and, for every $w \in W'$,

  - $V'(w) := V(w)$
  - $\mathsf{R}_i'(w) := \mathsf{R}_i(w)$
  - $\mathsf{PA}_i'(w) := \mathsf{PA}_i(w) \cup \mathsf{atm}(\chi)$
  - $\mathsf{A}_i'(w) := \mathsf{A}_i(w) \cup \{\chi\}$

  In words, the operation $j : \chi!$ removes worlds where $\mathrm{Ex}_j \chi$ does not hold, restricting the agents' accessibility relation and the valuation to the new domain. It also extends the agents' propositional availability sets with the atomic propositions occurring in χ and extends their access sets with χ itself, preserving rule sets as in the original model. ◀

While the first two operations affect the model components of just one agent, the third one affects those of all of them. Indeed, while the atomic awareness operation $\rightsquigarrow_q^j$ affects only agent j's PA-sets and the inference operation $\hookrightarrow_\sigma^j$ affects only agent j's A-sets, the announcement affects the accessibility relation as well as the PA- and A-sets of *every* agent. But affecting just the model-components of a single agent, like our first two operations do, does not imply that other agents' information does not change. In fact, the atomic awareness and inference operations behave similar to the 'public' consider operation of Chapter 3 (Definition 3.4) in that, by modifying a model component of one agent, they affect the information of the others. In the atomic awareness case, $\rightsquigarrow_q^j$ makes $[j]q$ true in every world in the model, therefore making it true in every world any agent i considers possible, that is, $\Box_i [j]q$ becomes true everywhere. This does not say that every agent has now explicit information about agent j being aware of q, but it does say that they will as soon as they become aware of q and have access to $[i]q$ in all the worlds they consider possible (the other two ingredients for explicit information). Something similar happens with the *inference* operation $\hookrightarrow_\sigma^j$ since it makes $\Box_i \mathrm{A}_j \mathsf{cn}(\sigma)$ true in every world of the model. *Private* versions of these operations can be defined following the action model approach of Section 3.6. We will omit details here.

The announcement operation deserves also extra words for two reasons, and the first is the worlds that the operation discards. Note how an announcement of χ by agent j preserves only the worlds where agent j is explicitly informed about χ (i.e., worlds where $\mathrm{Ex}_j \chi$ holds). This is different from the *observation* operation (Definition 1.5) and its explicit versions (Definitions 2.20 and 3.7), which preserve only worlds where the observed χ holds.

The second reason is how the operation affects the sets of formulas of the preserved worlds. After j announces χ, only χ is added to the A-sets. This choice represent situations in which the hearers acknowledge implicitly that the announcer indeed is explicitly informed about χ (hence only $\mathrm{Ex}_j\,\chi$-worlds survive), but they acknowledge explicitly only χ. There are other variations for defining an announcement; our choice has the advantage of taking the announcer into consideration and also making the hearers explicitly informed about the announced χ in $\mathbf{M}_K$-models (see Proposition 4.14 below).

Our three operations preserve models in $\mathbf{M}_K$.

Proposition 4.13 *If M is a $\mathbf{M}_K$-model, so are $M_{\rightsquigarrow^j_q}$, $M_{\hookrightarrow^j_\sigma}$ and $M_{j:\chi!}$.*

Proof. We just need to prove that the accessibility relations in the three new models are equivalence relations. This is immediate for the first two since neither the domain nor the relation are affected, and also immediate for the third because we go to a sub-model. ■

In order to express the effect of this operations over the agent's information, we extend the language $\mathcal{L}$ with three new existential modalities, $\langle \rightsquigarrow^j_q \rangle$, $\langle \hookrightarrow^j_\sigma \rangle$ and $\langle j : \chi! \rangle$, representing each one of our operations (their universal versions are defined as their corresponding dual, as usual). We call this language *extended* $\mathcal{L}$; the semantic interpretation of the new formulas is as follows.

Definition 4.10 (Semantic interpretation) Let $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{A}_i, \mathsf{R}_i \rangle$ be a semantic model, and take a world $w \in W$. Define the following formulas

$$\mathrm{Pre}_{\hookrightarrow^j_\sigma} := \Big(\bigwedge_{\psi \in \mathsf{pm}(\sigma)} \mathrm{Ex}_j\,\psi \Big) \wedge \mathrm{Ex}_j\,\sigma \qquad\qquad \mathrm{Pre}_{j:\chi!} := \mathrm{Ex}_j\,\chi$$

expressing the precondition for $\hookrightarrow^j_\sigma$ and $j : \chi!$, respectively. Then,

$$\begin{array}{lll}
(M,w) \Vdash \langle \rightsquigarrow^j_q \rangle \varphi & \text{iff} & (M_{\rightsquigarrow^j_q}, w) \Vdash \varphi \\
(M,w) \Vdash \langle \hookrightarrow^j_\sigma \rangle \varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{\hookrightarrow^j_\sigma} \text{ and } (M_{\hookrightarrow^j_\sigma}, w) \Vdash \varphi \\
(M,w) \Vdash \langle j : \chi! \rangle \varphi & \text{iff} & (M,w) \Vdash \mathrm{Pre}_{j:\chi!} \text{ and } (M_{j:\chi!}, w) \Vdash \varphi
\end{array}$$ ◄

Note how the precondition of each operation reflects its intuitive meaning. An agent can extend her language at any point; for applying an inference with σ she needs to know explicitly the rule and all its premises; for announcing χ, the agent simply needs to be explicitly informed about it.

Again, we use reduction axioms in order to provide a sound and complete axiom system for the extended language.

$\vdash \langle \rightsquigarrow^j_q \rangle p \leftrightarrow p$	$\vdash \langle \rightsquigarrow^j_q \rangle {}^{[i]}p \leftrightarrow {}^{[i]}p$ for $i \neq j$
$\vdash \langle \rightsquigarrow^j_q \rangle \neg\varphi \leftrightarrow \neg\langle \rightsquigarrow^j_q \rangle \varphi$	$\vdash \langle \rightsquigarrow^j_q \rangle {}^{[j]}p \leftrightarrow {}^{[j]}p$ for $p \neq q$
$\vdash \langle \rightsquigarrow^j_q \rangle (\varphi \vee \psi) \leftrightarrow \left(\langle \rightsquigarrow^j_q \rangle \varphi \vee \langle \rightsquigarrow^j_q \rangle \psi\right)$	$\vdash \langle \rightsquigarrow^j_q \rangle {}^{[j]}q \leftrightarrow \top$
$\vdash \langle \rightsquigarrow^j_q \rangle \Diamond_i \varphi \leftrightarrow \Diamond_i \langle \rightsquigarrow^j_q \rangle \varphi$	$\vdash \langle \rightsquigarrow^j_q \rangle \mathrm{A}_i \varphi \leftrightarrow \mathrm{A}_i \varphi$
If $\vdash \varphi$, then $\vdash [\rightsquigarrow^j_q] \varphi$	$\vdash \langle \rightsquigarrow^j_q \rangle \mathrm{R}_i \rho \leftrightarrow \mathrm{R}_i \rho$
$\vdash \langle \hookrightarrow^j_\sigma \rangle p \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge p$	$\vdash \langle \hookrightarrow^j_\sigma \rangle {}^{[i]}p \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge {}^{[i]}p$
$\vdash \langle \hookrightarrow^j_\sigma \rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge \neg\langle \hookrightarrow^j_\sigma \rangle \varphi)$	$\vdash \langle \hookrightarrow^j_\sigma \rangle \mathrm{A}_i \varphi \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge \mathrm{A}_i \varphi$ for $i \neq j$
$\vdash \langle \hookrightarrow^j_\sigma \rangle (\varphi \vee \psi) \leftrightarrow \left(\langle \hookrightarrow^j_\sigma \rangle \varphi \vee \langle \hookrightarrow^j_\sigma \rangle \psi\right)$	$\vdash \langle \hookrightarrow^j_\sigma \rangle \mathrm{A}_j \varphi \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge \mathrm{A}_j \varphi$ for $\varphi \neq \mathsf{cn}(\sigma)$
$\vdash \langle \hookrightarrow^j_\sigma \rangle \Diamond_i \varphi \leftrightarrow (\mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge \Diamond_i \langle \hookrightarrow^j_\sigma \rangle \varphi)$	$\vdash \langle \hookrightarrow^j_\sigma \rangle \mathrm{A}_j \mathsf{cn}(\sigma) \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma}$
If $\vdash \varphi$, then $\vdash [\hookrightarrow^j_\sigma] \varphi$	$\vdash \langle \hookrightarrow^j_\sigma \rangle \mathrm{R}_i \rho \leftrightarrow \mathrm{Pre}_{\hookrightarrow^j_\sigma} \wedge \mathrm{R}_i \rho$
$\vdash \langle j : \chi! \rangle p \leftrightarrow \mathrm{Pre}_{j:\chi!} \wedge p$	$\vdash \langle j : \chi! \rangle {}^{[i]}p \leftrightarrow \mathrm{Pre}_{j:\chi!} \wedge {}^{[i]}p$ for $p \notin \mathsf{atm}(\chi)$
$\vdash \langle j : \chi! \rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{j:\chi!} \wedge \neg\langle j : \chi! \rangle \varphi)$	$\vdash \langle j : \chi! \rangle {}^{[i]}p \leftrightarrow \mathrm{Pre}_{j:\chi!}$ for $p \in \mathsf{atm}(\chi)$
$\vdash \langle j : \chi! \rangle (\varphi \vee \psi) \leftrightarrow \left(\langle j : \chi! \rangle \varphi \vee \langle j : \chi! \rangle \psi\right)$	$\vdash \langle j : \chi! \rangle \mathrm{A}_i \varphi \leftrightarrow \mathrm{Pre}_{j:\chi!} \wedge \mathrm{A}_i \varphi$ for $\varphi \neq \chi$
$\vdash \langle j : \chi! \rangle \Diamond_i \varphi \leftrightarrow (\mathrm{Pre}_{j:\chi!} \wedge \Diamond_i \langle j : \chi! \rangle \varphi)$	$\vdash \langle j : \chi! \rangle \mathrm{A}_i \chi \leftrightarrow \mathrm{Pre}_{j:\chi!}$
If $\vdash \varphi$, then $\vdash [j : \chi!] \varphi$	$\vdash \langle j : \chi! \rangle \mathrm{R}_i \rho \leftrightarrow \mathrm{Pre}_{j:\chi!} \wedge \mathrm{R}_i \rho$

Table 4.6: Extra axioms for extended $\mathcal{L}$ w.r.t. $\mathbf{M}_K$

Theorem 4.3 (Reduction axioms for dynamic modalities) *The valid formulas of the language extended $\mathcal{L}$ in $\mathbf{M}_K$-models are exactly those provable by the axioms and rules for the static base language (Tables 4.1 and 4.4) plus the reduction axioms and modal inference rules listed in Table 4.6 (with $\top$ the always true formula).* ■

For the existential modalities of the three operations, $\langle \rightsquigarrow^j_q \rangle$, $\langle \hookrightarrow^j_\sigma \rangle$ and $\langle j : \chi! \rangle$, the reduction axioms in the case of atomic propositions p, negations $\neg$, disjunctions $\vee$ and existential relational modalities $\Diamond_i$ (left column of the table) are standard: the operations do not affect atomic propositions, distribute over $\vee$ and commute with $\neg$ and $\Diamond_i$ modulo their respective preconditions.

The interesting cases are those expressing how propositional availability, access and rule sets are affected (right column of the table). For the $\rightsquigarrow^j_q$ operation, the axioms indicate that only q is added exactly to the propositional availability sets of agent j, leaving the rest of the components of the model as before. For the $\hookrightarrow^j_\sigma$ operation, the axioms indicate that only the access sets of agent j are modified, and the modification consist in adding the conclusion of the applied rule. Finally, axioms for the $j : \chi!$ operation indicate that while rule sets are not affected, propositional availability sets *of every agent* are extended with the atoms of χ and access sets are extended with χ itself.

Basic operations We have introduced only those operations that have a direct interpretation in our setting. One can easily imagine situations, like our running example, in which becoming aware, applying inference and talking to people

are the relevant actions that change the agents' information. Nevertheless, from a technical point of view, our defined inference and announcement operations can be decomposed into more basic ones.

Our *inference* operation modifies access sets A, adding the conclusion of the rule whenever its premises and rule itself are present. But following ideas from van Benthem (2008c) and our Chapter 3, we can define a more basic operation, $+\chi^j_\psi$, that adds an arbitrary formula χ to the access set of agent j on those worlds satisfying certain condition ψ. The formal definition of this model operation is straightforward. For the language, we can introduce a modality $\langle +\chi^j_\psi \rangle$ whose semantic interpretation is given by

$$(M, w) \Vdash \langle +\chi^j_\psi \rangle \varphi \quad \text{iff} \quad (M_{+\chi^j_\psi}, w) \Vdash \varphi$$

Now we can define our inference operation as the conjunction of its precondition and a formula expressing the result of adding the rule's conclusion to the agent's access sets, that is,

$$\langle \hookrightarrow^j_\sigma \rangle \varphi := \mathrm{Pre}_{\hookrightarrow^j_\sigma} \ \wedge \ \langle +\mathsf{cn}(\sigma)^j_\zeta \rangle \varphi$$

with $\zeta := \mathrm{R}_j\, \sigma \wedge \mathrm{A}_j\, \mathsf{pm}(\sigma)$. In words, the above definition says that it is possible for agent j to apply an inference with σ after which φ will be the case, $\langle \hookrightarrow^j_\sigma \rangle \varphi$, if and only if she knows explicitly the rule and all its premises, $\mathrm{Pre}_{\hookrightarrow^j_\sigma}$, and, after adding the rule's conclusion to the access sets of those worlds in which the agent has the rule and its premises, φ is the case, $\langle +\mathsf{cn}(\sigma)^j_\zeta \rangle \varphi$.

Our *announcement* operation removes those worlds in which the announcer is not informed explicitly about the announcement, adding the announced formula's atoms to the PA_i-sets and the announced formula itself to the A_i-sets, for every agent i. But following the *implicit observation* of the previous chapter, we can define a more basic *restriction* operation, $\chi!!$, that simply removes those worlds that do not satisfy the given χ. Again, the formal definition of this model operation is straightforward, and for the language we can introduce a modality $\langle \chi!! \rangle$ whose semantic interpretation is given by

$$(M, w) \Vdash \langle \chi!! \rangle \varphi \quad \text{iff} \quad (M_{\chi!!}, w) \Vdash \varphi \qquad [6]$$

Then, our announcement operation $j : \chi!$ can be defined as a conjunction of its precondition and a formula expressing the result of a sequence of operations: a restriction with $\mathrm{Ex}_j\, \chi$ and then awareness operations (one for every atom in χ) and an addition of χ to the A-sets of every agent. Assuming a finite set of them $i_1, \ldots, i_m$, we have

$$\langle j : \chi! \rangle \varphi := \mathrm{Pre}_{j:\chi!} \ \wedge \ \langle \mathrm{Ex}_j\, \chi!! \rangle \Big(\langle \rightsquigarrow^{i_1}_{q_1} \rangle \cdots \langle \rightsquigarrow^{i_1}_{q_n} \rangle \langle +\chi^{i_1}_{\top} \rangle \Big) \cdots \Big(\langle \rightsquigarrow^{i_m}_{q_1} \rangle \cdots \langle \rightsquigarrow^{i_m}_{q_n} \rangle \langle +\chi^{i_m}_{\top} \rangle \Big) \varphi$$

[6] Note how this operation is not the implicit observation of before; different from it, a restriction lacks a precondition.

with $q_1, \ldots, q_n$ the atomic propositions occurring in χ. Note that once the *restriction* operation $\mathrm{Ex}_j\,\chi!!$ has taken place, the rest of the operations can be performed in any order, yielding exactly the same model. They can even be performed at the same time, suggesting the idea of *parallel* model operations that, though interesting, will not be pursued here.

4.4.1 Some properties of the operations

Our three operations behave as expected, witness the following proposition.

Proposition 4.14

- *The formula* $[\rightsquigarrow^j_q]\,\mathrm{Aw}_j\,q$ *is valid in the general class of* $\mathbf{M}$*-models: after* $\rightsquigarrow^j_q$*, agent* j *is aware of* q*.*
- *The formula* $[\hookrightarrow^j_\sigma]\,\mathrm{Ex}_j\,\mathsf{cn}(\sigma)$ *is valid in the general class of* $\mathbf{M}$*-models: after* $\hookrightarrow^j_\sigma$*, agent* j *is explicitly informed about* $\mathsf{cn}(\sigma)$*.*
- *For* χ propositional *and any agent* i*,* $[j : \chi!]\,\mathrm{Ex}_i\,\chi$ *is valid in the class of* $\mathbf{M}_K$*-models: after* $j : \chi!$ *any agent* i *is explicitly informed about* χ*.*

Proof. Pick any pointed semantic model (M, w). The first property is straightforward: the operation puts q in the PA_j-set of every world in the model, so in particular $\Box_j\,{}^{[j]}q$ is true at w.

For the second one, we cover the three ingredients for explicit information. After the inference operation the agent is aware of $\mathsf{cn}(\sigma)$ because the precondition of the operation tells us that she was already aware of σ; this gives us $\Box_j\,{}^{[j]}\mathsf{cn}(\sigma)$. Moreover, after the operation, $\mathsf{cn}(\sigma)$ itself is in the A_j-set of every world that already had σ and its premises, so in particular it is in every world R_j-accessible from w since the precondition of the operation requires that σ and its premises were already there; this gives us $\Box_j \mathrm{A}_j\,\mathsf{cn}(\sigma)$. Finally, observe that the $\hookrightarrow^j_\sigma$ operation only affects formulas containing $\mathrm{A}_j\,\mathsf{cn}(\sigma)$; hence, $\mathsf{cn}(\sigma)$ itself cannot be affected. Because of the precondition, we know that $\mathsf{cn}(\sigma)$ holds in every world R_j-accessible from w in the initial model M; then, it is still true at every world R_j-accessible from w *in the resulting model* $M_{\hookrightarrow^j_\sigma}$; this gives us $\Box_j\mathsf{cn}(\sigma)$. Therefore, $\mathrm{Ex}_j\,\mathsf{cn}(\sigma)$ holds at w in $M_{\hookrightarrow^j_\sigma}$.

The third case is also straightforward. The operation guarantees that, after it, $\Box_i({}^{[i]}\chi \wedge \mathrm{A}_i\,\chi)$ will be true at w. Moreover, the new model contains only worlds that satisfy $\mathrm{Ex}_j\,\chi$ in M, and since the relation is reflexive, they also satisfy χ in M. Now, propositional formulas depend just on the valuations, which are not affected by the announcement operation; then the surviving worlds will still satisfy χ in the resulting model $M_{j:\chi!}$. Hence, $\Box_i\chi$ is true at w and therefore we have $\mathrm{Ex}_i\,\chi$ true at w in $M_{j:\chi!}$. ■

The property for announcements cannot be extended to arbitrary χs because of the well-know Moore-type formulas, $p \wedge \neg\Box_i p$, that become false in every world of a model after it has been restricted to those that satisfied it.

It is interesting, though, to note a slight difference between how standard *PAL* and our agents react to announcements in the Moore spirit. In *PAL*, after an announcement of *"p is the case and agent j does not know it"*, $p \wedge \neg\Box_j p$, only p-worlds are left, and therefore $\Box_i p$ is true: every agent i knows that p holds. But in our setting, though an announcement of *"p is the case and agent j does not know it explicitly"*, $p \wedge \neg \mathrm{Ex}_j\, p$, does leave just p-worlds, there is no guarantee that the agents will be informed about *p explicitly*. This is because, though the announcement puts $p \wedge \neg \mathrm{Ex}_j\, p$ in the A_i-set of every world w, nothing guarantees us that p will also be there. Agents may need a further inference step to 'break down' the conjunction and then make p explicit information.

4.4.2 A brief look at a finer form of observation

Before applying the developed framework to our running example, we will spend some words discussing an interesting variant of the removing-worlds operations: our announcements and observations. For simplicity, we will work with observations and we will not deal with the awareness notion. As we have mentioned, our fine-grained framework gives us several possibilities for defining such operations, and we have defined explicit versions that, besides removing the worlds where the observation is false, also add the observed formula to the A-sets of the preserved worlds (Definitions 2.20 and 3.7).

However, as it has been indicated by Hamami (2010b), these definitions presuppose some form of omniscience from the agent. When χ is observed, the agent gets to know that χ is true, and therefore she discards those worlds that she recognizes as $\neg\chi$-worlds. In the omniscient *PAL*, this amounts for the agent to eliminate all worlds where $\neg\chi$ is the case, but in our non-omniscient setting the agent may not acknowledge all the information each possible world provides. In particular, she may not recognize a $\neg\chi$-world as such because she may not acknowledge that $\neg\chi$ is true in that world. Then, intuitively, if the agent does not identify a world as a $\neg\chi$-one, she should not eliminate it when observing χ. So how can we address this issue?

One interesting possibility for a finer observation operation is the following. In Epistemic Logic, the statement *"the agent knows φ"* is represented by the formula $\Box\,\varphi$; then a standard observation of a certain χ removes those worlds that the agent recognizes as $\neg\chi$-worlds, that is, worlds where $\neg\chi$ is the case. In our framework, the statement *"the agent knows explicitly φ"* is represented by the formula $\Box\,(\varphi \wedge \mathrm{A}\,\varphi)$; then a finer explicit observation of a certain χ will remove those worlds that the agent recognizes as $\neg\chi$-worlds, that is, worlds where $\neg\chi \wedge \mathrm{A}\,\neg\chi$ is the case. Again, to deal with our finer knowledge representation,

the operation will add χ to the A-sets of all remaining worlds. This definition is closer to the spirit of our framework.

Note how, with such a definition, though some $\neg\chi$-worlds will be discarded (those the agent recognizes as $\neg\chi$-worlds), some of them will not (those the agent *does not* recognize as $\neg\chi$-worlds), agreeing with our intuition. But now there is another issue: our intuition also tells us that an explicit observation of χ should allow the agent to discard *all* $\neg\chi$ worlds!

The two intuitions are reconciliated through the following observation. A finer explicit observation of χ should definitely allow the agent to eliminate all $\neg\chi$-worlds, but only the ones recognized as such should be eliminated immediately after the observation. The remaining ones can be eliminated *only after they have been identified as* $\neg\chi$*-worlds*. In order to do this, we can introduce a further operation, *contradiction removal*, that discards worlds in which the agent has acknowledged both a formula φ and its negation as true, that is, worlds where $\mathrm{A}\,\varphi \wedge \mathrm{A}\,\neg\varphi$ is the case.

Now we can sketch the full story. After χ is observed, the agent eliminates the $\neg\chi$-worlds she has identified so far. Some $\neg\chi$-worlds will survive, but by adding χ to the A-sets of all the worlds that are not eliminated, the agent acknowledges that χ should be the case in all of them. Then, by further reasoning (e.g., by inference), the agent might recognize a $\neg\chi$-worlds as such, thereby realizing that the world contradicts the previous observation. At this point the *contradiction removal* can be invoked, and then the agent will be able to discard that world, as expected.

4.4.3 The information dynamics of the example

We conclude this chapter by going back to Example 4.1 and the formalization of its underlying information state provided in Example 4.2. The dynamic framework we have introduced allows us to 'press play' to see how the agents interact and how their information evolves as a result of the interaction.

Example 4.3 (Information flow among the jurors) We can formalize the dynamics of in Example 4.1 by singling out six different stages.

Stage 1. Juror H's action of scratching his nose makes A aware of both `mkns` and `gls`. Then, he (A) becomes aware of the three relevant rules (he was already questioning the eyesight of the woman, `esq`), and that is enough to make the rules part of his explicit knowledge. More importantly, he also gets explicit knowledge about `mkns`, since he had that information before but simply disregarded the issue (see Table 4.5 of Example 4.2).

$$\langle \rightsquigarrow^{A}_{\mathtt{mkns}} \rangle \langle \rightsquigarrow^{A}_{\mathtt{gls}} \rangle \Big(\mathrm{Aw}_A\,\mathtt{mkns} \wedge \mathrm{Aw}_A\,\mathtt{gls} \wedge \mathrm{Ex}_A\,(\mathtt{mkns} \Rightarrow \mathtt{gls}) \wedge \mathrm{Ex}_A\,(\mathtt{gls} \Rightarrow \mathtt{esq}) \wedge \mathrm{Ex}_A\,(\mathtt{esq} \Rightarrow \neg\mathtt{glt}) \wedge \mathrm{Ex}_A\,\mathtt{mkns} \Big)$$

Stage 2. Juror A has become aware of $\mathtt{mkns}$ so he can now introduce it to the discussion by announcing the possibility. Since he also knows explicitly that $\mathtt{mkns}$ is the case, he can also announce it, giving explicit knowledge about it to all the members of the Jury.

$$\langle A : \mathrm{Aw}_A\, \mathtt{mkns}!\rangle \Big(\mathrm{Aw}_{JURY}\, \mathtt{mkns} \wedge \mathrm{Ex}_A\, \mathtt{mkns} \wedge \langle A : \mathtt{mkns}!\rangle\, \mathrm{Ex}_{JURY}\, \mathtt{mkns}\Big)$$

Stage 3. In particular, the simple introduction of $\mathtt{mkns}$ to the discussion makes it part of G's explicit knowledge, since he was just unaware of it.

$$\Box_G(\mathtt{mkns} \wedge \mathrm{A}_G\, \mathtt{mkns}) \wedge \neg \mathrm{Aw}_G\, \mathtt{mkns} \wedge \langle A : \mathrm{Aw}_A\, \mathtt{mkns}!\rangle\, \mathrm{Ex}_G\, \mathtt{mkns}$$

Stage 4. Now, A can apply the rule $\mathtt{mkns} \Rightarrow \mathtt{gls}$ since after stage 1 he got explicit knowledge about the rule and its premise. After doing it, he announces the conclusion $\mathtt{gls}$. This very act also makes aware every member of the Jury about the possibility of questioning the eyesight of the witness.

$$\langle \hookrightarrow^{A}_{\mathtt{mkns}\Rightarrow\mathtt{gls}} \rangle \Big(\mathrm{Ex}_A\, \mathtt{gls} \wedge \langle A : \mathtt{gls}!\rangle\, (\mathrm{Ex}_{JURY}\, \mathtt{gls} \wedge \langle \rightsquigarrow^{JURY}_{\mathtt{esq}} \rangle\, \mathrm{Aw}_{JURY}\, \mathtt{esq})\Big)$$

Stage 5. Now aware of $\mathtt{gls}$ and $\mathtt{esq}$, C has explicit knowledge about $\mathtt{gls} \Rightarrow \mathtt{esq}$. Moreover, he knows $\mathtt{gls}$ explicitly from A's announcement. Then he can perform an inference step, announcing after it that $\mathtt{esq}$ is indeed the case.

$$\langle \hookrightarrow^{C}_{\mathtt{gls}\Rightarrow\mathtt{esq}} \rangle \Big(\mathrm{Ex}_C\, \mathtt{esq} \wedge \langle C : \mathtt{esq}!\rangle\, \mathrm{Ex}_{JURY}\, \mathtt{esq}\Big)$$

Stage 6. Finally B, now explicitly knowing $\mathtt{esq} \Rightarrow \neg\mathtt{glt}$ and its premise, draws the last inference and announces the conclusion.

$$\langle \hookrightarrow^{B}_{\mathtt{esq}\Rightarrow\neg\mathtt{glt}} \rangle\, (\mathrm{Ex}_B\, \neg\mathtt{glt} \wedge \langle B : \neg\mathtt{glt}!\rangle\, \mathrm{Ex}_{JURY}\, \neg\mathtt{glt})$$

Stages 1-6 can be compounded in one formula and, given Proposition 4.14, it is not difficult to check that such formula is a logical consequence of the information state formalized in Example 4.2. ◀

4.5 Remarks

In this chapter we have discussed a notion of explicit information that combines two requirements examined separately in the two previous chapters. First, in order to have explicit information, the agent should be aware of that information, and in this chapter we have understood awareness as a language-related notion: being aware of some information means that the agent is able to express that information with her current language. Second, the agent should also acknowledge that the information is indeed true. These two requirements are captured by the two components of each possible world: while the PA-sets provide us with the atomic propositions the agent can use at each possible

Notion	Definition	Model req.
Availability of formulas.	$^{[]}\varphi$	—
Local access to formulas.	$\mathrm{A}\,\varphi$	—
Local access to rules.	$\mathrm{R}\,\rho$	—
Awareness about formulas.	$\Box\, ^{[]}\varphi$	—
Awareness about rules.	$\Box\, ^{[]}\mathsf{tr}(\rho)$	—
Implicit information about formulas.	$\Box\,(^{[]}\varphi \wedge \varphi)$	—
Implicit information about rules.	$\Box\,(^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho))$	—
Explicit information about formulas.	$\Box\,(^{[]}\varphi \wedge \varphi \wedge \mathrm{A}\,\varphi)$	—
Explicit information of rules.	$\Box\,(^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho)$	—
Implicit knowledge about formulas.	$\Box\,(^{[]}\varphi \wedge \varphi)$	Equiv. relation.
Implicit knowledge about rules.	$\Box\,(^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho))$	Equiv. relation.
Explicit knowledge about formulas.	$\Box\,(^{[]}\varphi \wedge \varphi \wedge \mathrm{A}\,\varphi)$	Equiv. relation.
Explicit knowledge about rules.	$\Box\,(^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho)$	Equiv. relation.

Table 4.7: Static notions of information.

world (therefore defining the agent's language at it), the A-sets provides us with the formulas the agent has accepted as true, again at each possible world. Based on combinations of these components, we can define several notions of information; the ones we have worked with are listed in Table 4.7.

There are other notions that correspond to other combinations of access to worlds with access to formulas and propositional availability. As mentioned before, a good intuition about them can be obtained by reading them in terms of what they miss in order to become explicit information. For example, $\Box\,(\varphi \wedge \mathrm{A}\,\varphi)$ expresses that φ is a piece of information that will become explicit when the agent considers the atoms in φ. In other words, φ is information that the agent is not currently paying attention to, i.e., forgotten information.

In the dynamic part, we have 'updated' two of the main informational actions of the previous chapter to the new setting. An act that increase awareness is now given not in terms of adding a specific formula, but in terms of adding an atomic proposition. For the act of inference there has been not an important change; given our definition of awareness as a language-related notion, the fact that the agent has explicit information about a rule and its premises also indicates that she is already aware of the rule's conclusion, and therefore the action representing the rule's application only needs to deal with acknowledging the conclusion as true. Finally, we have also reviewed the act of 'broadcasted

explicit observation' (i.e., an announcement). In the version we have defined, an announcement not only restricts the domain to the worlds where the announcer knows explicitly the announcement, adding the announced formula to the acknowledgement set of every agent: it also makes the hearers aware of the atomic propositions involved. We have also sketched a non-omniscient version of an observation that fits better the spirit of our work. Table 4.8 shows a summary of the properly defined actions.

Action	**Description**
Becoming aware.	The agent becomes aware of an atom (and therefore of all formulas built from it).
Truth-preserving inference.	Turns implicit knowledge into explicit knowledge.
Public announcement.	An agent announces something she knows explicitly to everyone.

Table 4.8: Actions and their effects.

Though the general framework presented in this chapter deals with information that does not need to be true, we have spent extra time reviewing the particular case in which the information has this property. In other words, we have focused on the notion of *knowledge*, for which we have now the implicit and explicit counterparts, as well as two main actions that affect them: observations (announcements) and knowledge-based inference.

But in the introduction of this dissertation we also argued for notions of information that does not need to be true. In fact in many situations, like our "12 Angry Men" example, it is the agents' *beliefs* what are more relevant, rather that their *knowledge*.[7] And once we take beliefs seriously, we should look for a real analysis of finer notions of information in the setting of dynamic logics for acts of *belief revision* that work over epistemic plausibility models (van Benthem 2007; Baltag and Smets 2008). And not only that. It also makes sense to look at how our finer reasoning act of inference behaves in a *beliefs* setting.

[7]Nevertheless, even restricted to the notion of knowledge, our finer representation sheds some light on the small steps that leads to the final result.

Chapter 5
Dynamics of implicit and explicit beliefs

In the previous chapters we have explored the notions of implicit and explicit information and their dynamics, focusing in particular on the cases of implicit and explicit *knowledge*. But in our daily life we usually work with incomplete information, and therefore very few things are completely certain for us. The public transport in Amsterdam is highly reliable and usually behaves according to its time schedule, and nevertheless we cannot say in the absolute sense that we *know* the bus will be at the bus stop on time: many unpredictable factors, like flooded streets, snow, mechanical failure or even a car crash may take place. In fact, if we had to act based only on what we know, we would have very little maneuvering space. Fortunately, our attitudes toward information are more than just 'knowing' and 'not knowing'. Most of our behaviour is leaded not by what we know, but rather by what we *believe*.

This chapter will focus on the study of the notions of implicit and explicit *beliefs*, as well as their dynamics. We will start by recalling two existing frameworks for representing beliefs in a possible worlds models. Then we will present our setting for representing *implicit and explicit* beliefs, discussing briefly some of their properties. We emphasize that, just like in the frameworks of the previous chapters there are several possibilities for defining explicit knowledge, the framework we will work with now offers us several possibilities for defining *explicit* beliefs (e.g., Velázquez-Quesada (2009b)). The one we will use follows the idea of defining an *explicit* notion of information as what is true *and accepted as true* in all worlds relevant for the notion (Chapters 3 and 4).

Once the static framework is settled, we will move on to discuss dynamics of implicit and explicit beliefs. First, we will recall an existing notion of *belief revision* in a *DEL* setting, refining it to put it in harmony with our non-omniscient approach. Then we will move to the new action our non-omniscient agent can perform: inferences that involve not only knowledge but also *beliefs*.

In order to simplify the analysis of this chapter, we will focus on the single-agent case. Moreover, we will assume that our single agent is aware of all atomic propositions of the language, and therefore aware of all formulas in the language generated by it. Nevertheless, we still keep our non-omniscient spirit: though the agent will have full attention, she does not need to recognize as true all the formulas that are so, and therefore her explicit information (in this case her explicit beliefs) does not need to be closed under logical consequence.

5.1 Approaches for representing beliefs

Let us start by reviewing two alternatives for representing beliefs within the possible worlds framework.

5.1.1 The *KD45* approach

The classical approach for modelling *knowledge* in *EL* is to define it as what is true in all the worlds the agent considers possible. To get a proper representation, it is asked for the accessibility relation to be at least *reflexive* (making $\Box\varphi \rightarrow \varphi$ valid: if the agent knows φ, then φ is true), and often to be also *transitive* and *euclidean* (giving the agent full positive and negative introspection).

The idea behind the *KD45* approach for representing beliefs is similar. The notion is again defined as what holds in all the worlds the agent can access from the current one, but now the accessibility relation R is asked to satisfy weaker properties. While knowledge is usually required to be true, beliefs are usually required to be just *consistent*. This is achieved by asking for R to be not reflexive but just *serial*, making the D formula $\neg\Box\bot$ valid. The additional transitivity (*4*) and euclideanity (*5*) give us full introspection.

5.1.2 Plausibility models

But beliefs are different from knowledge. Intuitively, we do not believe something because it is true in all possible situations; we believe it because it is true in the ones we consider most likely to be the case (Grove 1988; Segerberg 2001). This idea suggest that we should add further structure to the worlds that an agent consider possible. They should be given not just by a plain set, like what we get when we consider equivalence relations for representing knowledge; there should be also a *plausibility order* among them, indicating which worlds the agent considers more likely to be the case. This idea has led to the development of variants of possible worlds models (Board 2004; van Benthem 2007), similar to those used for conditional logics (Lewis 1973; Veltman 1985; Lamarre 1991; Boutilier 1994b). Here we recall the models we will use: the *plausibility models* of Baltag and Smets (2008) with small modifications.

The first question we should ask is, which properties should this *plausibility order* satisfy? There are several options. The minimum found in Burgess (1984) and Veltman (1985) are reflexivity and transitivity; some other authors (e.g., Lewis (1973)) impose also connectedness. There is no ideal choice, but we should be sure that the properties of our relation are enough to provide a proper definition of the notions we want to deal with.

The idea behind plausibility models is to define beliefs as what holds in the *most plausible* worlds. If we want consistent beliefs, we need to be sure that this set of maximal worlds is always properly defined. This can be done by asking for the relation to be a locally well-preorder.

Definition 5.1 (Locally well-preorder) Let $M = \langle W, \leq \rangle$ be a possible worlds frame (that is, a possible worlds model minus the atomic valuation) in which the accessibility relation is denoted by $\leq$.

For every world $w \in W$, denote by V_w the *comparability class* of w, that is, the set of worlds $\leq$-comparable to ($\leq$-above, $\leq$-equivalent or $\leq$-below) w:

$$V_w := \{u \mid w \leq u \text{ or } u \leq w\}$$

For every $U \subseteq W$, denote by $\mathrm{Max}_{\leq}(U)$ the set of $\leq$-maximal worlds of U, that is, the set of those worlds in U that are better than all the rest in U:

$$\mathrm{Max}_{\leq}(U) := \{v \in U \mid \text{for all } u \in U, u \leq v\}$$

The accessibility relation $\leq$ is said to be a *locally well-preorder* if and only if it is a preorder (a reflexive and transitive relation) such that, for each comparability class V_w and for every non-empty $U \subseteq V_w$, the set of maximal worlds in U is non-empty, that is, $\mathrm{Max}_{\leq}(U) \neq \emptyset$. ◄

Note how the existence of maximal elements in every $U \subseteq V_w$ implies the already required reflexivity by just taking U as a singleton. Moreover, it also implies *connectedness* inside V_w (*local connectedness*): for any two-worlds set $U = \{w_1, w_2\}$, the $\mathrm{Max}_{\leq}(U) \neq \emptyset$ requirement forces us to have $w_1 \leq w_2$ (so w_2 is the maximal), $w_2 \leq w_1$ (w_1 is the maximal), or both. In particular, if two worlds w_2 and w_3 are more plausible than a given w_1 ($w_1 \leq w_2$ and $w_1 \leq w_3$), then these two worlds should be $\leq$-related ($w_2 \leq w_3$ or $w_3 \leq w_2$ or both). Finally, the requirement also implies that each comparability class is conversely well-founded, since there should be at least one element that is above the rest.

Summarizing, a locally well-preorder over a set W partitions it in one or more comparability classes, each one of them being a connected preorder that has maximal elements. In other words, a locally well-preorder is the same as a *locally connected and conversely well-founded preorder*. In particular, because of local connectedness, the notion of "most plausible" is global inside each comparability class, that is, the maximal worlds in each comparability class are the same from the perspective of any world belonging to it.

Now we can define what a plausibility model is.

Definition 5.2 (Plausibility model) A *plausibility model* is a possible worlds model $M = \langle W, \leq, V \rangle$ in which the accessibility relation, denoted now by $\leq$ and called the *plausibility relation*, is a locally well-preorder over W. ◀

Note how, given a world w, the comparability class V_w actually defines all the worlds the agent cannot distinguish from w. Of course, some worlds in V_w might be less plausible than w (those u for which we have $u \leq w$ and $w \not\leq u$), some might be more plausible (those u for which $w \leq u$ and $u \not\leq w$) and some others even equally-plausible (those satisfying both $w \leq u$ and $u \leq w$). But precisely because of that, the agent cannot rule them out if w were the real one. Then, the *union* of $\leq$ and its converse $\geq$ gives us an equivalence relation (denoted by $\sim$) that corresponds to the agent's *epistemic indistinguishability* (i.e.,comparability) relation.

Before discussing the needed modalities to express beliefs our plausibility models, it is illustrative to justify the properties of the plausibility relation.

As we mentioned, reflexivity and transitivity are the minimal requirements found in the literature. Our first important choice is to allow models with multiple comparability classes instead of a single one, and the reason is that, though we will focus on the single-agent case in which a unique class is enough, considering multiple classes already will make smoother the transition to multi-agent situations, a case that is left for further analysis.

Now, why is it asked for *every subset* U of every comparability class V_w to have maximal elements instead of asking for this requirement just for every V_w? The reason is, again, further developments: though we will work with the notion of plain beliefs, asking for this requirement will allow us future extensions that involve the more general notion of *conditional beliefs*. This notion expresses not what the agent believes about the current situation, but what she would believe it was the case if she would learn that some ψ was true, and plain beliefs can be defined as the particular case in which the condition ψ is the always true $\top$. Now, in conditional beliefs, learning ψ is understood as not considering those worlds where ψ does not hold, similar to what the observation operation does (Definition 1.5). Nevertheless, while this operation allows us to express what *will be* the case after the learning, conditional beliefs describe what *was* the case before the learning took place. This is achieved by looking just at those worlds that satisfy the given ψ in each comparability class, *without discarding the rest of them*. For this definition to work, we need to be sure that any such restriction will yield a set of worlds for which there are maximal elements; hence the requirement.

It is time to review the options we have for expressing the notion of belief. The first possibility is to work with the more general notion of conditional beliefs as a primitive by means of a modality of the form $B^{\psi}\varphi$ (*"the agent believes φ conditionally to ψ"*), and then define the notion of belief as the particular case where the condition ψ is the always true $\top$.

But recall that, because of the properties of our plausibility relation, each comparability class is in fact a connected preorder that has maximal elements. Then, as observed in Stalnaker (2006) and Baltag and Smets (2008), φ is true in the most plausible worlds from w (i.e., the maximal ones in V_w) if and only if w can $\leq$-see a world from which all $\leq$-successors are φ worlds. Hence, we can use a standard modality for the relation $\leq$ and define plain beliefs with it in the following way:

$$B\,\varphi := \langle \leq \rangle\, [\leq]\, \varphi \qquad [1]$$

As a remark, note how the notion of *conditional belief* cannot be defined just in terms of a modality for $\leq$, but it can be defined if we include (1) a universal modality U and a strict plausibility modality $\langle < \rangle$ [2], or (2) a universal modality[3], or (3) a modality for the epistemic indistinguishability relation $\sim$ [4].

5.2 Representing non-omniscient beliefs

Our framework for representing implicit and explicit beliefs combines the ideas used in previous chapters for defining implicit and explicit knowledge, with the just introduced plausibility models.

The language has two components: formulas and rules. Formulas are given by a propositional language extended, first, with formulas of the form $\mathrm{A}\,\varphi$ and $\mathrm{R}\,\rho$, where φ is a formula and ρ a rule (as in Chapters 2 and 4), and second, with modalities $\langle \leq \rangle$ and $\langle \sim \rangle$. Rules, on the other hand, are pairs consisting of a finite set of formulas, the rule's premises, and a single formula, the rule's conclusion (again, as in Chapters 2 and 4). The formal definition is as follows.

Definition 5.3 (Language $\mathcal{L}$) Given a set of atomic propositions $\mathtt{P}$, formulas φ, ψ and rules ρ of the *plausibility-access* language $\mathcal{L}$ are given, respectively, by

$$\varphi ::= p \mid \mathrm{A}\,\varphi \mid \mathrm{R}\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \langle \leq \rangle\, \varphi \mid \langle \sim \rangle\, \varphi$$
$$\rho ::= (\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi)$$

where $p \in \mathtt{P}$. Once again, formulas of the form $\mathrm{A}\,\varphi$ are read as *"the agent has acknowledged (accepted) that formula φ is true"*, and formulas of the form $\mathrm{R}\,\rho$ as *"the agent has acknowledged (accepted) that rule ρ is truth-preserving"*. For the modalities, $\langle \leq \rangle\, \varphi$ is read as *"there is a more plausible world where φ holds"*, and $\langle \sim \rangle\, \varphi$ as *"there is an epistemically indistinguishable world where φ holds"*. Other boolean connectives as well as the universal modalities $[\leq]$ and $[\sim]$ are defined as usual. We denote by $\mathcal{L}_f$ the set of formulas of $\mathcal{L}$, and by $\mathcal{L}_r$ its set of rules. ◄

[1] Note that $[\leq]\,\varphi$ is not adequate since it holds at w when φ is true *in all the worlds that are more plausible than w*, and that includes not only the most plausible ones, but also all those laying between them and w. In fact, $[\leq]\,\varphi$ stands for the so-called notion of *safe belief*.

[2] $B^{\psi}\,\varphi := U\,((\psi \wedge \neg\langle < \rangle\, \psi) \rightarrow \varphi)$; see Girard (2008).

[3] $B^{\psi}\,\varphi := U\,(\psi \rightarrow \langle \leq \rangle\, (\psi \wedge [\leq]\, (\psi \rightarrow \varphi)))$; see van Benthem and Liu (2007).

[4] $B^{\psi}\,\varphi := \langle \sim \rangle\, \psi \rightarrow \langle \sim \rangle\, (\psi \wedge [\leq]\, (\psi \rightarrow \varphi))$; see Boutilier (1994a); Baltag and Smets (2008).

For the semantic model, we extend plausibility models with two functions, indicating the formulas and the rules the agent can access (i.e., has acknowledged as true and truth-preserving, respectively) at each possible world, just like in the semantic models of Chapters 2 and 4.

Definition 5.4 (Plausibility-access model) Let $\mathtt{P}$ be a set of atomic propositions. A *plausibility-access (PA) model* is a tuple $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ where $\langle W, \leq, V\rangle$ is a plausibility model over $\mathtt{P}$ and

- $\mathsf{A} : W \to \wp(\mathcal{L}_f)$ is the *access set function*, indicating the formulas the agent has acknowledged as true (i.e., accepted) at each possible world.
- $\mathsf{R} : W \to \wp(\mathcal{L}_r)$ is the *rule set function*, indicating the rules the agent has acknowledged as truth-preserving (i.e., accepted) at each possible world.

Recall that if two worlds are $\leq$-related (comparable), then in fact they are epistemically indistinguishable. Then, we define the indistinguishability relation $\sim$ as the union of $\leq$ and its converse, that is, $\sim := \leq \cup \geq$. In other words, the agent cannot distinguish between two worlds if and only if she considers one of them more plausible than the other. Note that $\sim$ is different from the *equal plausibility* relation, which is given by the *intersection* between $\leq$ and $\geq$.

A *pointed plausibility-access model* (M, w) is a plausibility-access model with a distinguished world $w \in W$. ◀

Now for the semantic evaluation. The modalities $\langle\leq\rangle$ and $\langle\sim\rangle$ are interpreted via their corresponding relation in the usual way, and formulas of the form $\mathrm{A}\,\varphi$ and $\mathrm{R}\,\rho$ are interpreted with our two extra functions.

Definition 5.5 (Semantic interpretation) Let (M, w) be a pointed *PA* model with $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$. Atomic propositions and boolean operators are interpreted as usual. For the remaining cases,

$$
\begin{array}{lll}
(M,w) \Vdash \mathrm{A}\,\varphi & \text{iff} & \varphi \in \mathsf{A}(w) \\
(M,w) \Vdash \mathrm{R}\,\rho & \text{iff} & \rho \in \mathsf{R}(w) \\
(M,w) \Vdash \langle\leq\rangle\varphi & \text{iff} & \text{there is a } u \in W \text{ such that } w \leq u \text{ and } (M,u) \Vdash \varphi \\
(M,w) \Vdash \langle\sim\rangle\varphi & \text{iff} & \text{there is a } u \in W \text{ such that } w \sim u \text{ and } (M,u) \Vdash \varphi
\end{array}
$$ ◀

In order to characterize syntactically the formulas in $\mathcal{L}_f$ valid in plausibility access models, we follow Theorem 2.5 of Baltag and Smets (2008). The important observation is that a locally well-preorder is a *locally connected and conversely well-founded preorder*. Then, by standard results on canonicity and modal correspondence (Chapter 4 of Blackburn et al. (2001)), the axiom system of Table 5.1 is sound and (weakly) complete for formulas of our language $\mathcal{L}$ *with respect to 'non-standard' plausibility-access models*: those in which $\leq$ is reflexive, transitive and locally connected (axioms $T_{\leq}$, $4_{\leq}$ and *LC*, respectively)

and $\sim$ is the symmetric extension of $\leq$ (axioms $T_\sim$, $4_\sim$, $B_\sim$ and *Inc*). But such models have the *finite model* property with respect to formulas in our language, so completeness with respect to plausibility-access models follows from the fact that every finite strict preorder is conversely well-founded.

Prop	$\vdash \varphi$ for φ a propositional tautology	*MP*	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$
$K_\leq$	$\vdash [\leq](\varphi \to \psi) \to ([\leq]\varphi \to [\leq]\psi)$	$K_\sim$	$\vdash [\sim](\varphi \to \psi) \to ([\sim]\varphi \to [\sim]\psi)$
$Dual_\leq$	$\vdash \langle\leq\rangle\varphi \leftrightarrow \neg[\leq]\neg\varphi$	$Dual_\sim$	$\vdash \langle\sim\rangle\varphi \leftrightarrow \neg[\sim]\neg\varphi$
$Nec_\leq$	If $\vdash \varphi$, then $\vdash [\leq]\varphi$	$Nec_\sim$	If $\vdash \varphi$, then $\vdash [\sim]\varphi$
$T_\leq$	$\vdash [\leq]\varphi \to \varphi$	$T_\sim$	$\vdash [\sim]\varphi \to \varphi$
$4_\leq$	$\vdash [\leq]\varphi \to [\leq][\leq]\varphi$	$4_\sim$	$\vdash [\sim]\varphi \to [\sim][\sim]\varphi$
		$B_\sim$	$\vdash \varphi \to [\sim]\langle\sim\rangle\varphi$
LC	$(\langle\sim\rangle\varphi \wedge \langle\sim\rangle\psi) \to \big(\langle\sim\rangle(\varphi \wedge \langle\leq\rangle\psi) \vee \langle\sim\rangle(\psi \wedge \langle\leq\rangle\varphi)\big)$		
Inc	$\langle\leq\rangle\varphi \to \langle\sim\rangle\varphi$		

Table 5.1: Axiom system for $\mathcal{L}$ with respect to plausibility-access models.

5.2.1 Implicit and explicit beliefs, and their basic properties

It is time to define the notions of implicit and explicit beliefs. Again, just like there are several ways of defining explicit knowledge (the $A\varphi$ of Duc (1995); Jago (2006a); van Benthem (2008c) and our Chapter 2; the $\Box\varphi \wedge A\varphi$ of Fagin and Halpern (1988) and van Ditmarsch and French (2009); the $\Box A\varphi$ of Velázquez-Quesada (2009b); the $\Box(\varphi \wedge A\varphi)$ of our Chapter 3), there are also several possibilities for defining *explicit* beliefs. The definitions we will use in this chapter, shown in Table 5.2 below, combine the ideas for defining beliefs as what is true in the most plausible situations (see the mentioned references in Subsection 5.1) with the ideas for defining notions of explicit information as what is true and has been recognized by the agent as true in the relevant set of worlds (see Chapters 3 and 4).

Definition 5.6 The notions of *implicit* and *explicit belief* about formulas and rules are provided in Table 5.2. In words, the agent believes *implicitly* the formula φ (the rule ρ) if and only if φ ($\mathsf{tr}(\rho)$) is true in the most plausible worlds, and she believes φ (ρ) *explicitly* if, in addition, she acknowledges it as true (truth-preserving) in these 'best' worlds. ◄

But we also have an epistemic indistinguishability relation $\sim$, so we can also define implicit and explicit *knowledge* by using its universal modality, just like we did in Chapters 3 and 4. Table 5.3 shows these definitions once again.

The agent believes *implicitly* the *formula* φ	$B_{\mathrm{Im}}\varphi := \langle\leq\rangle\,[\leq]\,\varphi$
The agent believes *explicitly* the *formula* φ	$B_{\mathrm{Ex}}\varphi := \langle\leq\rangle\,[\leq]\,\big(\varphi \wedge \mathrm{A}\,\varphi\big)$
The agent believes *implicitly* the *rule* ρ	$B_{\mathrm{Im}}\rho := \langle\leq\rangle\,[\leq]\,\mathsf{tr}(\rho)$
The agent believes *explicitly* the *rule* ρ	$B_{\mathrm{Ex}}\rho := \langle\leq\rangle\,[\leq]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$

Table 5.2: Implicit and explicit *beliefs* about formulas and rules.

The agent knows *implicitly* the *formula* φ	$K_{\mathrm{Im}}\varphi := [\sim]\,\varphi$
The agent knows *explicitly* the *formula* φ	$K_{\mathrm{Ex}}\varphi := [\sim]\,\big(\varphi \wedge \mathrm{A}\,\varphi\big)$
The agent knows *implicitly* the *rule* ρ	$K_{\mathrm{Im}}\rho := [\sim]\,\mathsf{tr}(\rho)$
The agent knows *explicitly* the *rule* ρ	$K_{\mathrm{Ex}}\rho := [\sim]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$

Table 5.3: Implicit and explicit *knowledge* about formulas and rules.

The current framework allows us to represent implicit/explicit forms of knowledge/beliefs about formulas and rules. Moreover, the following validities, which follow from the contrapositive of axioms *Inc* and $T_{\leq}$ of Table 5.1 ($[\sim]\,\varphi \to [\leq]\,\varphi$ and $\varphi \to \langle\leq\rangle\,\varphi$, respectively), indicates that implicit and explicit knowledge imply implicit and explicit beliefs, respectively.

For formulas:	$K_{\mathrm{Im}}\varphi \to B_{\mathrm{Im}}\varphi$	For rules:	$K_{\mathrm{Im}}\rho \to B_{\mathrm{Im}}\rho$
	$K_{\mathrm{Ex}}\varphi \to B_{\mathrm{Ex}}\varphi$		$K_{\mathrm{Ex}}\rho \to B_{\mathrm{Ex}}\rho$

Modulo the awareness notion (not considered in this chapter), the just defined notions of implicit and explicit knowledge have the properties stated in Subsection 4.3.3. Let us now review some properties of the notions of implicit/explicit belief. Again, though we will focus on the case of formulas, properties for rules can be obtained in a similar way.

The notions are global Note how the notions of implicit/explicit knowledge/beliefs are global in each comparability class. This is obvious for implicit knowledge because this notion is defined as what it is true *in all the worlds the agent considers possible*, that is, in all the worlds *in the comparability class*. Then, if the agent knows implicitly a given φ, this φ is true in all the worlds of the comparability class, and therefore the agent knows it implicitly in any world in it. Moreover, since explicit knowledge is defined as implicit knowledge plus a requirement *in all epistemically indistinguishable worlds*, the notion is global too. More precisely, we have the following proposition.

Proposition 5.1 *Let (M, w) be a pointed PA model. (1) If $(M, w) \Vdash K_{\mathrm{Im}}\varphi$ then, for all worlds $u \in V_w$, $(M, u) \Vdash K_{\mathrm{Im}}\varphi$. (2) If $(M, w) \Vdash K_{\mathrm{Ex}}\varphi$ then, for all worlds $u \in V_w$, $(M, u) \Vdash K_{\mathrm{Ex}}\varphi$. These two statements are abbreviated in the following validities:*

$$K_{\mathrm{Im}}\varphi \to [\sim] K_{\mathrm{Im}}\varphi \qquad K_{\mathrm{Ex}}\varphi \to [\sim] K_{\mathrm{Ex}}\varphi$$ ■

But the notion of belief is also global inside each comparability class in its implicit and its explicit version. The main reason for this is that each such class is connected, and therefore even if the plausibility order branches at some point, these ramifications should converge to a topmost layer that exists because the relation is conversely well-founded.

Proposition 5.2 *Let (M, w) be a pointed PA model. (1) If $(M, w) \Vdash B_{\mathrm{Im}}\varphi$ then, for all worlds $u \in V_w$, $(M, u) \Vdash B_{\mathrm{Im}}\varphi$. (2) If $(M, w) \Vdash B_{\mathrm{Ex}}\varphi$ then, for all worlds $u \in V_w$, $(M, u) \Vdash B_{\mathrm{Ex}}\varphi$. Again, these two statements correspond to the following validities:*

$$B_{\mathrm{Im}}\varphi \to [\sim] B_{\mathrm{Im}}\varphi \qquad B_{\mathrm{Ex}}\varphi \to [\sim] B_{\mathrm{Ex}}\varphi$$ ■

Basic properties First, explicit beliefs are obviously implicit beliefs.

Proposition 5.3 *If φ is explicitly believed, then it is also implicitly believed, that is, the following formula is valid in PA models:*

$$B_{\mathrm{Ex}}\varphi \to B_{\mathrm{Im}}\varphi$$ ■

But, different from implicit and explicit *knowledge*, and though $\leq$ is reflexive, neither implicit nor explicit beliefs have to be true. The reason is that the real world does not need to be among the most plausible ones.

Fact 5.1 *The formula $B_{\mathrm{Ex}}\varphi \wedge \neg\varphi$ is satisfiable in PA models.* ■

Nevertheless, reflexivity makes implicit (hence explicit) beliefs consistent.

Proposition 5.4 *Implicit and explicit beliefs are consistent, that is, the following formula is valid in PA models:*

$$\neg B_{\mathrm{Im}}\bot$$

Proof. The validity can be derived with the axiom system from $\top \to \langle\leq\rangle \top$ (an instance of the contrapositive of $T_{\leq}$), *Prop*, *MP*, $Nec_{\leq}$ and then two applications of instances of $Dual_{\leq}$ and *MP*. ■

Omniscience Implicit beliefs are omniscient.

Proposition 5.5 *All logical validities are implicitly believed and, moreover, implicit beliefs are closed under logical consequence. This gives us the following:*

- *If φ is valid, then $B_{\mathrm{Im}}\varphi$.*
- $B_{\mathrm{Im}}(\varphi \to \psi) \to (B_{\mathrm{Im}}\varphi \to B_{\mathrm{Im}}\psi)$.

Proof. The argument for these statements is simple. For the first, if φ is valid, then it is true in every world of every model; in particular, it is true in the most plausible worlds from any world in any model. For the second, if the most plausible worlds satisfy both $\varphi \to \psi$ and φ, then they also satisfy ψ. ■

But, again, explicit beliefs do not need to have these properties because the A-sets do not need to have any closure property. Nothing forces the A-sets to contain all validities, and having φ and $\varphi \to \psi$ does not guarantee to have ψ.

Introspection Now let us review the introspection properties. First, implicit beliefs are positively introspective.

Proposition 5.6 *In PA-models, implicit beliefs have the* positive *introspection property, that is, the following formula is valid:*

$$B_{\mathrm{Im}}\varphi \to B_{\mathrm{Im}}B_{\mathrm{Im}}\varphi$$

Proof. Here is a derivation:

$$\begin{array}{lcll} B_{\mathrm{Im}}\varphi & \to & \langle\leq\rangle\,[\leq]\,\varphi & \text{by definition} \\ & \to & \langle\leq\rangle\,[\leq]\,[\leq]\,[\leq]\,\varphi & \text{by two applications of } 4_{\leq} \\ & \to & \langle\leq\rangle\,[\leq]\,\langle\leq\rangle\,[\leq]\,\varphi & \text{by } [\leq]\,\varphi \to \langle\leq\rangle\,\varphi\text{, derivable from } T_{\leq} \\ & \to & B_{\mathrm{Im}}B_{\mathrm{Im}}\varphi & \text{by definition} \end{array}$$ ■

They are also negatively introspective.

Proposition 5.7 *In PA-models, implicit beliefs have the* negative *introspection property, that is, the following formula is valid:*

$$\neg B_{\mathrm{Im}}\varphi \to B_{\mathrm{Im}}\neg B_{\mathrm{Im}}\varphi$$

Proof. Here is a derivation:

$$\begin{array}{lcll} \neg B_{\mathrm{Im}}\varphi & \to & \neg\langle\leq\rangle\,[\leq]\,\varphi & \text{by definition} \\ & \to & [\leq]\,\neg[\leq]\,\varphi & \text{by } \mathit{Dual}_{\leq} \\ & \to & [\leq]\,[\leq]\,[\leq]\,\neg[\leq]\,\varphi & \text{by two applications of } 4_{\leq} \\ & \to & \langle\leq\rangle\,[\leq]\,[\leq]\,\neg[\leq]\,\varphi & \text{by } [\leq]\,\varphi \to \langle\leq\rangle\,\varphi \\ & \to & \langle\leq\rangle\,[\leq]\,\neg\langle\leq\rangle\,[\leq]\,\varphi & \text{by } \mathit{Dual}_{\leq} \\ & \to & B_{\mathrm{Im}}\neg B_{\mathrm{Im}}\varphi & \text{by definition} \end{array}$$ ■

Explicit beliefs do not have these properties in the general case, again because the A-sets do not need to have any closure property. Nevertheless, we can get introspection by asking for additional requirements, like we did in Subsection 4.3.3 for the knowledge case.

For positive introspection, we need that if the agent has acknowledged that φ is true, then she has also acknowledged that she believes explicitly in it.

Proposition 5.8 *In PA models in which* $\mathrm{A}\,\varphi \to \mathrm{A}\,B_{\mathrm{Ex}}\varphi$ *is valid, explicit beliefs have the positive introspection property, that is, the following formula is valid:*

$$B_{\mathrm{Ex}}\varphi \to B_{\mathrm{Ex}}B_{\mathrm{Ex}}\varphi$$

Proof. Here is a derivation:

$$\begin{array}{rcll}
B_{\mathrm{Ex}}\varphi & \to & \langle\leq\rangle\,[\leq]\,(\varphi \wedge \mathrm{A}\,\varphi) & \text{by definition} \\
& \to & \langle\leq\rangle\,[\leq]\,[\leq]\,(\varphi \wedge \mathrm{A}\,\varphi \wedge \mathrm{A}\,\varphi) & \text{by } 4_{\leq} \text{ and } \mathit{Prop} \\
& \to & \langle\leq\rangle\,[\leq]\Big([\leq]\,(\varphi \wedge \mathrm{A}\,\varphi) \wedge [\leq]\,\mathrm{A}\,\varphi\Big) & \text{by dist. of } [\leq] \text{ over } \wedge \\
& \to & \langle\leq\rangle\,[\leq]\Big([\leq]\,[\leq]\,(\varphi \wedge \mathrm{A}\,\varphi) \wedge \mathrm{A}\,\varphi\Big) & \text{by } 4_{\leq} \text{ and } T_{\leq} \\
& \to & \langle\leq\rangle\,[\leq]\Big(\langle\leq\rangle\,[\leq]\,(\varphi \wedge \mathrm{A}\,\varphi) \wedge \mathrm{A}\,\varphi\Big) & \text{by } [\leq]\,\varphi \to \langle\leq\rangle\,\varphi \\
& \to & \langle\leq\rangle\,[\leq]\Big(B_{\mathrm{Ex}}\varphi \wedge \mathrm{A}\,\varphi\Big) & \text{by definition} \\
& \to & \langle\leq\rangle\,[\leq]\Big(B_{\mathrm{Ex}}\varphi \wedge \mathrm{A}\,B_{\mathrm{Ex}}\varphi\Big) & \text{by the assumption} \\
& \to & B_{\mathrm{Ex}}B_{\mathrm{Ex}}\varphi & \text{by definition}
\end{array}$$

■

For negative introspection, explicit belief about a given φ may fail because φ is not even implicitly believed, or because the agent has not acknowledged φ in the most plausible worlds. If φ is indeed an implicit belief, then explicit belief is negatively introspective if the agent acknowledges $\neg B_{\mathrm{Ex}}\varphi$ in all the best worlds every time she does not acknowledge φ in all of them.

Proposition 5.9 *In PA models in which* $\neg B_{\mathrm{Im}}\mathrm{A}\,\varphi \to B_{\mathrm{Im}}\mathrm{A}\,\neg B_{\mathrm{Ex}}\varphi$ *is valid, the following formula is valid:*

$$(\neg B_{\mathrm{Ex}}\varphi \wedge B_{\mathrm{Im}}\varphi) \to B_{\mathrm{Ex}}\neg B_{\mathrm{Ex}}\varphi$$

Proof. By using the definitions, $Dual_{\leq}$ and distributing $[\leq]$ over $\wedge$, the implication's antecedent becomes $\langle\leq\rangle\,[\leq]\,\varphi \wedge \neg\langle\leq\rangle\,[\leq]\,\mathrm{A}\,\varphi$, its right side being $\neg B_{\mathrm{Im}}\mathrm{A}\,\varphi$. Then, the extra assumption gives us $B_{\mathrm{Im}}\mathrm{A}\,\neg B_{\mathrm{Ex}}\varphi$. But Proposition 5.10 below takes us from the left conjunct of the antecedent to $B_{\mathrm{Im}}\neg B_{\mathrm{Ex}}\varphi$. Then we have $B_{\mathrm{Im}}\neg B_{\mathrm{Ex}}\varphi \wedge B_{\mathrm{Im}}\mathrm{A}\,\neg B_{\mathrm{Ex}}\varphi$, i.e., $B_{\mathrm{Im}}(\neg B_{\mathrm{Ex}}\varphi \wedge \mathrm{A}\,\neg B_{\mathrm{Ex}}\varphi)$, which abbreviates as $B_{\mathrm{Ex}}\neg B_{\mathrm{Ex}}\varphi$. ■

Nevertheless, though in the general case explicit beliefs do not have neither positive nor negative introspection, they do have them in a weak form.

Proposition 5.10 *The following formulas are valid in PA-models:*

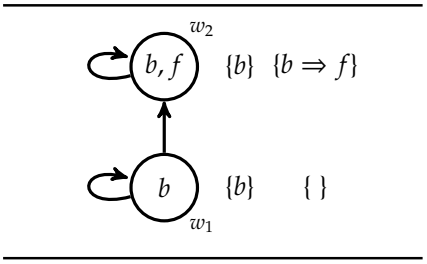
$$B_{\mathrm{Ex}}\varphi \to B_{\mathrm{Im}}B_{\mathrm{Ex}}\varphi \qquad \text{and} \qquad \neg B_{\mathrm{Ex}}\varphi \to B_{\mathrm{Im}}\neg B_{\mathrm{Ex}}\varphi$$

Proof. Similar to the proofs of Propositions 5.6 and 5.7. ■

5.2.2 An example

We close the definition of the static framework for beliefs with a simple example.

Example 5.1 Consider the following plausibility-access model where the A- and the R-set of each world appears to their right in that order. The atomic propositions b and f have the reading *"Chilly Willy is a bird"* and *"Chilly Willy flies"*, respectively. In the model, (1) the agent knows explicitly that Chilly Willy is a bird and, moreover, she believes explicitly the rule stating that if it is a bird, then it flies. Nevertheless, (2) though she believes implicitly that Chilly Willy flies, (3) this belief is not explicit. All this is indicated by the formulas on the right. Since these notions are global inside each comparability class, we do not refer to some particular evaluation point.



(1) $K_{\mathrm{Ex}}\, b \;\wedge\; B_{\mathrm{Ex}}(b \Rightarrow f)$

(2) $B_{\mathrm{Im}}\, f$

(3) $\neg B_{\mathrm{Ex}}\, f$

◄

After defining a framework for representing implicit and explicit forms of beliefs, we will now turn our attention to processes that transform them.

5.3 Belief revision

The first operation that we will review is the one that corresponds to *belief revision*, an action that occurs when an agent's beliefs change in order to incorporate new external information in a consistent way (Gärdenfors 1992; Gärdenfors and Rott 1995; Williams and Rott 2001; Rott 2001). The study of this process and its properties can be traced back to the early 1980s, with the seminal work of Alchourrón et al. (1985) considered to mark the birth of the field.

Traditionally, there have been two approaches to study belief revision. The first one, what we could call the *postulational approach*, analyzes belief change

without committing to any fixed mechanism, proposing instead abstract general principles that a "rational" belief revision process should satisfy. Most of the initial work in the field follows this approach, with the so called *AGM* theory (Alchourrón et al. 1985) being the most representative one. In most of these proposals, an agent's beliefs are represented by a set of formulas closed under logical consequence (i.e., a complete *theory*), and in all of them three are the most relevant operations: (1) *expansion* of beliefs with a given χ, consisting technically in adding χ to the set of formulas and then closing it under logical consequence; (2) *contracting* the beliefs with respect to χ, consisting in removing some formulas such that the closure under logical consequence of the resulting set does not contain χ; (3) *revising* the beliefs with respect to χ, consisting in contracting with $\neg\chi$ and then expanding with χ.

On the other hand, some works have approached belief revision from a more constructive way, presenting concrete mechanisms that change the agent's beliefs. Among these approaches we can mention the epistemic entrenchment functions of Gärdenfors and Makinson (1988): an ordering among formulas that indicates how strong is the agent's belief about them, and therefore provide a way to encode factors that determine which beliefs should be discarded when revising with respect to a given χ. More interesting are the approaches that represent beliefs in a different way, like Grove (1988) which uses a structure called *a system of spheres* (based on the earlier work of Lewis (1973)) to construct revision functions. Like an epistemic entrenchment, a system of spheres is essentially a preorder, but now the ordered objects are no longer formulas, but complete theories.

On its most basic form, belief revision involves an agent with her beliefs, and study the way these beliefs change when new information appears. Then, it is very natural to look for a belief revision approach within the *DEL* framework. Here we review briefly the main idea behind the most relevant proposals.

5.3.1 The *DEL* approach

The main idea behind plausibility models is that the set of worlds the agent considers possible has in fact a further internal structure: an order indicating how plausible each possible world is. Then, an agent believes what is true in the most plausible worlds, i.e., those she considers more likely to be the case.

Now here is the key idea (van Ditmarsch 2005; van Benthem 2007; Baltag and Smets 2008): if beliefs are represented by a plausibility order, then changes in beliefs can be represented by changes in this order. In particular, the act of *revising* beliefs in order to accept χ can be seen as an operation that puts χ-worlds at the top of the plausibility order. Of course, there are several ways in which such a new order can be defined, but each one of them can be seen as a different policy for *revising* beliefs. Here is one of the many possibilities.

Definition 5.7 (Upgrade operation) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model and let χ be a formula in $\mathcal{L}_f$. The *upgrade* operation produces the *PA* model $M_{\chi\Uparrow} = \langle W, \leq', V, \mathsf{A}, \mathsf{R}\rangle$, differing from M just in the plausibility order, given by

$$\leq' := \underbrace{(\leq; \chi?)}_{(1)} \cup \underbrace{(\neg\chi?; \leq)}_{(2)} \cup \underbrace{(\neg\chi?; \sim; \chi?)}_{(3)} \qquad \blacktriangleleft$$

The new plausibility relation is given in a *PDL* style. It states that, after an upgrade with χ, *"all χ-worlds become more plausible than all $\neg\chi$-worlds, and within the two zones, the old ordering remains"* (van Benthem 2007). More precisely, in $M_{\chi\Uparrow}$ we will have $w \leq' u$ if and only if in M (1) $w \leq u$ and u is a χ-world, or (2) w is a $\neg\chi$-world and $w \leq u$, or (3) $w \sim u$, w is a $\neg\chi$-world and u is a χ-world. Again, there are many other definitions for a new plausibility relation that put χ-worlds at the top (e.g., van Benthem (2007); van Eijck and Wang (2008)). The presented one, so-called *radical upgrade*, only shows one of many options.

But not all relation-changing operations are technically adequate. We are interested in those that preserve the required model properties, and therefore keep us in the relevant class of models. In our case, we are interested in operations that do yield a locally well-preorder.

Proposition 5.11 *If M is a PA model, so is $M_{\chi\Uparrow}$.*

Proof. We need to show that if $\leq$ is a locally well-preorder, so is $\leq'$. The proof can be found in Appendix A.9. ■

Note the effect of the upgrade operation on the agent's beliefs. It does not affect the A-sets, so we cannot expect for it to create *explicit* beliefs about χ, since nothing guarantees that χ will be present in the most plausible worlds.

But even if we modify the definition to force χ to be present, the operation puts on top those worlds that satisfy χ *in the original model M* (if there are none, the plausibility order will stay the same), but these worlds do not need to satisfy χ *in the resulting model $M_{\chi\Uparrow}$*. In other words, an upgrade with χ does not necessarily make the agent believe in χ, even *implicity*; this is because, besides beliefs about facts, our agent also has high-order beliefs, that is, beliefs about beliefs and so on. Since the plausibility relation changes, the agent's beliefs change, and so her beliefs about beliefs. This corresponds to the well-known Moore-like sentences (*"p is the case and the agent does not know it"*) in *Public Announcement Logic* (Plaza 1989; Gerbrandy 1999) that become false after being announced, and therefore cannot be known by the agent.

Nevertheless, the operation makes the agent believe *implicitly* in χ if χ is a propositional formula. An upgrade does not change valuations, so if χ is purely propositional, the operation will put on top those worlds that satisfy χ in the original model M, and these worlds will still satisfy χ in the resulting model $M_{\chi\Uparrow}$. Then, the agent will believe χ *implicitly*.

In order to represent this operation within the language, we add the existential modality $\langle \chi \Uparrow \rangle$, with its universal version defined as its dual in the standard way. Formulas of the form $\langle \chi \Uparrow \rangle \varphi$ are read as *"it is possible for the agent to upgrade her beliefs with χ in such a way that after doing it φ is the case"*, with their semantic interpretation given in the following way.

Definition 5.8 (Semantic interpretation) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R} \rangle$ be a *PA* model and χ a formula in $\mathcal{L}_f$. Then,

$$(M, w) \Vdash \langle \chi \Uparrow \rangle \varphi \quad \text{iff} \quad (M_{\chi\Uparrow}, w) \Vdash \varphi$$

Note how the upgrade operation is a total function: it can always be executed (there is no precondition[5]) and it always yields one and only one model. Then, the semantic interpretation of the universal upgrade modality collapses to

$$(M, w) \Vdash [\chi \Uparrow] \varphi \quad \text{iff} \quad (M_{\chi\Uparrow}, w) \Vdash \varphi \qquad \blacktriangleleft$$

Finally, in order to provide a sound and complete axiom system for the language with the new modality, we use reduction axioms once again.

Theorem 5.1 (Reduction axioms for the upgrade modality) *The valid formulas of the language $\mathcal{L}_f$ plus the upgrade modality in PA models are exactly those provable by the axioms and rules for the static base language (Table 5.1) plus the reduction axioms and modal inference rules listed in Table 5.4.* ■

$\Uparrow_p \ \vdash \langle \chi \Uparrow \rangle p \leftrightarrow p$	$\Uparrow_{\mathrm{A}} \vdash \langle \chi \Uparrow \rangle \mathrm{A}\varphi \leftrightarrow \mathrm{A}\varphi$
$\Uparrow_\neg \ \vdash \langle \chi \Uparrow \rangle \neg\varphi \leftrightarrow \neg\langle \chi \Uparrow \rangle \varphi$	$\Uparrow_{\mathrm{R}} \vdash \langle \chi \Uparrow \rangle \mathrm{R}\rho \leftrightarrow \mathrm{R}\rho$
$\Uparrow_\vee \ \vdash \langle \chi \Uparrow \rangle (\varphi \vee \psi) \leftrightarrow \big(\langle \chi \Uparrow \rangle \varphi \vee \langle \chi \Uparrow \rangle \psi\big)$	
$\Uparrow_{\langle\leq\rangle} \vdash \langle \chi \Uparrow \rangle \langle\leq\rangle \varphi \leftrightarrow \langle\leq\rangle\big(\chi \wedge \langle \chi \Uparrow \rangle \varphi\big) \vee \big(\neg\chi \wedge \langle\leq\rangle \langle \chi \Uparrow \rangle \varphi\big) \vee \big(\neg\chi \wedge \langle\sim\rangle (\chi \wedge \langle \chi \Uparrow \rangle \varphi)\big)$	
$\Uparrow_{\langle\sim\rangle} \vdash \langle \chi \Uparrow \rangle \langle\sim\rangle \varphi \leftrightarrow \langle\sim\rangle \langle \chi \Uparrow \rangle \varphi$	
$\Uparrow_N$ If $\vdash \varphi$, then $\vdash [\chi \Uparrow] \varphi$	

Table 5.4: Axioms and rule for the upgrade modality.

Atomic valuation and access and rule sets are not affected by an upgrade, and the reduction axioms for atomic propositions, access and rule formulas reflect this. The reduction axioms for negation and disjunction are standard (in the case of negation recall that the operation does not have precondition), and those for the indistinguishability modality $\langle\sim\rangle$ reflects the fact that the operation

[5]It can be argued that for the agent to upgrade her beliefs with respect to χ, she needs to consider χ possible. This corresponds to $\langle\sim\rangle \chi$ as precondition for the operation.

just changes the order *within* each comparability class, so two worlds will be comparable after an upgrade if and only if they were comparable before.

The interesting axiom is the one for the plausibility modality $\langle \leq \rangle$. It is obtained with techniques from van Benthem and Liu (2007): if the new relation can be defined in terms of the original one with by means of a *PDL*-expression, $\leq' := \alpha(\leq)$, then in the new model a world w can $\leq'$-reach a world that satisfies a given φ, $\langle \chi \Uparrow \rangle \langle \leq \rangle \varphi$, if and only if in the original model w could $\alpha(\leq)$-reach a world that will satisfy φ after the operation, $\langle \alpha(\leq) \rangle \langle \chi \Uparrow \rangle \varphi$. Then, if the *PDL* expression $\alpha(\leq)$ does not use iteration, the *PDL* axioms for sequential composition ($\langle a; b \rangle \varphi \leftrightarrow \langle a \rangle \langle b \rangle \varphi$), non-deterministic choice ($\langle a \cup b \rangle \varphi \leftrightarrow (\langle a \rangle \varphi \vee \langle b \rangle \varphi)$) and test ($\langle \chi ? \rangle \varphi \leftrightarrow (\chi \wedge \varphi)$) can be successively applied to the formula $\langle \alpha(\leq) \rangle \langle \chi \Uparrow \rangle \varphi$ until only modalities with the original relation $\leq$ are left. In our particular case, the axiom simply translates the three-cases *PDL* definition of the new plausibility relation: after an upgrade with χ there is a $\leq$-reachable world where φ holds if and only if before the operation (1) there is a $\leq$-reachable χ-world that will become φ after the upgrade, or (2) the current is a $\neg\chi$-world that can $\leq$-reach another that will turn into a φ-one after the operation, or (3) the current is a $\neg\chi$-world that can $\sim$-reach another that is χ and will become φ after the upgrade.

A reduction axiom for the notion of conditional belief in terms of the modalities for $\leq$ and $\sim$ can be also obtained by unfolding the stated definition.

5.3.2 Our non-omniscient case

Recall our discussion about finer observations (Subsection 4.4.2). We emphasized that, in our approach, the agent may not have direct access to all the information each possible world provides. In other words, in general the agent has not only uncertainty about which one is the real world, but also about what holds in each one of them. So even if she considers possible a single world where some φ holds, she may not recognize it as a φ-world because she may not have acknowledged that φ is indeed the case.

We already discussed how this affects the intuitive idea of an *observation*. Now, how does this affect the idea behind an upgrade?

The intuition behind any operation that changes the plausibility relation is that the agent rearranges the worlds according to what holds in each one of them. In particular, in the operation we defined, the agent puts the worlds she recognizes as χ-worlds, that is, those where χ holds, on top of the rest of them, that is, those where $\neg\chi$ holds. But in our non-omniscient setting, the agent may not be able to tell whether a given world satisfies χ or not. Besides the worlds she identifies as χ-worlds, that is, those satisfying $\chi \wedge \mathrm{A}\,\chi$, and the worlds she identifies as $\neg\chi$-worlds, that is, those satisfying $\neg\chi \wedge \mathrm{A}\,\neg\chi$, she can also see χ-uncertain worlds, that is, those that do not satisfy neither $\chi \wedge \mathrm{A}\,\chi$

nor $\neg\chi \wedge \mathrm{A}\neg\chi$. From this perspective, the definition of the new relation is not reasonable anymore, since it assumes that the agent can identify whether χ holds or not in each possible world.

If the intuitive idea behind an upgrade operation with χ is that the agent will put on top those worlds she recognizes as χ-worlds, then a non-omniscient version should reflect this. Define $G_\chi := \chi \wedge \mathrm{A}\chi$. Then,

Definition 5.9 (Non-omniscient upgrade operation) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model and let χ be a formula in $\mathcal{L}_f$. The *non-omniscient upgrade* operation produces the *PA* model $M_{\chi^{+}\Uparrow} = \langle W, \leq', V, \mathsf{A}, \mathsf{R}\rangle$, differing from M just in the plausibility order, given now by

$$\leq' := (\leq; G_\chi?) \cup (\neg G_\chi?; \leq) \cup (\neg G_\chi?; \sim; G_\chi?) \qquad \blacktriangleleft$$

In words, this revision policy states that the agent will put the worlds *she recognizes* as χ-worlds on top of the rest of them, keeping the old ordering between the two zones. Note how, just like there are several definitions for omniscient upgrades that put χ-worlds at the top, there are several definitions for non-omniscient variations that do the same with G_χ-worlds. The one we have provided guarantees that, if χ is propositional, the agent will believe it explicitly after the upgrade, as we will discuss below.

The defined operation differs from its omniscient counterpart just in the 'upgraded' formula, but the structure of the new order is exactly the same. Therefore, this non-omniscient operation preserves *PA* models too.

Syntactically, we have the following.

Definition 5.10 (Semantic interpretation) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model and χ a formula in $\mathcal{L}_f$. Then,

$$(M, w) \Vdash \langle \chi^{+} \Uparrow\rangle \varphi \quad \text{iff} \quad (M_{\chi^{+}\Uparrow}, w) \Vdash \varphi$$

The non-omniscient upgrade operation is a total function too[6], so the semantic interpretation of the universal non-omniscient upgrade modality collapses to

$$(M, w) \Vdash [\chi^{+} \Uparrow] \varphi \quad \text{iff} \quad (M_{\chi^{+}\Uparrow}, w) \Vdash \varphi \qquad \blacktriangleleft$$

Note now the effect of this non-omniscient upgrade operation. Again, we cannot expect for it to create even implicit beliefs about χ, since once that the plausibility order has changed, the worlds that we have put on top, those that satisfied χ in the original model, may not satisfy it anymore.

[6]Now the requirement would be for the agent to consider χ *explicitly* possible. This corresponds to $\langle\sim\rangle(\chi \wedge \mathrm{A}\chi)$ as precondition for this non-omniscient operation.

But consider now the cases in which χ is a propositional formula. A non-omniscient upgrade with χ puts on top of the ordering not those worlds that satisfy χ, but those worlds *the agent recognizes* as χ worlds, that is, worlds satisfying $\chi \wedge \mathsf{A}\,\chi$. Then, since χ is propositional and the A-sets are not affected by the operation, every world satisfying $\chi \wedge \mathsf{A}\,\chi$ in the original model will also satisfy it after the operation. Therefore, after a non-omniscient upgrade, the agent will believe χ not only implicitly, but also *explicitly*.

Finally, for an axiom system, the non-omniscient upgrade modality has exactly the same reduction axioms the upgrade modality has in the cases of atomic propositions, negation, disjunction, indistinguishability modality and access and rule set (Table 5.4). They key reduction axiom, the one for the plausibility relation, is simply the earlier one with G_χ filled in:

$$\langle \chi \Uparrow \rangle \langle \leq \rangle \varphi \;\leftrightarrow\; \langle \leq \rangle \big(G_\chi \wedge \langle \chi \Uparrow \rangle \varphi \big) \vee \big(\neg G_\chi \wedge \langle \leq \rangle \langle \chi \Uparrow \rangle \varphi \big) \vee \big(\neg G_\chi \wedge \langle \sim \rangle (G_\chi \wedge \langle \chi \Uparrow \rangle \varphi) \big)$$

5.4 Belief-based inference

The just discussed action, *upgrade* in its omniscient and non-omniscient versions, was borrowed from standard *DEL*. But our non-omniscient agent can perform actions omniscient agents cannot; in particular, she can perform *inference*. We have already a representation of this act in its *knowledge-based* form; let us review the main ideas behind it before going into the *belief-based* case.

The intuition behind the action of *knowledge-based inference* is that, if the agent *knows explicitly* a rule and all its premises, then an inference will make her *know explicitly* the rule's conclusion. This action has been semantically defined as an operation that adds the rule's conclusion to the A-set of those worlds where the agent has access to the rule and all its premises (Definition 4.9). But since the precondition of the operation is for the agent to know explicitly the rule and its premises, what the operation actually does is to add the rule's conclusion to the A-set of those worlds in which the agent *knows the rule and its premises*.

But take a closer look at the operation. What it actually does is discard those worlds in which the agent knows explicitly the rule and all its premises, and replace them with copies that are almost identical, the only difference being that their A-sets now contain the conclusion of the applied rule. And this is reasonable because, under the assumption that knowledge is true information, knowledge-based inference (inference with a known rule and known premises) is simply *deductive* reasoning: the premises are true and the rule preserves the truth, so the conclusion *should* be true. In fact, knowledge-based inference can be seen as the act of recognizing two things. First, since the applied rule is truth-preserving and its premises are true, its conclusion *must* be true; second, situations where the premises are true and the rule is truth-preserving but the conclusion does not hold *are not possible*.

The case is different when the inference involve beliefs. Consider, for example, a situation in which the premises of a rule are explicitly known, but the rule itself is only explicitly believed. In such cases it is reasonable to consider very likely a situation in which the premises and the conclusion hold. Nevertheless, the agent should not discard a situation where the premises hold but the conclusion fails, and therefore the rule is not truth-preserving. Technically, an operation representing such action should *split* the current possibilities into two. One of them, the most plausible one, standing for the case in which the rule's conclusion is indeed true; the other, the less plausible one, standing for the case in which the rule's conclusion is false and the rule is not truth-preserving.

More generally, an inference that involves beliefs creates new possibilities, and an operation representing it should be faithful to this. But, how to do it? The action models and product update of Baltag et al. (1999), already used in Section 3.6 for multi-agent situations, will be useful once again. This time our proposal will be based in its *plausibility* version.

5.4.1 Plausibility-access action models

Recall the intuition behind the *action models* of Baltag et al. (1999): just as the agent can be uncertain about which one is the real world, she can also be uncertain about which event has taken place. In such situations, her uncertainty about the action can be represented with a model similar to that used for representing her uncertainty about the situation. *Action models* are possible-worlds-like structures in which the agent considers different events as possible, and her uncertainty *after the action* is an even combination of her uncertainty about the situation *before the action* and her uncertainty *about the action*.

This idea has been extended in order to match richer structures that indicate now only the worlds the agent considers possible but also a plausibility order among them. A first approach was made in Aucher (2003), then generalized in van Ditmarsch (2005). But these two works are based in *quantitative* plausibility orders that use plausibility ordinals in order to express degrees of belief. In contrast, the *plausibility action models* of Baltag and Smets (2008) are purely *qualitative*, and therefore provide a more natural extension to be used with the matching plausibility models.

Just like we did in Section 3.6, we will extend these plausibility action models in order to deal with our access and rule sets function. Here is the formal definition, *for the single-agent case*.

Definition 5.11 (Plausibility-access action model) A single agent *plausibility-access (PA) action model* is a tuple $C = \langle E, \leqslant, \mathsf{Pre}, \mathsf{Pos}_\mathsf{A}, \mathsf{Pos}_\mathsf{R} \rangle$ where

- $\langle E, \leqslant, \mathsf{Pre} \rangle$ is a plausibility action model (Baltag and Smets 2008) with E a finite non-empty set of *events*, $\leqslant$ a *plausibility order* on E (with the

same requirements as those for a plausibility order in *PA* models) and $\mathsf{Pre} : E \to \mathcal{L}_f$ a *precondition* function indicating the requirement each event should satisfy in order to take place. This requirement is given in terms of a formula in our language $\mathcal{L}_f$, so it can include not only facts about the real world but also about the agent's implicit/explicit knowledge/beliefs.

- $\mathsf{Pos_A} : (E \times \wp(\mathcal{L}_f)) \to \wp(\mathcal{L}_f)$ is the *new access set* function, indicating the set of formulas the agent will accept after the action according what she accepted before it and the event that has taken place.
- $\mathsf{Pos_R} : (E \times \wp(\mathcal{L}_r)) \to \wp(\mathcal{L}_r)$ is the *new rule set* function, indicating the set of rules the agent will accept after the action according what she accepted before it and the event that has taken place.

This time, we define three new relations: *strict plausibility*, $\prec := \preccurlyeq \cap \overline{\succcurlyeq}$, *equal plausibility*, $\cong := \preccurlyeq \cap \succcurlyeq$, and *epistemic indistinguishability* (i.e., comparability), $\approx := \preccurlyeq \cup \succcurlyeq$. A pointed *PA* action model (C, e) has a distinguished event $e \in E$. ◂

Now, for the definition of the product update, note that both the static and the action model are preorders with further properties. There are two natural ways of building the order of their cartesian product: we can give the priority either the preorder of the static model, or else to that of the action model. The second option, to give priority to the order of the action, is closer to the intended spirit in which it is the *action* the one that will modify the agent's static plausibility order. The formal definition of this case is as follows.

Definition 5.12 (Product update) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model and $C = \langle E, \preccurlyeq, \mathsf{Pre}, \mathsf{Pos_A}, \mathsf{Pos_R}\rangle$ be a *PA* action model. The *product update* operation $\otimes$ yields the *PA* model $M \otimes C = \langle W', \leq', V', \mathsf{A}', \mathsf{R}'\rangle$, given by

- $W' := \left\{(w, e) \in (W \times E) \mid (M, w) \models \mathsf{Pre}(e)\right\}$
- $(w_1, e_1) \leq' (w_2, e_2)$ iff $\left(e_1 \prec e_2 \text{ and } w_1 \sim w_2\right)$ or $\left(e_1 \cong e_2 \text{ and } w_1 \leq w_2\right)$

and, for every $(w, e) \in W'$,

- $V'(w, e) := V(w)$
- $\mathsf{A}'(w, e) := \mathsf{Pos_A}(e, \mathsf{A}(w))$
- $\mathsf{A}'(w, e) := \mathsf{Pos_R}(e, \mathsf{R}(w))$ ◂

Again, the set of worlds of the new plausibility-access model is given by the restricted cartesian product of W and E; a pair (w, e) will be a world in the new model if and only if event e can be executed at world w. Valuations and the access and rule set of the new worlds are also just as before. First, a world in the new model inherits the atomic valuation of its static component, that is,

an atom p holds at (w, e) if and only if p holds at w. Then, the agent's access set at world (w, e) is given by the function Pos_A with the event e and her access set at w as parameters. The case for rule sets is similar.

The important difference is that new plausibility order is now built following the so-called 'action-priority' rule. A world (w_2, e_2) will be more plausible than (w_1, e_1) if and only if either e_2 is *strictly* more plausible than e_1 and w_1, w_2 are already comparable (i.e., *epistemically indistinguishable*), or else e_1, e_2 are *equally plausible* and w_2 is more plausible than w_1.

Observe what our *PA* action models can do. If we define the new access set and new rule set functions as the identity functions (that is, $\mathsf{Pos}_\mathsf{A}(e, X) := X$ and $\mathsf{Pos}_\mathsf{A}(e, Y) := Y$ for all events $e \in E$), then we get a pure plausibility model that can modify the worlds the agent considers possible and the plausibility order among them.

But pure plausibility models cannot modify the model-component that allows us to represent finer notions of information: access and rule sets. Here is precisely where our new access set and new rule set functions, a generalization of the *substitution function* in van Benthem et al. (2006) for representing factual change, come into play. Our *PA* action models can modify the formulas and rules that the agent has acknowledged as true and truth-preserving, respectively, and therefore they can also modify the agent's *explicit* beliefs.

More importantly, the main virtue of a *PA* action model and its product update is not that they can modify the semantic component of our static model on one hand and the syntactic component on the other. They can modify both of them *together*, allowing us to truly represent acts that change not only the situations the agent considers possible, but also what she has acknowledged as true in each one of them, as we will see in Subsection 5.4.2.

It is not hard to verify that *product update* preserves *PA* models.

Proposition 5.12 . *If M is a plausibility-access model and C an plausibility-access action model, then $M \otimes C$ is a plausibility-access model.*

Proof. We need to prove that if the plausibility orders in M and C are locally well-preorders, then so is the plausibility order of $M \otimes C$. The proof can be found in Appendix A.10. ■

In order to express how product updates affect the agents' information, we extend our language with modalities for each pointed plausibility-access action model (C, e), allowing us to build formulas of the form $\langle C, e\rangle \varphi$, whose semantic interpretation is given below.

Definition 5.13 (Semantic interpretation) Let (M, w) be a pointed *PA* model and let (C, e) be a pointed *PA* action model with Pre its precondition function.

$$(M, w) \Vdash \langle C, e\rangle \varphi \quad \text{iff} \quad (M, w) \Vdash \mathsf{Pre}(e) \text{ and } (M \otimes C, (w, e)) \Vdash \varphi \qquad \blacktriangleleft$$

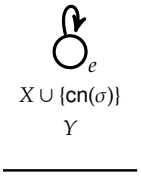
It is now time to introduce the inferences that can be represented with *PA* action models.

5.4.2 Some examples of action models

Just like a public announcement in *PAL* corresponds to a single-event action model (Baltag et al. 1999), the action of knowledge-based inference can be represented with a single-event plausibility-access action model.

Definition 5.14 (Inference with known premises and known rule) Let σ be a rule. The action of *knowledge-based inference*, that is, inference with known premises and known rule, is given by the *PA* action model C^{σ}_{KK} whose definition (left) and diagram showing events, plausibility relation and the way rule and access sets are affected (right) are given by

- $E := \{e\}$
- $\preccurlyeq := \{(e, e)\}$
- $\mathsf{Pre}(e) := \bigwedge_{\psi \in \mathsf{pm}(\sigma)} K_{\mathrm{Ex}}\psi \wedge K_{\mathrm{Ex}}\sigma$
- $\mathsf{Pos}_{\mathsf{A}}(e, X) := X \cup \{\mathsf{cn}(\sigma)\}$
- $\mathsf{Pos}_{\mathsf{R}}(e, Y) := Y$



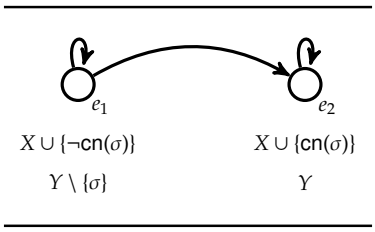
This action model has a single event, with its precondition being for the agent to know explicitly the rule and its premises. In the resulting model, the agent will acknowledge the rule's conclusion in all worlds satisfying the precondition. Moreover, since the premises are true and the rule is truth-preserving in all epistemically indistinguishable ($\sim$-accessible) worlds, the conclusion of the rule must be true in them *in the original model M*. But, just like the inference operation of Definition 4.9, this *PA* action model only affects formulas containing A $\mathsf{cn}(\sigma)$; hence, $\mathsf{cn}(\sigma)$ itself cannot be affected and will still be true in all $\sim'$-accessible worlds in the resulting model $M \otimes C^{\sigma}_{KK}$. Hence, the agent will know explicitly the rule's conclusion. ◀

But our *PA* action models allow us to represent more. Following our previous discussion, here is the action model for inference with known premises and believed rule.

Definition 5.15 (Inference with known premises and believed rule) Let σ be a rule. The action of *inference with known premises and believed rule* is given by the *PA* action model C^{σ}_{KB} whose definition is the following.

- $E := \{e_1, e_2\}$
- $\preccurlyeq := \{(e_1, e_1), (e_1, e_2), (e_2, e_2)\}$
- $\mathsf{Pre}(e_i) := \bigwedge_{\psi \in \mathsf{pm}(\sigma)} K_{\mathrm{Ex}}\psi \wedge B_{\mathrm{Ex}}\sigma$
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}}(e_1, X) := X \cup \{\neg\mathsf{cn}(\sigma)\} \\ \mathsf{Pos}_{\mathsf{A}}(e_2, X) := X \cup \{\mathsf{cn}(\sigma)\} \end{cases}$
- $\begin{cases} \mathsf{Pos}_{\mathsf{R}}(e_1, Y) := Y \setminus \{\sigma\} \\ \mathsf{Pos}_{\mathsf{R}}(e_2, Y) := Y \end{cases}$

The diagram below shows this two-event model. The event on the right, the most plausible one, extends the agent's access set with the rule's conclusion, leaving the rule set intact; it corresponds to the case in which the conclusion of the rule holds. The one on the left, the less plausible one, extends the agent's access set with the *negation* of the rule's conclusion, removing the rule itself from the rule set; it corresponds to the case in which the conclusion of the rule does not hold. In both events the precondition is the same: the agent should know explicitly σ's premises and believe explicitly σ itself.



◀

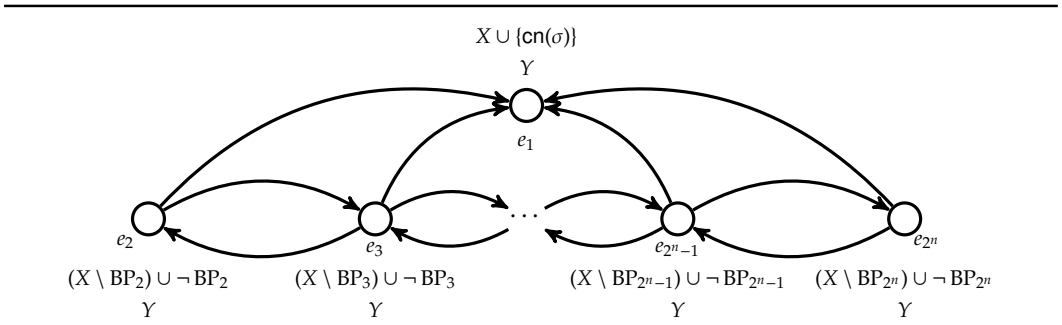
We can also represent a similar situation in which the rule is explicitly known, but one or more of the premises are just explicitly believed. In this case, the best scenario is in which all believed premises are true, but one or more of them may be false, producing an extra number of situations the agent should consider. The following definition provides a model in which all these extra situations are equally plausible, but different orders can be represented.

Definition 5.16 (Inference with believed premises and known rule) Let σ be a rule, and let $\{\psi_1, \ldots, \psi_n\} \subseteq \mathsf{pm}(\sigma)$ be the premises of σ that are believed but not known. Moreover, list as $\mathrm{BP}_2, \ldots, \mathrm{BP}_{2^n}$ each one of the non-empty subsets of $\{\psi_1, \ldots, \psi_n\}$, and denote by $\neg\,\mathrm{BP}_i$ the set that contains the negation of all formulas in BP_i. The action of *inference with believed premises and known rule* is given by the *PA* action model C^{σ}_{BK} whose definition is

- $E := \{e_1, \ldots, e_{2^n}\}$
- $\preccurlyeq \;:=\; \{(e_i, e_1) \mid i = 1, \ldots, 2^n\} \cup \Big((E \setminus \{e_1\}) \times (E \setminus \{e_1\})\Big)$
- $\mathsf{Pre}(e_i) := \bigwedge_{\psi \in \mathsf{pm}(\sigma)} B_{\mathrm{Ex}}\psi \wedge K_{\mathrm{Ex}}\sigma$
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}}(e_1, X) := X \cup \{\mathsf{cn}(\sigma)\} \\ \mathsf{Pos}_{\mathsf{A}}(e_i, X) := (X \setminus \mathrm{BP}_i) \cup \neg\,\mathrm{BP}_i \quad \text{for } i = 2, \ldots, 2^n \end{cases}$
- $\mathsf{Pos}_{\mathsf{R}}(e_i, Y) := Y$

This time the rule is known, but some premises are just believed; then the model has one event for each combination of failing premises. Event e_1 is the one in which no premise fails, and therefore the rule's conclusion is accepted. Events e_2 to e_{2^n} are those in which at least one premise fails, and therefore the

agent rejects them, accepting now their negation. A diagram of this *PA* model, with reflexive and transitive arrows omitted, appears below.



◀

We can even represent a third scenario in which some premises and the rule are just believed. Besides the precondition, this case differs from the previous one (believed premises and known rule) in that there is another possibility: all the premises are indeed true, but the rule is not truth-preserving.

The previous *PA* action models act 'globally' in the sense that they extend the agent's explicit information based on what she has *in a set of relevant worlds*: the epistemically indistinguishable ones when we talk about knowledge and the most plausible ones when we talk about beliefs. But it can also be the case that the agent performs a *local* inference in which she extends what she acknowledges about some particular world based only on the information she has about it. This accounts for situations in which the agent just look at one of the possibilities she considers, and performs inference on it without looking at the rest. Such situations can also be represented with a *PA* action model.

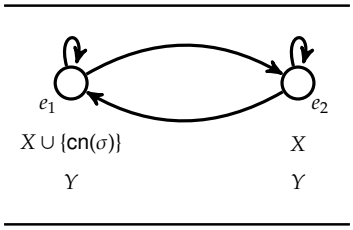
Definition 5.17 (Weak local inference) Let σ be a rule, and define the following abbreviation, stating that the agent has acknowledged a rule σ and its premises

$$\mathrm{Pre}_\sigma := \left(\bigwedge_{\psi \in \mathsf{pm}(\sigma)} \mathrm{A}\,\psi \right) \wedge \mathrm{R}\,\sigma$$

The action of *weak local inference* is given by the following *PA* action model C^σ:

- $E := \{e_1, e_2\}$
- $\preccurlyeq := E \times E$
- $\begin{cases} \mathsf{Pre}(e_1) := \mathrm{Pre}_\sigma \\ \mathsf{Pre}(e_2) := \neg \mathrm{Pre}_\sigma \end{cases}$
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}}(e_1, X) := X \cup \{\mathsf{cn}(\sigma)\} \\ \mathsf{Pos}_{\mathsf{A}}(e_2, X) := X \end{cases}$
- $\mathsf{Pos}_{\mathsf{R}}(e_i, Y) := Y$

With this action the agent works *locally*. Note how any given world satisfies either the precondition of e_1 or the precondition of e_2, but not both. Then, after the operation, we will get a model that differs from the original static one only in that the agent will have accepted the rule's conclusion exactly in those worlds in which she already accepted the rule and its premises.[7] The diagram of this *PA* action model appears below.



◀

A stronger form of local inference can be obtained by strengthening the precondition in the following way.

Definition 5.18 (Strong local inference) Let σ be a rule. The action of *strong local inference* is given by a *PA* action model that differs from the one representing weak local inference only in the definition of the formula Pre_σ, which is now strengthen in the following way:

$$\mathrm{Pre}_\sigma := \Big(\bigwedge_{\psi \in \mathsf{pm}(\sigma)} (\psi \wedge \mathrm{A}\,\psi) \Big) \wedge \Big(\mathsf{tr}(\sigma) \wedge \mathrm{R}\,\sigma \Big)$$

The precondition of event e_1 now requires not only for the agent to accept the rule and the premises, but also for the premises to be true and the rule to be truth-preserving. The access sets of such worlds will be extended with the rule's conclusion, and the rest of the worlds will remain the same. ◀

5.4.3 A further exploration

Plausibility-access action models allow us to represent more than what we have described. We will not go into details (further applications can be found in Chapter 6), but here are some notions that arise in this rich framework.

As observed by many authors, a plausibility relation with the specified properties generates a Grove's system of spheres (Grove 1988), that is, layers of equally-plausible elements with the layers themselves ordered according to their plausibility. The *PA* action models presented so far have at most two layers and in the most plausible one there are at most two events, but we do not have to restrict ourselves to them. With more than two layers we can generalize situations like the case of inference with believed premises and

[7] Note how the effect of a weak local inference can be achieved with the inference operations of previous chapters (Definitions 2.16 and 4.9) by setting the appropriate precondition.

known rule: the events in which at least one of the premises fails do not need to be equally plausible. With more than two events in the top layer we can represent inferences with rules that have more than one conclusion: in the top layer we can have one event for each situation in which one or more of the conclusions are accepted.

We can also classify inferences according to how the events of the action model affect the access sets. For example, *PA* action models in which for every two events e_1, e_2 we have that $e_1 \preccurlyeq e_2$ implies $\mathsf{Pos}_\mathsf{A}(e_1, X) \subseteq \mathsf{Pos}_\mathsf{A}(e_2, X)$ reflect the *optimism* of the agent about the conclusion: events that extend A-sets are more plausible. On the opposite side we have *PA* action models in which for every two events e_1, e_2 we have that $e_1 \preccurlyeq e_2$ implies $\mathsf{Pos}_\mathsf{A}(e_1, X) \supseteq \mathsf{Pos}_\mathsf{A}(e_2, X)$; they reflect the *pessimism* of the agent about the conclusion since events that extend A-sets are less plausible.

Finally, all the inferences we have discussed follow one direction: from the rule and its premises to its conclusion. But a rule can be used in many other ways. For example, the agent can also reason by contraposition: if she knows explicitly a rule and also knows explicitly that the conclusion fails, then she can infer that at least one of the premises fails. And we do not need to stick to deductive reasoning: if the agent knows explicitly a rule and its conclusion, then she can believe explicitly that the premises hold, performing in this way a form of *abductive* reasoning that will be discussed in more detail in Section 6.3.

All in all, *PA* action models are a powerful tool that allow us to represent diverse forms of inference that involve not only an agent's *knowledge* but also her *beliefs*, therefore giving us the possibility to represent not only *truth-preserving* inferences but also *non-truth-preserving* ones (see Chapter 6). Technically, the defined product update works not only on the *semantic* component of the agent's information, like traditional action models do, but also on the *syntactic* component (formulas and rules) we have worked with through this dissertation; this allows us to truly represent acts that change not only the situations the agent considers possible, but also what she has acknowledged as true in each one of them. Thus, our agents are equipped with a broad variety of actions, and with them we can provide a more precise representation of the fine steps that changes our information in real life situations, like we will show in our example of Section 5.5.

5.4.4 Completeness

We have shown how *PA* action models can represent diverse form of inference. Let us now turn to the syntactic characterization of validities involving the *PA* action model modalities. Following the strategy used through this dissertation, we will provide *reduction axioms* for the product update operation.

The reduction axioms for atomic propositions, negation, disjunction and plausibility and indistinguishability modalities of Baltag and Smets (2008) are inherited by our system. But when looking for reduction axioms for access and rule set formulas, the functions Pos_A and Pos_R present a problem. The reason is that they allow the new access and rule sets to be *any* arbitrary set. Let us compare this with other action models and product update definitions for which reduction axioms are provided.

The action models and product update of van Benthem et al. (2006), from which our access-changing functions Pos_A and Pos_R and our product update have evolved, allow us to change the atomic valuation, but the new set of worlds in which each atomic proposition will be true is not arbitrary: it is given by a *formula* of the language. And if we see each formula as a set of worlds (those in which the formula is true), then in fact the new set of worlds in which a given atom p will be true is given in terms of the original one (the set of worlds in which p was true) by means of certain operations: $\neg$ (complement), $\vee$ (union) and so on. But not only that: the static language is already expressive enough to deal with these operations.

Let us look at another definition of an action model and its corresponding product update. The approach of van Eijck and Wang (2008) allows us to change the accessibility relation, but again the new relation is not given in an arbitrary way: it is given in terms of the previous relations by using only *regular (PDL) operations*. And just as the previous case, the static language is already expressive enough to deal with these expressions.

Consider now our product update operation, which extends plausibility action models by allowing us to modify sets of formulas and sets of rules. By looking at the two mentioned cases, we can see that we can provide reduction axioms in the cases in which the definitions of the Pos_A and Pos_R functions are not given arbitrarily, but by means of some structured *expression* that can be already handled in the static language. We will focus on what we will call *set expressions*, and here is our strategy. First, we will extend our static language in order to deal with these expressions, providing not only their semantic interpretation but also the corresponding axioms for them. Then, with the help of these new formulas, we will provide reduction axioms for the class of *PA* action models in which the Pos_A and Pos_R functions are definable by means of these expressions.

As it is currently defined, our static language allows us to look for formulas only at A- and R-sets. What we will do now is to incorporate new formulas that allow us to look not only at these basic sets, but also at more complex ones.

Definition 5.19 (Extended $\mathcal{L}$) Given a set of atomic propositions $\mathtt{P}$, formulas φ, ψ, rules ρ, *set expressions over formulas* Φ, Ψ and *set expressions over rules* Ω, Υ of the *extended plausibility-access language* $\mathcal{L}$ are given, respectively, by

$$\begin{aligned}
\varphi &::= p \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\sim\rangle\varphi \mid \langle\leq\rangle\varphi \mid \lfloor\Phi\rfloor\varphi \mid \lfloor\Omega\rfloor\rho \\
\rho &::= (\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi) \\
\Phi &::= \mathrm{A} \mid \{\varphi\} \mid \overline{\Phi} \mid \Phi \cup \Psi \\
\Omega &::= \mathrm{R} \mid \{\rho\} \mid \overline{\Omega} \mid \Omega \cup \Upsilon
\end{aligned}$$

with p an atomic proposition in P. ◀

Formulas of the form $\mathrm{A}\,\varphi$ have disappeared, leaving their place to formulas of the form $\lfloor\Phi\rfloor\varphi$ where Φ is what we call a *set expression* over formulas. While the $\mathrm{A}\,\varphi$ formulas allowed us to look only at the contents of the A-sets, formulas of the form $\lfloor\Phi\rfloor\varphi$ allow us to look at the content of more complex sets Φ that are built from A and singletons $\{\varphi\}$ by means of complement and union. (The case of set expressions over rules Ω is analogous.) We emphasize that, even though our syntax for set expressions may suggest some strong semantic content, they are just a way of making syntactic comparisons between formulas and between rules, as we will see when analyzing their axiomatization.

The behavior of the new formulas is fixed by their semantic interpretation.

Definition 5.20 (Semantic interpretation) Let (M, w) be a pointed *PA* model with A and R the access and rule sets functions, respectively. The semantic interpretation for the new formulas is given by

$(M,w) \Vdash \lfloor\mathrm{A}\rfloor\varphi$	iff $\varphi \in \mathsf{A}(w)$	$(M,w) \Vdash \lfloor\mathrm{R}\rfloor\rho$	iff $\rho \in \mathsf{R}(w)$
$(M,w) \Vdash \lfloor\{\psi\}\rfloor\varphi$	iff $\varphi = \psi$	$(M,w) \Vdash \lfloor\{\varrho\}\rfloor\rho$	iff ρ is ϱ
$(M,w) \Vdash \lfloor\overline{\Phi}\rfloor\varphi$	iff $\varphi \notin \Phi$	$(M,w) \Vdash \lfloor\overline{\Omega}\rfloor\rho$	iff $\rho \notin \Omega$
$(M,w) \Vdash \lfloor\Phi \cup \Psi\rfloor\varphi$	iff $\varphi \in (\Phi \cup \Psi)$	$(M,w) \Vdash \lfloor\Omega \cup \Upsilon\rfloor\rho$	iff $\rho \in (\Omega \cup \Upsilon)$ ◀

Note, first, how $\lfloor\mathrm{A}\rfloor\varphi$ and $\lfloor\mathrm{R}\rfloor\rho$ are equivalent to the earlier $\mathrm{A}\,\varphi$ and $\mathrm{R}\,\rho$, respectively. Note also how we can even look at the contents of sets built with the intersection and difference operations following the standard definitions:

$$\Phi \cap \Psi := \overline{\overline{\Phi} \cup \overline{\Psi}} \qquad \Phi \setminus \Psi := \Phi \cap \overline{\Psi}$$

The earlier 'static' axiom system is not enough anymore. Though the A- and R-sets still lack any special closure property and there is still no restriction in the way they interact with each other, the additional set expressions have special behaviour, characterized by the following extra axioms.

Theorem 5.2 (Extra axioms for extended $\mathcal{L}$ w.r.t. *PA* models) *The axiom system of Table 5.1, together with the axioms from Table 5.5 is sound and (weakly) complete for the extended language $\mathcal{L}$ with respect to plausibility-access models.* ■

$SE^{\{\}}_{\mathsf{A}} \vdash \lfloor\{\psi\}\rfloor\psi$	$SE^{\{\}}_{\mathsf{R}} \vdash \lfloor\{\varrho\}\rfloor\varrho$
$SE^{\{\}}_{\mathsf{A}} \vdash \neg\lfloor\{\psi\}\rfloor\varphi$ for $\varphi \neq \psi$	$SE^{\{\}}_{\mathsf{R}} \vdash \neg\lfloor\{\varrho\}\rfloor\rho$ for $\rho \neq \varrho$
$SE^{-}_{\mathsf{A}} \vdash \lfloor\overline{\Phi}\rfloor\varphi \leftrightarrow \neg\lfloor\Phi\rfloor\varphi$	$SE^{-}_{\mathsf{R}} \vdash \lfloor\overline{\Omega}\rfloor\rho \leftrightarrow \neg\lfloor\Omega\rfloor\rho$
$SE^{\cup}_{\mathsf{A}} \vdash \lfloor\Phi\cup\Psi\rfloor\varphi \leftrightarrow \big(\lfloor\Phi\rfloor\varphi \vee \lfloor\Psi\rfloor\varphi\big)$	$SE^{\cup}_{\mathsf{R}} \vdash \lfloor\Omega\cup\Upsilon\rfloor\rho \leftrightarrow \big(\lfloor\Omega\rfloor\rho \vee \lfloor\Upsilon\rfloor\rho\big)$

Table 5.5: Axiom system for extended $\mathcal{L}$ w.r.t. plausibility-access models.

The new axioms reflect the behaviour of these sets operations. In the case of set expressions over formulas, axioms $SE^{\{\}}_{\mathsf{A}}$ indicate that ψ and only ψ is an element of $\{\psi\}$. Axiom SE^{-}_{A} says that φ is in the complement of a set if and only if it is not in the set; axiom $SE^{\cup}_{\mathsf{A}}$ says that φ is in the union of two sets if and only if it is in at least one of them. The axioms for set expressions over rules behave in a similar way.

Moreover, the axioms for complement and union actually tell us that $\lfloor\overline{\Phi}\rfloor\varphi$ and $\lfloor\Phi\cup\Psi\rfloor\varphi$ are not really needed, since they can be defined as $\neg\lfloor\Phi\rfloor\varphi$ and $\lfloor\Phi\rfloor\varphi \vee \lfloor\Psi\rfloor\varphi$, respectively. In fact, from the axioms we can see that all we really need are expressions that allow us to verify syntactic identity between formulas on one side, and syntactic identity between rules on the other, like formulas of the form $\lfloor\{\psi\}\rfloor\varphi$ and $\lfloor\{\varrho\}\rfloor\rho$ do (see their axioms). With such extension, our original *PA* language $\mathcal{L}$ (Definition 5.3) is enough for defining these new expressions. Nevertheless, we will keep this 'syntactic sugar' in order to make easier the reading of the formulas and, more importantly, to simplify the reduction axioms that will be provided.

With the extended language it is easy to formulate reduction axioms for *PA* action models that provide the new access and rule sets by means of set expressions. First, we provide a proper definition of this class.

Definition 5.21 (SE-definable *PA* action model) A *set-expression (SE) definable PA action model* is a *PA* action model in which, for each event e, the new access set function $\mathsf{Pos}_{\mathsf{A}}(e)$ is given by a set expression over formulas, and the new rule set function $\mathsf{Pos}_{\mathsf{R}}(e)$ is given by a set expression over rules. ◀

Note how all the *PA* action models presented in Subsection 5.4.2 are *SE* definable. For example, in the action model for inference with known premises and believed rule (Definition 5.15), we have

$$\begin{array}{lll} \text{Event } e_1\text{:} & \mathsf{Pos}_{\mathsf{A}}(e_1) := \mathsf{A} \cup \{\neg\mathsf{cn}(\sigma)\}, & \mathsf{Pos}_{\mathsf{R}}(e_1) := \mathsf{R} \setminus \{\sigma\}. \\ \text{Event } e_2\text{:} & \mathsf{Pos}_{\mathsf{A}}(e_2) := \mathsf{A} \cup \{\mathsf{cn}(\sigma)\}, & \mathsf{Pos}_{\mathsf{R}}(e_2) := \mathsf{R}. \end{array}$$

Now we can provide reduction axioms for the modalities that involve *PA* action models and product update.

Theorem 5.3 *The axiom system built from Tables 5.1, 5.5 and Table 5.6 (with $\top$ and $\bot$ the always true and always false formula, respectively) provide a sound and (weakly) complete axiom system for formulas in the extended language $\mathcal{L}$ plus modalities for action models with respect to PA models and SE-definable PA action models.* ■

$\vdash \langle C,e\rangle p \leftrightarrow \mathsf{Pre}(e) \wedge p$
$\vdash \langle C,e\rangle\neg\varphi \leftrightarrow \mathsf{Pre}(e) \wedge \neg\langle C,e\rangle\varphi$
$\vdash \langle C,e\rangle(\varphi \vee \psi) \leftrightarrow \big(\langle C,e\rangle\varphi \vee \langle C,e\rangle\psi\big)$
$\vdash \langle C,e\rangle\langle\leq\rangle\varphi \leftrightarrow \Big(\mathsf{Pre}(e) \wedge \big(\bigvee_{e\prec e'}\langle\sim\rangle\langle C,e'\rangle\varphi \vee \bigvee_{e\cong e''}\langle\leq\rangle\langle C,e''\rangle\varphi\big)\Big)$
$\vdash \langle C,e\rangle\langle\sim\rangle\varphi \leftrightarrow \Big(\mathsf{Pre}(e) \wedge \bigvee_{e\approx e'}\langle\sim\rangle\langle C,e'\rangle\varphi\Big)$
If $\vdash \varphi$, then $\vdash [C,e]\varphi$
$\vdash \langle C,e\rangle\lfloor \mathrm{A}\rfloor\varphi \leftrightarrow \mathsf{Pre}(e) \wedge \lfloor \mathsf{Pos}_{\mathsf{A}}(e)\rfloor\varphi$
$\vdash \langle C,e\rangle\lfloor\{\psi\}\rfloor\psi \leftrightarrow \mathsf{Pre}(e) \wedge \top$
$\vdash \langle C,e\rangle\lfloor\{\psi\}\rfloor\varphi \leftrightarrow \mathsf{Pre}(e) \wedge \bot$ for $\varphi \neq \psi$
$\vdash \langle C,e\rangle\lfloor\overline{\Psi}\rfloor\varphi \leftrightarrow \langle C,e\rangle\neg\lfloor\Psi\rfloor\varphi$
$\vdash \langle C,e\rangle\lfloor\Psi \cup \Phi\rfloor\varphi \leftrightarrow \langle C,e\rangle(\lfloor\Phi\rfloor\varphi \vee \lfloor\Psi\rfloor\varphi)$
$\vdash \langle C,e\rangle\lfloor \mathrm{R}\rfloor\varphi \leftrightarrow \mathsf{Pre}(e) \wedge \lfloor \mathsf{Pos}_{\mathsf{R}}(e)\rfloor\varphi$
$\vdash \langle C,e\rangle\lfloor\{\varrho\}\rfloor\varrho \leftrightarrow \mathsf{Pre}(e) \wedge \top$
$\vdash \langle C,e\rangle\lfloor\{\varrho\}\rfloor\rho \leftrightarrow \mathsf{Pre}(e) \wedge \bot$ for $\rho \neq \varrho$
$\vdash \langle C,e\rangle\lfloor\overline{\Omega}\rfloor\rho \leftrightarrow \langle C,e\rangle\neg\lfloor\Omega\rfloor\rho$
$\vdash \langle C,e\rangle\lfloor\Omega \cup \Upsilon\rfloor\rho \leftrightarrow \langle C,e\rangle(\lfloor\Omega\rfloor\rho \vee \lfloor\Upsilon\rfloor\rho)$

Table 5.6: Axioms and rules for *SE*-definable action models.

On the first block, the first three axioms are standard: $\langle C,e\rangle$ does not affect atomic valuations, commute with negations (modulo the precondition) and distributes over disjunctions. The fourth, inherited from Baltag and Smets (2008), states that a (C,e) product update after which there is a more plausible φ-world can be performed if and only if the evaluation point satisfies e's precondition, and in the original model there is an *epistemically indistinguishable* world that will satisfy φ after a product update with a *strictly more plausible* e', or there is a more plausible world that will satisfy φ after a product update with an *equally plausible* e''. Finally, the fifth reduction axiom indicates that the comparability class does not change: a (C,e) product update after which there is an epistemically indistinguishable φ-world can be performed if and only if the evaluation point satisfies e's precondition and there is an epistemically indistinguishable world that will satisfy φ after a product update with an *indistinguishable* e'.

The second block contains the axioms for set expressions over formulas, with the first one being the key. After a (C, e) product update, φ will be in the agent's access set if and only if e's precondition is satisfied an φ is in the set expression that defines the new access set at event e:

$$\langle C, e\rangle\lfloor \mathsf{A}\rfloor \varphi \;\leftrightarrow\; \mathsf{Pre}(e) \wedge \lfloor \mathsf{Pos}_{\mathsf{A}}(e)\rfloor \varphi$$

The simplicity of the axiom takes advantage of the fact that our extended $\mathcal{L}$ language can deal with set expressions. As mentioned before, the original language $\mathcal{L}$ plus expressions for syntactic identity is powerful enough to express the membership of a given formula in a set defined from A-sets and singletons by means of complement and union. Then, reduction axioms without set expressions can be provided, but we would need an inductive translation from the expression $\mathsf{Pos}_{\mathsf{A}}(e)$ to the formula that express the membership of φ in it.

The remaining axioms of the second block simply unfold the static axioms for the remaining set-expressions over formulas. The third block, containing axioms for set expressions over rules, behave exactly the same.

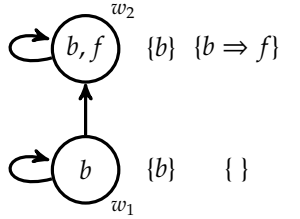
Again, a reduction axiom for the notion of conditional belief in terms of the modalities for $\leq$ and $\sim$ can be obtained by unfolding the stated definition.

5.5 An example in motion

We close this chapter with an example of the operations we have defined.

Example 5.2 Recall the situation of Example 5.1, whose diagram appears below on the left. The agent (1) knows explicitly that Chilly Willy is a bird and believes explicitly that if it is a bird, then it flies. Nevertheless, (2) her belief about Chilly Willy being able to fly is just implicit. The formulas below the diagram express this. We also assume that the rule $f \Rightarrow \neg\neg f$ is present in the R-sets of both worlds, though it will not be indicated in the picture for simplicity.

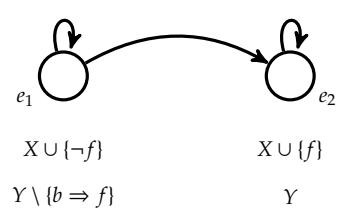
Now the agent decides to use the explicitly believed rule $b \Rightarrow f$, whose premise she knows explicitly. This corresponds to the *PA* action model $C_{KB}^{b\Rightarrow f}$, whose diagram and precondition for both worlds appears on the right.



(1) $K_{\mathrm{Ex}}\, b \;\wedge\; B_{\mathrm{Ex}}(b \Rightarrow f)$

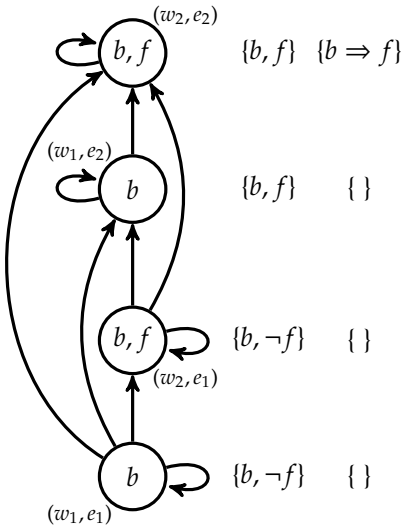
(2) $B_{\mathrm{Im}}\, f \;\wedge\; \neg B_{\mathrm{Ex}}\, f$

$\otimes$



$K_{\mathrm{Ex}} b \;\wedge\; B_{\mathrm{Ex}}(b \Rightarrow f)$

The two worlds of the static model satisfy the precondition of the two events of the action model, so the resulting *PA* model has four worlds, as indicated in the diagram below.



Note how (w_2, e_2) and (w_1, e_2) form a copy of the original static model in which the access sets of the worlds have been extended with the rule's conclusion, according to what event e_2 indicates. The two remaining worlds, (w_2, e_1) and (w_1, e_1), form a copy of the original static model in which the access sets of the worlds have been extended with the *negation* of the rule's conclusion and the rule has been removed from the rule sets, according to what event e_1 indicates. The e_2-partition is above the e_1-partition because e_2 is more plausible than e_1 in the action model.

In the resulting model, (1) the agent still knows explicitly that Chilly Willy is a bird and still believes explicitly the rule stating that if it is a bird then it flies. But now (2) she also believes explicitly that it flies. Nevertheless, the rule she just applied is not known, just believed; then, conscious that the rule might fail, (3) the agent does not know neither explicitly nor implicitly that Chilly Willy flies. In fact, (4) she considers explicitly a possibility in which Chilly Willy does not fly. All this is expressed by the following formulas.

(1) $K_{\mathrm{Ex}}\, b \;\wedge\; B_{\mathrm{Ex}}(b \Rightarrow f)$

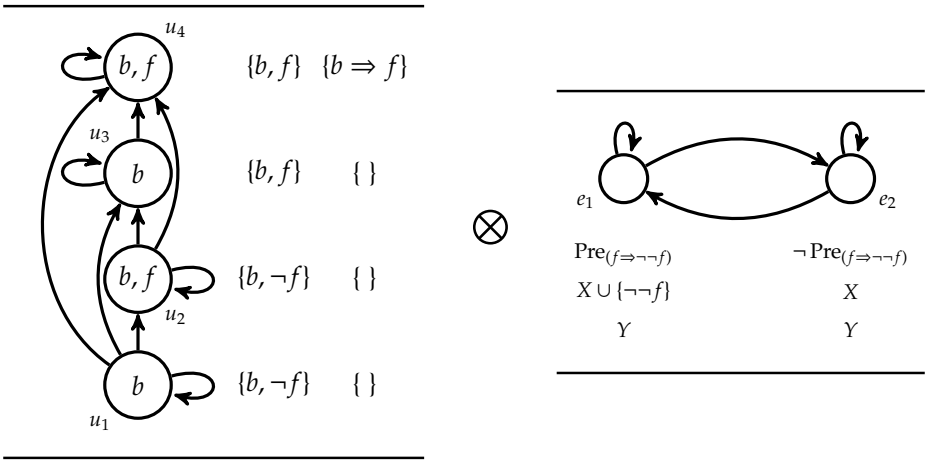
(2) $B_{\mathrm{Ex}}\, f$

(3) $\neg K_{\mathrm{Ex}}\, f \;\wedge\; \neg K_{\mathrm{Im}}\, f$

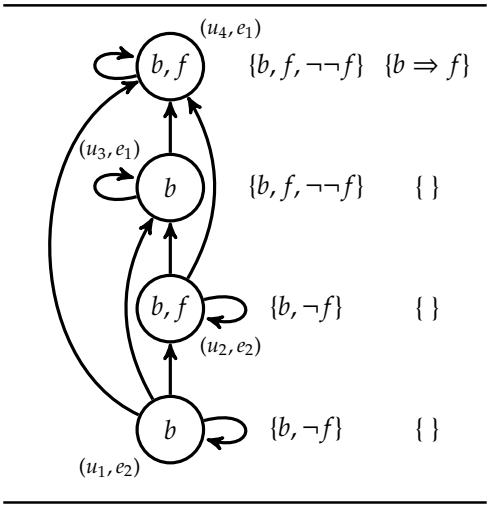
(4) $\widehat{K}_{\mathrm{Ex}}\, \neg f$

where $\widehat{K}_{\mathrm{Ex}}\varphi$ is the 'diamond' version of the explicit knowledge notion $K_{\mathrm{Ex}}\varphi$, that is, $\widehat{K}_{\mathrm{Ex}}\varphi := \langle\sim\rangle\,(\varphi \wedge \mathrm{A}\,\varphi)$.

While waiting for information that confirms or refutes her beliefs, our agent decides to perform weak local inference. She realizes that in the situations in which she has accepted f, she should also accept $\neg\neg f$. This action corresponds to the product update between the previous four-worlds static model (below to the left with worlds renamed) and the *PA* action model $C^{f \Rightarrow \neg\neg f}$ (below to the right). The diagram of the action model includes now the different preconditions for each event, with $\mathrm{Pre}_{(f \Rightarrow \neg\neg f)}$ standing for the formula $\mathrm{A}\, f \;\wedge\; \mathrm{R}\,(f \Rightarrow \neg\neg f)$.

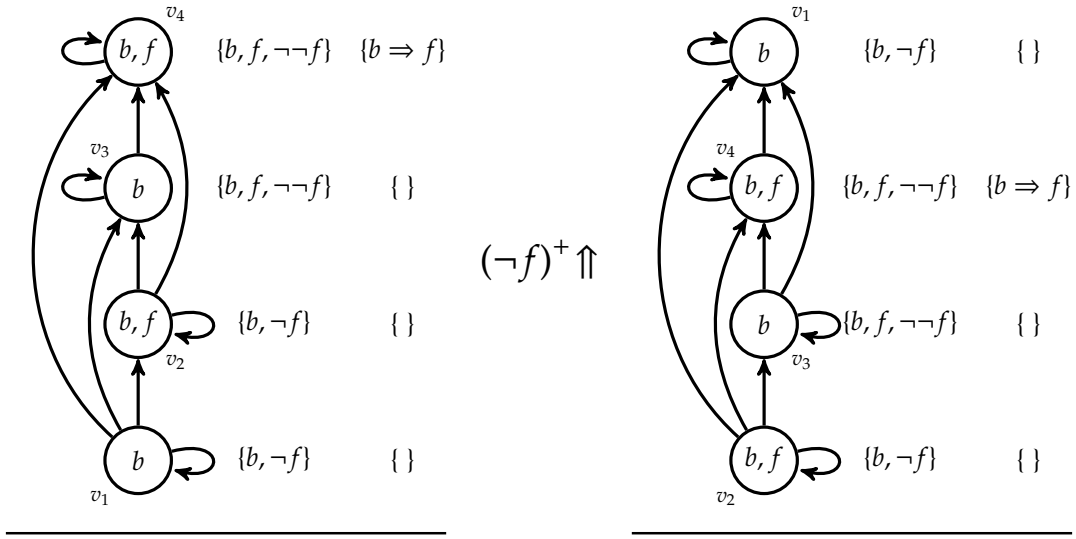


The action simply extends with $\neg\neg f$ those worlds in which the agent has accepted the rule $f \Rightarrow \neg\neg f$ and its premise f. Remember that we have assumed the rule was already present in the R-sets of the initial static model, so it is also in the R-sets of the worlds u_1 to u_4. Since u_3 and u_4 satisfy $\mathrm{Pre}_{(f \Rightarrow \neg\neg f)}$, they will be extended with $\neg\neg f$, following event e_1 of the action model; since u_1 and u_2 satisfy $\neg\, \mathrm{Pre}_{(f \Rightarrow \neg\neg f)}$, they will stay the same, following event e_2. Note how, since the events are equally plausible and their preconditions are complementary, what we get is an exact copy of the static model in which the worlds that satisfy $\mathrm{Pre}_{(f \Rightarrow \neg\neg f)}$ are extended with $\neg\neg f$, and worlds not satisfying it stay the same. The diagram of this resulting model appears below.



Finally, our agent gets new information: a reliable and yet fallible source tells her that in fact Chilly Willy does not fly ($\neg f$). Since the source is fallible, our agent should not discard those worlds where this 'soft' information does not hold; since the source is reliable, she should consider those satisfying the information more likely to be the case.

She can handle this information by *revising* her beliefs with respect to $\neg f$. The operation will put the worlds the agent recognizes as $\neg f$, that is, those satisfying $\neg f \wedge \mathrm{A}\,\neg f$, on top of the rest, keeping the ordering in the two zones as before. In this particular case, the only world the agent recognizes as $\neg f$ is v_1 (formerly called (u_1, e_2)); then, the operation produces the following result.



In the resulting model, (1) the agent still knows explicitly that Chilly Willy is a bird, but now she does not believe (neither explicitly nor implicitly) anymore that if it is a bird then it flies. Moreover, (2) she does not believe (neither explicitly nor implicitly) that it flies; in fact, she believes explicitly that Chilly Willy does not fly ($\neg f$). Nevertheless, she recognizes that this does not need to be the case, and therefore (3) she does not know (neither explicitly nor implicitly) that Chilly Willy does not fly. Actually, (4) she still recognizes explicitly the possibility for it to fly.

(1) $K_{\mathrm{Ex}}\, b \;\wedge\; \neg B_{\mathrm{Ex}}(b \Rightarrow f) \;\wedge\; \neg B_{\mathrm{Im}}(b \Rightarrow f)$
(2) $(\neg B_{\mathrm{Ex}}\, f \wedge \neg B_{\mathrm{Im}}\, f) \;\wedge\; B_{\mathrm{Ex}}\, \neg f$
(3) $\neg K_{\mathrm{Ex}}\, \neg f \;\wedge\; \neg K_{\mathrm{Im}}\, \neg f$
(4) $\widehat{K}_{\mathrm{Ex}}\, f$ ◀

5.6 Remarks

After previous chapters have explored some variations of the notions of implicit and explicit information with particular emphasis on the *knowledge* cases, this chapter has focused on the notions of implicit and explicit *beliefs*.

On the static side, we have provided a representation for implicit and explicit beliefs by combining the ideas for representing non-omniscient agents discussed in the previous chapters, with ideas for representing beliefs in a possible worlds setting (specifically, we have used *plausibility models*). Table 5.7 shows the introduced notions.

Notion	Definition	Model requirements
Implicit belief about formulas.	$\langle\leq\rangle[\leq]\varphi$	$\leq$ a locally well-preorder.
Implicit belief about rules.	$\langle\leq\rangle[\leq]\mathsf{tr}(\rho)$	$\leq$ a locally well-preorder
Explicit belief about formulas.	$\langle\leq\rangle[\leq]\left(\varphi \wedge \mathrm{A}\,\varphi\right)$	$\leq$ a locally well-preorder.
Explicit belief about rules.	$\langle\leq\rangle[\leq]\left(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\right)$	$\leq$ a locally well-preorder.

Table 5.7: Static notions of information.

We have also defined, again, the notions of implicit and explicit knowledge. The definitions are the same as those of Chapter 4 minus the awareness requirement (a notion not considered in this chapter for simplicity). The main difference is that the relation that defines the notion of knowledge, the *epistemic indistinguishability relation* $\sim$, is not a primitive in the model anymore: it is defined as the union of the plausibility order $\leq$ and its converse $\geq$. This states that the agent cannot distinguish between two worlds if she considers one of them more plausible than the other.

On the dynamic side, we have reviewed the existing *DEL* approach for the act of *belief revision*, presenting a variant that is closer to the non-omniscient spirit of our work. But the main part of this chapter has been devoted to the study of inferences that involve beliefs. After arguing why such notion should allow the agent to create new possibilities, we have shown how the combination of existing plausibility action models with the action models that deal with the syntactic components (formulas and rules) of our extended possible worlds model (Section 3.6.2) provide us with a powerful tool that can represent different forms of inference that involve not only knowledge but also beliefs. In particular, we have introduced *PA* action models that represent the actions of inference with known rule and known premises, believed rule and known premises, known rule and believed premises, and strong and weak local inference. Table 5.8 summarizes the actions defined in this chapter.

Thus, our setting provides a very general perspective on the workings of inferences that mix knowledge and belief, far beyond the specifics of particular consequence relations. Though we have provided just some examples of the forms of inferences we can represent, we have by no means exhausted the possibilities. Just like the original action models allow us to represent a broad variety of announcements (public, private, hidden, etc.), our *PA* action models allow us to represent a broad variety of inferences, including some forms that resemble *default* and *abductive* reasoning, as we will discuss in Chapter 6.

There is an important issue in which our *PA* action models can shed some light. The so-called "scandal of deduction" (Hintikka 1973; Sequoiah-Grayson

Action	Description
Upgrade	The agent puts on top of her order those worlds satisfying a given formula.
Non-omniscient upgrade	The agent puts on top of her order those worlds *she recognizes* as satisfying a given formula.
Knowledge-based inference	Inference with explicitly known premises and explicitly known rule. This is truth-preserving inference, that is, deduction.
Belief-based inference	Inference in which the rule or at least one of the premises is not explicitly known, but explicitly believed.

Table 5.8: Actions and their effects.

2008; D'Agostino and Floridi 2009) comes from the idea that deductive reasoning does not provide new information because whatever is concluded was already present in the information given by the premises. In fact, it has been argued that only non truth-preserving inferences can be considered *ampliative* since, if the concluded information is genuinely new, its truth cannot be guaranteed by the old information (Hintikka and Sandu 2007).

In our approach, both truth-preserving and non-truth-preserving inferences are *ampliative*, but in a different sense. Truth-preserving inference (i.e., deduction) is definitely ampliative because, though the agent does not get new implicit knowledge, her explicit knowledge is increased. This has already been recognized by the distinction between *surface* information (our explicit information) and *depth* information (our implicit one). More precisely, we can say that deductive inference is *internally* ampliative because, though it does not change the number of situations the agent considers (no change in implicit information), it does increase the information the agent has about each one of these possibilities. On the other hand, non-truth-preserving inference is ampliative in a different way: it adds more possibilities. More precisely, we can say that non-truth-preserving inference is *externally* ampliative because it increases the number of possibilities the agent considers.

In fact, in our syntactic-semantic setting we can see a nice interplay of four main informational activities. Hard external information, i.e., observations, remove situations the agent considered possible (the observation and announcements operations), but soft external information only rearranges them (the upgrade operations). This already happens in standard omniscient *DEL*, but our non-omniscient setting allows us to represent new actions, the most important one being that of *inference*. Our truth-preserving inference does not remove situations and does not rearrange them; what it does is to extend the informa-

tion the agent has about each one of them. Then, our non-truth-preserving inference is what allows the agent to generate new possible situations.

Finally, through this chapter we have interpreted the A-sets as what the agent has acknowledged as true in each possible world, and for simplicity we have left the notion of awareness out of the picture. By incorporating the notion of awareness in this belief setting, we can provide a richer picture of the agent's attitudes, combining what she believes with what she knows and what she is aware of. In particular, for the notion of awareness, we have now two candidates, the general notion defined in Chapter 3, or the language-based notion of Chapter 4. More importantly, in this chapter we have shown how our extended version of action models can deal with syntactic 'acknowledgement' dynamics, but in Chapter 3 we already showed how a similar structure can deal with syntactic 'awareness' dynamics. Then, just like in the static part, a combination of both can provide us with a richer setting, this time of dynamic actions that affect at the same time what the agent is paying attention to and what she acknowledges as true.

Chapter 6

Connections with other forms of reasoning

The framework developed in the previous chapters has two main virtues. First, it allows us to define finer notions of information; second, and more importantly, it allows us to represent fine informational acts. And, although we have examined non-omniscient versions of the already *DEL*-studied actions of observation and upgrade as well as actions that increase or reduce the agent's awareness, our main focus has been on diverse notions of inference, and we have presented settings for truth-preserving (knowledge-based) and non-truth-preserving (belief-based) inference.

In fact, these two forms of syntactic inference are related with the two semantic informational actions in classical *DEL*. In a purely semantic setting with knowledge and beliefs, we have two kinds of 'incoming information': *hard* knowledge-generating information, that is, observations that remove the worlds in which the incoming observation is not the case, and *soft* belief-generating information, that is, upgrades that do not delete the words where the incoming information does not hold, but nevertheless makes worlds that satisfy it the most plausible ones. Our proposal allows us represent acts of inference, and they also come in a 'hard' and 'soft' flavor: while the 'hard' *truth-preserving* inference of Chapters 2 and 4 makes the agent to accept the conclusion of the applied rule in all the worlds she considers possible, therefore generating explicit knowledge, the 'soft' *non-truth-preserving* inference of Chapter 5 makes the agent to accept the rule's conclusion only in the most plausible worlds, generating in this way only explicit beliefs.

This chapter looks at connections of the acts of inference presented so far with known forms of reasoning. We discuss how our framework relates to deductive, default and abductive reasoning. Then, for belief revision, we first examine the relation between our implicit/explicit beliefs and belief sets/bases, and then review how our setting deals with contradictions.

6.1 Deductive reasoning

The most extensively studied form of reasoning is that of *deductive* reasoning, also known as *valid inference* and *logical* or *classical consequence*. Its characteristic property is that it is a truth-preserving form of reasoning: the conclusion of the inference is true in every single case in which all the premises are true. In other words, when the reasoning is deductive, the truth of the premises guarantee the truth of the conclusion.

This form of reasoning corresponds directly to the truth-preserving forms of inference presented in Chapters 2 and 4. In both of them, the requisite for the application of an inference with a given rule σ is for the agent to know explicitly σ and its premises. We have assumed that knowledge is true information, which in the case of formulas means that they are true, and in the case of rules means that they are truth-preserving, that is, their translation as an implication produces a true formula. From this it follows that the precondition of the operation guarantees that the conclusion of the rule is true and, moreover, implicitly known. Then, the operation only needs to make this knowledge explicit by adding the formula to the corresponding A-sets.

Here is the straightforward translation of deductive reasoning into our setting. Suppose that the following rule σ states a valid inference, that is, if the premises are true, then the conclusion is true.

$$\frac{\psi_1, \ldots, \psi_n}{\varphi}$$

In our setting, this is stated by the following validity (notation of Chapter 4)

$$\Big(\bigwedge_{\psi \in \mathsf{pm}(\sigma)} K_{\mathrm{Ex}} \psi \wedge K_{\mathrm{Ex}} \sigma \Big) \to \langle \hookrightarrow_\sigma \rangle \, K_{\mathrm{Ex}} \mathsf{cn}(\sigma)$$

The main difference is that our setting represents deductive reasoning as a dynamic action that requires not only the rule's premises but also the rule itself. The rule's conclusion is already implicit knowledge, but the agent does not get that information in an explicit form automatically: she should perform a reasoning step.

6.2 Default reasoning

Though reasoning with knowledge is useful in certain areas (e.g., mathematics), most of the information we real agents deal with is not absolutely certain, but only very plausible. Instead of having information stating *"φ is true"*, we usually have information of the form *"φ is plausible"*. A classical situation is the one used in the running example of Chapter 5: birds typically fly.

The question that originates *default reasoning* is, how can we represent this fact? Given the unquestionable information that Chilly Willy is a bird, we would like to have some mechanism that allows us to infer that it flies. But in a deductive (i.e., truth-preserving) approach, the premises would include, besides the *"Chilly Willy is a bird"* requirement, an extra number of them, each one discarding one of the (possibly infinite) reasons for which Chilly Willy might not fly, like being an ostrich, being a penguin, having broken wings and so on. And the problem is not only that we would need to deal with a possibly infinite number of premises, but also that, in order for the agent to derive that Chilly Wily flies, she would need to verify that none of these 'flying-impossibilities' situations holds. More precisely, in order for the agent to derive that Chilly Willy flies, she would need to *know* that it is not an ostrich, it is not a penguin, it does not have broken wings, and so on.

Default reasoning aims to represent this reasoning based on plausible situations. As Reiter states it, "what is required is somehow to allow [Chilly Willy] to fly *by default*" (Reiter 1980). His *default logic* interprets this 'default' as "If [Chilly Willy] is a bird, then in the absence of any information to the contrary, infer that [Chilly Willy] can fly" (Reiter 1980). Following this intuition, he defines a *default rule* as an expression of the following form, where ψ is the *prerequisite* of the rule, each ϕ_i is a *justification*, and χ is the conclusion.

$$\frac{\psi : \phi_1, \ldots, \phi_n}{\chi}$$

The idea of a default rule is that, if the agent has the prerequisite *and the justification is consistent with her information*, then she can accept the conclusion. Usually, the "justification" part of a rule, $\phi_1, \ldots, \phi_n$, is abbreviated as simply χ, so the rule is read as *"if ψ is the case and χ is consistent with the information, then accept the latter"*. For example, with the atomic propositions used in Subsection 5.2.2 (b stands for *"Chilly Willy is a bird"*, and f stands for *"it flies"*), the default rule *"birds typically fly"* is given by a rule with b as prerequisite, f as justification, and f itself as conclusion.

Note how default reasoning is non-monotonic. Though the prerequisite has to be true, the justifications do not need to: they just need to be consistent with the *current* information. Then, further information can invalidate the use of a default rule, and therefore the conclusion may need to be retracted.

Default reasoning and other forms of non-monotonic reasoning have been usually studied from a purely syntactic point of view. The study has been based on "sub-structural" consequence relations, that is, consequence relations that do not satisfy the five structural rules the classical consequence relation satisfies: reflexivity, permutation, contraction, monotonicity and cut (see Subsection 2.4.2). Another approach, closer to the *DEL* spirit of our work, is not to look at

consequence relations with different properties, but rather to consider the different informational attitudes that these reasoning processes involve (Boutilier 1994c; Meyer and van der Hoek 1995). In the case of default logic, Reiter himself already mentioned that the result of inferences with default rules should have the status of a *belief*, subject to change in the light of further information.

Default reasoning as belief upgrade Introducing epistemic notions highlights other possible readings of a default rule. From an epistemic and doxastic point of view, it can be read as *"if the agent* knows *the prerequisite ψ and the justifications $\phi_1, \ldots, \phi_n$ are* consistent with her knowledge *(i.e., she considers $\phi_1, \ldots, \phi_n$ explicitly possible), then after applying the reasoning step she will* believe *χ"*.[1]

The framework for beliefs introduced in Section 5.4 allow us to represent the action described by this new reading. From this perspective, default reasoning can be seen as a change in beliefs that, under the condition that ψ is explicitly known and every ϕ_i is explicitly possible, will modify the agent's plausibility relation in order to put on top those worlds she recognizes as χ-worlds. This reasoning step can be expressed with the formula $\langle \mathrm{Def}_{\chi}^{\psi:\phi_1,\ldots,\phi_n} \rangle \varphi$, defined as

$$\langle \mathrm{Def}_{\chi}^{\psi:\phi_1,\ldots,\phi_n} \rangle \varphi \;:=\; K_{\mathrm{Ex}}\psi \wedge \left(\widehat{K}_{\mathrm{Ex}}\phi_1 \wedge \cdots \wedge \widehat{K}_{\mathrm{Ex}}\phi_n \right) \wedge \langle \chi^{+} \Uparrow \rangle \varphi$$

Thus, default reasoning can be seen as belief upgrade with a specific precondition. In fact, this idea can be already handled in standard *DEL* by dropping the "explicit" part in the precondition and using the omniscient upgrade.

Default reasoning as inference with believed rule But our framework with implicit/explicit knowledge/beliefs about formulas/rules gives us another option. Recall the definition of inference with known premises and believed rule (Definition 5.15): based on the explicit *knowledge* of the premises and the explicit *belief* in the rule, it produces an explicit *belief* in the conclusion of the applied rule, just like what a default rule does. This represents, again, certain form of default reasoning, but the followed strategy is different.

Consider the "birds typically fly" situation. What our setting proposes is that we can work with a rule that concludes flying abilities from bird nature as long as we recognize that this rule works *only in the most plausible situations*. In other words, instead of using a truth-preserving rule whose premises need to discard every single situation in which Chilly Willy might not fly (the *deductive* approach), or using a default rule that ask for the agent information to be consistent with the conclusion (the *default logic* approach), we can use a simple rule of the form "if it is a bird, it flies". But then, different from the deductive and default logic approaches, we do not ask for the rule to be *known*: what we assume is that the rule itself is just *believed*. Following the intuitive effect of such a reasoning step, an inference with this rule should make the agent *believe* that Chilly Willy actually flies. Nevertheless, this conclusion shout not

[1] Similar *dynamic* readings of defaults reasoning steps have been studied in Veltman (1996).

get a *knowledge* status and, in fact, the inference should also make the agent to acknowledge a possibility in which Chilly Willy does not fly. This is exactly what the *PA* action model of Definition 5.15 does.

In general, this second approach to default reasoning proposes the following. Given a default rule of the form described before, its effect can be mimicked by the application of an inference with the rule $\psi \Rightarrow \chi$, provided that ψ is explicitly known and $\psi \Rightarrow \chi$ is explicitly believed, that is,

$$\langle \mathrm{Def}_{\chi}^{\psi:\phi_1,\ldots,\phi_n} \rangle \varphi \;:=\; K_{\mathrm{Ex}}\psi \;\wedge\; B_{\mathrm{Ex}}(\psi \Rightarrow \chi) \;\wedge\; \langle C_{KB}^{\psi \Rightarrow \chi}, e_1 \rangle \varphi$$

where C_{KB}^{σ} is the *PA* action model of Definition 5.15.

But, where have the justifications gone? They are now embedded in the involved notions of information. Intuitively, the justifications are precisely what allow the agent to make the inference, so if any of them fails, the agent should not be able to perform the latter: if Chilly Willy is an ostrich, or a penguin, or has its wings broken, then it does not fly. This says that the agent should believe, at least implicitly, that if any of the justification ϕ_i fails, Chilly Willy does not fly, that is, she should believe, at least implicitly, that for every justification ϕ_i, formulas of the form $\neg\phi_i \rightarrow \neg\chi$ hold. With our notation, this is

$$B_{\mathrm{Im}}(\neg\phi_i \rightarrow \neg\chi) \qquad \text{for each } \phi_i$$

stating that every $\neg\phi_i \rightarrow \neg\chi$ is true the agent's most plausible worlds.

Suppose that indeed some justification ϕ_k fails and the agent knows it explicitly, that is,

$$K_{\mathrm{Ex}}\neg\phi_k$$

This makes $\neg\phi_k$ true in all possible worlds, and given the agent's stated implicit belief, $\neg\chi$ is true in the most plausible ones. Now, if the agent wants to perform an inference step based on a default rule in the just described style, she needs to know explicitly the prerequisite, that is,

$$K_{\mathrm{Ex}}\psi$$

This makes ψ true in all the worlds she considers possible, and then the most plausible ones satisfy ψ, but also $\neg\chi$. But now she cannot believe $\psi \Rightarrow \chi$ neither implicitly nor explicitly, and therefore she cannot apply the inference step.

$$\neg B_{\mathrm{Im}}(\psi \Rightarrow \chi)$$

Even if she does not know that some justification fails and simply believes it explicitly, $B_{\mathrm{Ex}}\neg\phi_k$, this would still make ψ and $\neg\chi$ true in the most plausible situations, so again the inference could not be applied.

By representing default reasoning with an inference based on known premises and believed rule, we do not need to list the justifications anymore; all we need is for the agent to believe implicitly that the failure of any of them will invalidate the inference. Then, it is enough for her to believe explicitly that one justification has failed in order for the inference to be blocked, as expected.

6.3 Abductive reasoning

All the inferences mentioned in Chapter 5 follow one direction: from the rule and its premises to its conclusion. But this is not the only way in which a rule can be used. In fact, one of the most prominent non-monotonic reasoning processes, abductive reasoning, is usually described as 'backwards deduction'.

The process of *abductive reasoning*, introduced into modern logic by Charles S. Peirce (see Aliseda (2006) for a more recent study of the subject), is usually described as the process of looking for an explanation of a given observation, and it has been recognized as one of the most commonly used in our daily activities. Classical examples go from Sherlock Holmes' stories (observing that Mr. Wilson's right cuff is very shiny for five inches and the left one has a smooth patch near the elbow, Holmes assumes that Mr. Wilson has done a considerable amount of writing lately) to medical diagnosis (given the symptoms A and B, a doctor suspects that the patient suffers from C). In Peirce's own words (Hartshorne and Weiss 1934), abduction can be described in the following way:

> The surprising fact χ is observed.
> But if ψ were true, χ would be a matter of course.
> Hence, there is reason to suspect that ψ is true.

Pierce himself did not remain quite convinced that one logical form covers all cases of abductive reasoning (Peirce 1911). Indeed, different kinds of abductive problems arise when we consider agents with different omniscient and reasoning abilities, and even more appear when we combine different notions of information (Soler-Toscano and Velázquez-Quesada 2010). Among all of them, some can be represented with the framework for belief-based inference introduced in Section 5.4.

Intuitively, the idea behind abductive reasoning is the following. The agent observes a fact that cannot be justified by her current information. Then, she looks for an *explanation*: one or several pieces of information that, if true, would make the observation something expected. Consider, for example, the Sherlock Holmes situation. Holmes observes that while Mr. Wilson's right cuff is very shiny, the left one has a patch near the elbow. Then, in order to explain these observations, Holmes assumes that Mr. Wilson has done a considerable amount of writing lately. These assumptions, if knew before, would have allowed him to predict the observations. In words closer to our terminology, Holmes knows a piece of information (the status of Mr. Wilson's cuffs) and he also knows how he could have derived it (If Mr. Wilson has been writing a lot, then his cuffs will have such status). Then, Holmes believes that what fires such derivation could have been the case (he believes that Mr. Wilson has writing a lot lately).

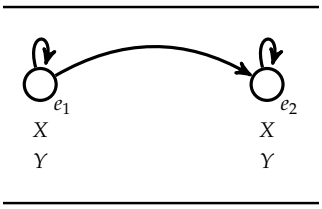
This kind of abductive reasoning can be represented in our setting with a *PA* action model. The idea behind this action model is that this form of abductive

reasoning can be seen as a change in beliefs fired by the agent's inferential tools: if she *knows explicitly* a formula that is the conclusion of an *explicitly known* rule; then, there is reason to *believe explicitly* in the premises of the rule.

Definition 6.1 (Knowledge-based abduction) Let σ be a rule, and recall that for a given formula φ, the worlds the agent recognizes as φ-worlds are those satisfying $G_\varphi := \varphi \wedge \mathrm{A}\,\varphi$. The action of *knowledge-based abduction* (that is, abduction with known rule and known conclusion) is given by the *PA* action model $C_{KK}^{Abd(\sigma)}$ whose definition is the following.

- $E := \{e_1, e_2\}$
- $\preccurlyeq := \{(e_1, e_1), (e_1, e_2), (e_2, e_2)\}$
- $\mathsf{Pre}(e_1) := \Big(K_{\mathrm{Ex}}\sigma \wedge K_{\mathrm{Ex}}\mathsf{cn}(\sigma)\Big) \wedge \neg\Big(\bigwedge_{\psi\in\mathsf{pm}(\sigma)} G_\psi\Big)$
- $\mathsf{Pre}(e_2) := \Big(K_{\mathrm{Ex}}\sigma \wedge K_{\mathrm{Ex}}\mathsf{cn}(\sigma)\Big) \wedge \Big(\bigwedge_{\psi\in\mathsf{pm}(\sigma)} G_\psi\Big)$
- $\begin{cases} \mathsf{Pos}_{\mathsf{A}}(e_1, X) := X \\ \mathsf{Pos}_{\mathsf{A}}(e_2, X) := X \end{cases}$
- $\begin{cases} \mathsf{Pos}_{\mathsf{R}}(e_1, Y) := Y \\ \mathsf{Pos}_{\mathsf{R}}(e_2, Y) := Y \end{cases}$

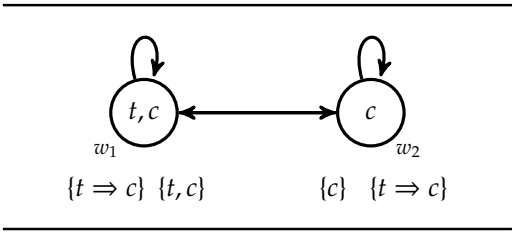
The diagram below shows this two-event model. Note that no event affects neither the formulas nor the rules the agent has accepted; in fact, the only difference between the events, besides their plausibility, is their precondition. The precondition for the most plausible event, e_2, is not only for the agent to know explicitly the rule and its conclusion (what we call the *strong abductive precondition*), but also to recognize each one of the rule's premises as true. The precondition for the least plausible event, e_1, is not only for the agent to know explicitly the rule and its conclusion (the strong abductive precondition), but also for the agent to *not to* recognize all the premises as true. What this action model does is just rearrange the ordering of the worlds. Those that the agent recognizes as satisfying *all premises of the rule* will be on top of the rest, and within the two zones the old ordering will remain.



◀

Here is an example of how this *PA* action model works.

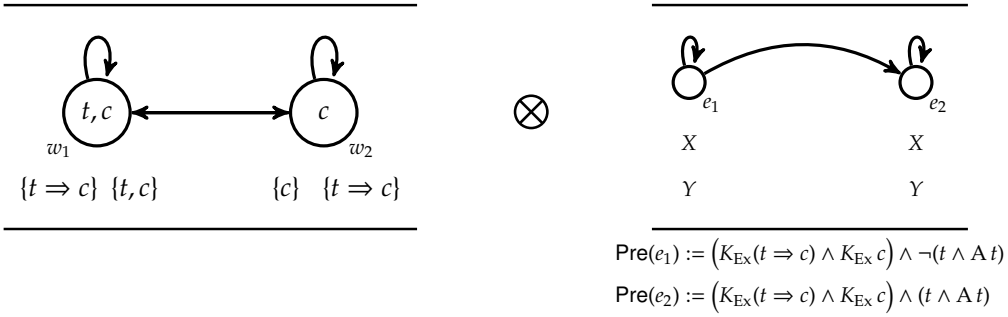
Example 6.1 Consider the static *PA* model below. In it our agent, Sherlock Holmes, (1) knows explicitly the status of Mr. Wilson's cuffs (c), and also knows explicitly that if Mr. Wilson has been doing a considerable amount of writing lately (t), then the status of his cuff's would follow. Nevertheless, (2) Holmes does not believe, neither explicitly nor implicitly, that Mr. Wilson has been writing lately. The formulas on the right of the diagram express all this.



(1) $K_{\mathrm{Ex}}\, c \ \wedge\ K_{\mathrm{Ex}}(t \Rightarrow c)$

(2) $\neg B_{\mathrm{Ex}}\, t \ \wedge\ \neg B_{\mathrm{Im}}\, t$

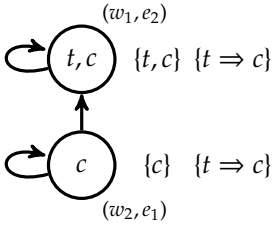
Now Holmes applies abductive reasoning in order to explain the status of Mr. Wilson's cuffs. He knows that if he knew that Mr. Wilson has being doing a considerable amount of writing lately, he would have been able to conclude the observed state of his cuffs. Then, there is reason for Holmes to believe it.



$\mathsf{Pre}(e_1) := \left(K_{\mathrm{Ex}}(t \Rightarrow c) \wedge K_{\mathrm{Ex}}\, c\right) \wedge \neg(t \wedge \mathrm{A}\, t)$

$\mathsf{Pre}(e_2) := \left(K_{\mathrm{Ex}}(t \Rightarrow c) \wedge K_{\mathrm{Ex}}\, c\right) \wedge (t \wedge \mathrm{A}\, t)$

The strong abductive precondition, $K_{\mathrm{Ex}}(t \Rightarrow c) \wedge K_{\mathrm{Ex}}\, c$, is true in both worlds of the *PA* model. Nevertheless, w_1 is the unique world that Holmes recognizes as a t-world; then, only w_1 satisfy e_2's precondition, and only w_2 satisfy e_1's precondition. As consequence, the resulting *PA* model (shown below) has only two worlds, (w_1, e_2) and (w_2, e_1). The components of these new worlds are exactly those of their static counterpart because atomic valuation does not change, and the postcondition functions of the two events do not make any change in the set of formulas and the set of rules the agent accepts. What has changed is the ordering of the worlds: now the unique world that Holmes recognizes as a t-world, w_1, has become more plausible than the rest. In this resulting model (1) Holmes still knows explicitly the status of Mr. Wilson's cuffs, and still knows explicitly the rule that links that with a considerable amount of writing. But, as a result of abductive reasoning, (2) Holmes now believes (both implicitly and explicitly) that the high amount of writing indeed is the case.



(1) $K_{\mathrm{Ex}}\, c \ \wedge\ K_{\mathrm{Ex}}(t \Rightarrow c)$

(2) $B_{\mathrm{Ex}}\, t \ \wedge\ B_{\mathrm{Im}}\, t$

◀

We have represented this form of abductive reasoning as a form of belief change driven by a rule: the agent knows explicitly the rule and its conclusion, so it is reasonable for her to believe explicitly all the premises.

Iterative abduction Abduction is a form of non-monotonic reasoning. The explanations are just hypothesis, and cannot be considered as absolute truth. In other words, abductive reasoning generates beliefs, not knowledge. But then, the form of abductive reasoning we have represented cannot be iterative. Though we require for the rule and its conclusion to be explicitly known, the process produces explicitly believed premises, and then the agent cannot look for an explanation of these premises themselves.

We can represent a form of abduction that allows iteration by weakening the abductive precondition. Instead of asking for the rule and the conclusion to be explicitly known, we can ask for the rule to be explicitly known, but for the conclusion only to be explicitly *believed*. After an abductive step the agent will believe the premises, and then she can look for an explanation for them.

Definition 6.2 (Belief-based abduction) Let σ be a rule, and recall that for any formula φ, the worlds the agent recognize as φ-worlds are those satisfying $G_\varphi := \varphi \wedge \mathrm{A}\,\varphi$. The action of *belief-based abduction* is given by the *PA* action model $C_{KB}^{Abd(\sigma)}$, differing from its knowledge-based counterpart $C_{KK}^{Abd(\sigma)}$ (Definition 6.1) only in the abductive precondition, which now becomes $K_{\mathrm{Ex}}\sigma \wedge B_{\mathrm{Ex}}\mathsf{cn}(\sigma)$. More precisely, the precondition of events e_1 and e_2 are now given by

- $\mathsf{Pre}(e_1) := \Big(K_{\mathrm{Ex}}\sigma \wedge B_{\mathrm{Ex}}\mathsf{cn}(\sigma)\Big) \wedge \neg\Big(\bigwedge_{\psi\in\mathsf{pm}(\sigma)} G_\psi\Big)$
- $\mathsf{Pre}(e_2) := \Big(K_{\mathrm{Ex}}\sigma \wedge B_{\mathrm{Ex}}\mathsf{cn}(\sigma)\Big) \wedge \Big(\bigwedge_{\psi\in\mathsf{pm}(\sigma)} G_\psi\Big)$ ◀

An even weaker notion of abduction can be defined by asking for the rule not to be explicitly known, but simply explicitly believed.

Finding abductive solutions Our approach allows us to represent some forms of abductive reasoning. But typically, works on abductive reasoning focus not only in defining an operation that incorporates the explanation to the agent's information (what a product update with the defined *PA* action models does), but also in *finding* such explanations and then selecting the 'best' of them.

From our perspective, the stage of looking for explanations should be guided by the inferential tools the agent has. If the aim of abductive reasoning is to incorporate hypothesis that, if knew before, would have allowed the agent to derive (i.e., predict) the observation, then it is reasonable to consider as explanations all those pieces of information that would have allowed the derivation. In our example there is only one rule whose conclusion is the observed fact, but in general the agent might have several rules that allow her to derive the observation, and therefore she could choose between many different explanations, each one of them being the premises of such rules.

Now, when looking for a criteria to decide which ones of these possible explanations are 'the best', our framework provides us with some options. Note that we cannot rely on how plausible is the rule that provides the explanation, since our precondition is that the rule is explicitly *known* (in a weaker case, believed), and therefore recognized as true in all possible worlds (in the most plausible ones). One possibility is to rely on the plausibility of the rule's premises: if the agent already believes in some of them, then it is reasonable to believe in the rest. This solution follows the "minimal change" approach, since the reordering needed to believe in all the premises is in principle less complicated than the reordering that would be needed to believe in all the premises when none of them are currently believed.

6.4 Belief bases in belief revision

Coherentism vs foundationalism As we have mentioned, belief revision deals with the different ways an agent's beliefs can change in order to incorporate external information in a consistent way. Classical approaches, like the mentioned *AGM* theory, assume that an agent's beliefs, her *belief set*, are given by a theory: a consistent set of formulas closed under logical consequence. From this perspective, the *coherentist* perspective, there is no distinction between the agent's beliefs: all of them come from the same source, all of them are equally supported, and all of them are equally relevant when they need to be revised.

Nevertheless, it has been argued (Alchourrón and Makinson 1982; Hansson 1989; Fuhrmann 1991; Hansson 1992) that not all beliefs in a belief set have the same status: there is a distinguished class of basic beliefs, the *belief base*, which are somehow given, and from which the rest of the beliefs can be derived by some inference process, typically a truth-preserving one. In the most general case, this belief base is a simple set of formulas that does not need to satisfy any logical constrain, like closure under logical consequence or even consistency. This *foundationalist* approach highlights the process of *inference* through which the agent generates the full belief set from the belief base.

Classical *EL* approaches for representing beliefs (Section 5.1) follow the *coherentist* idea. In the case of the *KD45* approach, the belief set of the agent corresponds to the set of formulas that are true in all the accessible worlds; in the case of the *plausibility models* approach, the belief set corresponds to the set of formulas that are true in the most plausible worlds. In both cases, there is no distinction among the believed formulas: they all have the same status and they all are equally relevant.

On the other hand, our non-omniscient approach to beliefs (Section 5.2) is closer to the *foundationalist* spirit. First, just like in the omniscient case, the

agent's belief set at world w in model M ($\text{BelBas}_{(M,w)}$) can be defined as the set of formulas that are true in the agent's most plausible worlds. In our terminology, these are exactly the formulas the agent believes *implicitly*. This gives us

$$\text{BelSet}_{(M,w)} := \{ \varphi \in \mathcal{L}_f \mid (M, w) \Vdash B_{\text{Im}}\varphi \}$$

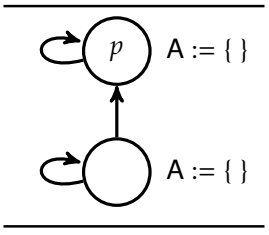
In other words, the formula φ is in the agent's belief set if and only if she believes it implicitly.

Then, the agent's belief base at w in M ($\text{BelBas}_{(M,w)}$) can be defined as those implicit beliefs the agent has acknowledged, that is, her *explicit* beliefs:

$$\text{BelBas}_{(M,w)} := \{ \varphi \in \mathcal{L}_f \mid (M, w) \Vdash B_{\text{Ex}}\varphi \}$$

Despite the similarities, our notion of explicit beliefs do not correspond directly to the notion of belief bases, and there are two main reasons for this.

The first reason is technical: our explicit beliefs do not need to be a set that generates the implicit beliefs. Consider, for example, the following model:



In this extreme situation, the agent does not have any explicit belief, and then the closure under logical consequence of this empty set will only generate the set of validities. But this set does not coincide with the agent's implicit beliefs, which additionally contains p and all its logical consequences. And there is more. Even if the agent acknowledges the truth-value of every atomic proposition in each possible world, there is still no guarantee that *she* can actually derive all the implicit beliefs. Her inferential abilities, that is, the *rules* she can apply, do not need to be *complete* in the sense that they may not be enough to derive all the logical consequences of her explicit information.

Our notions of implicit and explicit beliefs can be put in correspondence with the notions of belief set and belief base if we make these two assumptions:

1. the agent has acknowledged the truth-value of all atomic propositions in each possible world, that is, for all $p \in \mathsf{P}$ and for all $w \in W$,

$$p \in V(w) \text{ implies } p \in \mathsf{A}(w) \qquad \text{and} \qquad p \notin V(w) \text{ implies } \neg p \in \mathsf{A}(w)$$

2. the agent has complete reasoning abilities; in other words, if something is an implicit belief, then there is a finite sequence of reasoning steps

(i.e., rule applications) after which the belief will be explicit. This can be expressed with the formula

$$B_{\mathrm{Im}}\varphi \rightarrow \langle * \rangle B_{\mathrm{Ex}}\varphi$$

where the modality $\langle * \rangle$ stands for the reflexive and transitive closure of the application of inference steps.

The second reason is more conceptual, and highlights the top-down perspective of our framework. In the foundationalist approach, it is the *belief base* the one that is given; then the belief set is built by successive inference steps until we reach a stable situation in which no further step will add further information. But our notions of implicit and explicit beliefs, and in general our notions of implicit and explicit information follow the other direction. It is the *implicit* form the one that is given, usually by what is true in all the relevant worlds (the epistemically indistinguishable in the case of knowledge, the most plausible ones in the case of beliefs). Then, among the pieces of implicit information, we distinguish the ones that the agent has recognized and acknowledged; those are the *explicit* ones.

6.5 Dealing with contradictions

We have discussed connections of our framework with known forms of nonmonotonic reasoning, arguing that we can represent some of their forms by dealing explicitly with the weaker notion of information they involve: *beliefs*.

Now, when beliefs are considered, there is the possibility for the agent to have incorrect information and therefore to face contradictions. Let us revise which options our framework provides for dealing with such situations.

In general, an agent can face two different forms of contradiction.

External contradictions An agent can face a contradiction between her information and some external source. The typical belief revision case falls into this category: the agent believes that χ holds and then an external source suggests her that $\neg\chi$ is the case. There are also other possibilities, according to how strong is the agent's attitude towards χ (known or just believed) and how reliable is the external observation (infallible or just plausible). In out setting, the case in which the agent knows χ and gets informed that $\neg\chi$ certainly holds cannot happen, because χ cannot be both true and false at the same time (we have assumed true knowledge). But, putting this case aside, any of the other three situations is possible.

The way the agent deals with such contradictions depends on which one is the strongest: the agent's information or the observation. If the agent knows χ, then being suggested that $\neg\chi$ is the case will not affect neither her knowledge

nor her beliefs. On the other hand, if the agent believes χ, then having an irrefutable proof that $\neg\chi$ holds will make change her knowledge and hence also her beliefs (what an explicit observation does). Finally, in the case in which the agent believes χ and she gets informed from a reliable but fallible source that $\neg\chi$ is the case, the needed action depends on the reliability of the external source. If the external source is more reliable than the agent's beliefs, then we are in the typical belief revision case, which is solved by a change in the agent's beliefs (what a *DEL* upgrade and its non-omniscient version do) to agree with the external source. If, on the other hand, the agent's beliefs are more reliable, then there will be no change. Note how in the cases in which an action is needed, our setting has an operation that represents it.

Internal contradictions A more serious form of contradiction arises when the contradiction occurs *inside* the agent's information, that is, when the agent is informed (implicitly or explicitly) about both a formula and its negation.

In our setting of Chapter 5, the agent cannot have internal contradictions in her implicit/explicit knowledge/beliefs. For the case of beliefs, recall that the plausibility relation is a locally well-preorder, so inside each comparability class there are always maximal worlds; hence $B_{\mathrm{Im}}\varphi \wedge B_{\mathrm{Im}}\neg\varphi$ is not satisfiable, and therefore neither is $B_{\mathrm{Ex}}\varphi \wedge B_{\mathrm{Ex}}\neg\varphi$. For the case of knowledge, the indistinguishability relation is reflexive (because the plausibility relation is reflexive); hence $K_{\mathrm{Im}}\varphi \wedge K_{\mathrm{Im}}\neg\varphi$ is not satisfiable and therefore neither is $K_{\mathrm{Ex}}\varphi \wedge K_{\mathrm{Ex}}\neg\varphi$.

Note how implicit/explicit knowledge/beliefs cannot face internal contradictions because of semantic restrictions: the plausibility relations always have maximal worlds inside each comparability class and the indistinguishability relation is reflexive. Then, every single time there is at least one maximal world and at least one epistemically possible; hence the implicit forms of knowledge and belief are contradiction-free, and therefore so are their explicit forms.

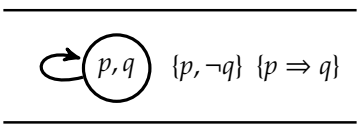
But we do not have any restriction for the formulas the agent has in her access sets.[2] Then, weaker notions of information that look only at the contents of such sets can face internal contradictions. In particular, our agent can consider as possible worlds in which she has acknowledged the truth of both a formula and its negation, that is, formulas of the form $\langle\sim\rangle(\mathrm{A}\,\varphi \wedge \mathrm{A}\,\neg\varphi)$ are satisfiable.

The question is now, how can our agent deal with these situations? In Section 4.4.2 we proposed to remove the world in which such contradiction occurs. The intuition behind the suggestion is that in a pure knowledge setting, that is, true observations and truth-preserving inference, the only reason such situation can occur is because the agent has observed that χ holds, and then has found out (via inference) that one of the possibilities she still considers is in fact a $\neg\chi$-one. Then discarding such possibility is simply the delayed effect of the previous observation.

[2] In the framework of Chapter 2 we ask for the formulas to be true, but we dropped this requisite in the next chapters when we changed our definition of explicit information.

In a setting that involves beliefs it is also reasonable to eliminate such contradictory possibilities if the beliefs have been built in a proper way. When beliefs are involved, the inferential acts we have proposed take care of generating the possibilities in which the assumptions fail. For example, our definition of inference with known premises but believed rule not only makes the agent to believe explicitly that the rule's conclusion is the case; it also generates an explicit possibility in which the conclusion is not. Hence, if there is an epistemically possible world in which the agent has acknowledged χ and then she truthfully and explicitly observes $\neg\chi$, there should be a copy of the world in which she has not acknowledged χ (and in fact she has acknowledged $\neg\chi$).

But the "removing the syntactically inconsistent world" proposal is not an option in situations like the following one:



In this model, the agent knows explicitly p and $p \Rightarrow q$. Then, she can perform an inference (in fact, deductive) step that makes her know q explicitly. There is no contradiction at the level of explicit knowledge, since the agent does not know $\neg q$, even implicitly. But, nevertheless, there is a local contradiction because the agent has accepted $\neg q$ as true before, and now she has just accepted q too.

Note how, if there is indeed a proper justification to have both the rule's conclusion and its negation in the A-set, then this 'simple' model represents a situation that, as mentioned in van Benthem (2009), *"challenges our dynamic approach to belief change so far"*. Not only the contradiction cannot be solved by a reordering of the worlds: then the whole theory itself becomes subject of revision, and fundamental changes may be needed.

Here we just mention briefly two possibilities for dealing with this situation, without going into further details. One of them is equip the A set with a further structure, an ordering among formulas, like it is done in syntactic belief revision. This further extension would allow us to decide which elements of the theory should be thrown away and which ones should be kept. Some examples of this are entrenchment functions in belief revision (Gärdenfors and Makinson 1988), ordered theory presentation (Ryan 1992) and structured belief bases Kahle (2002). There are also more recent proposals based on the idea of ordered preferences (Liu 2010). Another possibility is to consider worlds that contain not a single set of formulas, but several of them, ordered by some plausibility relation (van Benthem 2009). If a contradiction arises, then we can perform a reordering, but now not among the worlds, but among the theories themselves.

Chapter 7

Connections with other fields

The non-omniscient and dynamic setting presented in this dissertation adds an extra dimension to standard *Dynamic Epistemic Logic*: besides the studied acts of observation (update) and revision (upgrade), our agents can get information by means of finer informational actions, like changes in awareness and diverse forms of inference. Thus, our framework can give a different perspective and therefore shed some light in areas that deal, in some form or another, with an agent's information and its dynamics.

Though in this dissertation we will not pursue particular applications, the present chapter introduces some proposals for connections. We focus on areas in Linguistics and Cognitive Science as well as Game Theory. Our purpose is not to provide formal and deep proposals, but simply to show how the main ideas behind our setting have a wide range of applications.

7.1 Linguistics

Our work can be related with *Linguistics*, the formal study of natural language. Among all the linguistic areas, there are interesting connections with the study of the notions of *attention*, *questions* and *pragmatics*.

7.1.1 Attention

The interest on the study of a notion of "attention" arises from the observation that, though every day of our life we face an large amount (and probably an infinite number) of possibilities, we, as agents with limited resources, do not have the ability to work properly with all, and at any given moment we only deal with a small subset of them.

There are several ways of defining what an agent is paying attention to, and our approach can represent some of them.

Attention as awareness Attention can be understood as a language-related notion: we pay attention to the possibilities our current language allows us to express. For example, Natalia is looking for her car's keys, and she is just paying attention to the possibility for them to be either in the bedroom, or in the kitchen, or in the dinning room. From her point of view, there exists only three possibilities, and she is not considering the possibility for the keys to be in the bathroom because "bathroom" is not in her current language.

More generally, if an agent is only aware of two atomic propositions p and q, then she will identify at most four possibilities: the four different combinations of p and q's truth-values. But what she identifies as "the p and q possibility" may correspond to a number of them that differ from each other in the truth-value of atomic propositions the agent is not currently entertaining, like r, s and so on. In other words, some possibilities are not considered because the agent's language is not fine enough to identify them in the first place.

This notion of attention corresponds directly to the notion of awareness we have dealt with in Chapter 4, in which the set of formulas the agent is aware of is generated by the set of atomic propositions she has available in all the worlds she considers possible (Definition 4.6). This gives us the following definition:

$$\mathrm{Att}\,\varphi := \mathrm{Aw}\,\varphi$$

Attention given by beliefs There are others understandings of what it means to be "paying attention" to a given possibility. Following ideas about conscious belief and investigation presented in Stalnaker (1984), the dissertation de Jager (2009) relates the notions of attention and inattention not only to syntactic sources, like the agent's language, but also to semantic ones, like the agent's beliefs. For example, Natalia is still looking for her keys, but now she is also not paying attention to the possibility for them to be in the kitchen because she considers that situation very unlikely to be the case.

More generally, even though an agent might be aware of the atomic propositions p and q, she might be paying attention just to two possibilities (e.g, the one with p and q true, and the one with p true and q false). Though the other two possibilities are expressible, she is not paying attention to them because, according to her beliefs, they are very implausible.

When beliefs are involved, there are several possibilities for defining what the agent is paying attention to. One option is given by an agent that pays attention to a given φ if and only if φ is true in at least one epistemically possible situation: $\mathrm{Att}\,\varphi := \langle\sim\rangle\,\varphi$. But we can also have a more radical agent that pays attention only to those possibilities that occur in at least one world that is more plausible than the current one: $\mathrm{Att}\,\varphi := \langle\leq\rangle\,\varphi$. We can even have a very radical agent that pays attention only to those possibilities that hold in at least one of the most plausible situations: $\mathrm{Att}\,\varphi := [\leq]\langle\leq\rangle\,\varphi$.

7.1.2 Questions

There are close relations between our framework and the logical analysis of *questions*. In particular, we see two important connections.

Questions as aware raising mechanism In Chapters 3 and 4 we discussed two different understandings of the notion of awareness: as the formulas of an arbitrary set, and as those generated from the set of atoms the agent can use in all the worlds she considers possible, respectively. Our discussion was not only about awareness' representation, but also about its dynamics, and we provided mechanisms for raising and dropping awareness.

But, though our actions showed the effect that changes in awareness have in the agent's information, besides situations like the *Twelve Angry Men* example (Section 4.1), we did not provide concrete reasons for these awareness' modifications. In other words, though we describe *how* awareness change, we did not justify *why* these changes happen.

One of the most natural ways to change the current awareness of an agent and focus her attention on a specific issue is by asking a question. Intuitively, when a question is asked, the hearer switches her attention to focus on the just raised issue. This is the approach followed by some of the most prominent logical treatment of questions (Groenendijk 2007; van Benthem and Minică 2009): a question separates the current set of possibilities into several groups according to the possible answers, therefore changing the agent's attention.

From our fine grained perspective, we can think of a combination of questions and finer representations of information: we can understand a question as an action that increases some current set of 'relevant propositions' whose truth value needs to be determined.

A setting in which we can embed the ideas of our framework in a natural way is the *DEL* approach to questions of van Benthem and Minică (2009). Semantically, their *epistemic issue* model contains, besides the non-empty set of words, their atomic valuation and an equivalence indistinguishability relation denoted by $\sim$, an equivalence *abstract issue relation*, denoted by $\approx$, that divides the set of possibilities in areas in which the agent would like to be. Syntactically, the epistemic language is extended with a universal modality for the issue relation, $[\approx]$, and the universal modality, U.

The important actions in this framework are not only those that announce a fact (an *announcement*), but also those that raise an issue (a *question*). While an announcement "$\varphi!$" differs from that of standard *PAL* in that it just cut links between worlds that disagree on φ without deleting any of them, the effect of asking a question "$\varphi?$" is a refinement of the issue relation: each issue partition is split into (possibly) two: one with the worlds that satisfy φ, and another with those that satisfy $\neg\varphi$.

This setting allows us to define statements describing the effect of a question. For example, the formula $U([\approx]\psi \vee [\approx]\neg\psi)$ expresses that ψ is settled as an issue across the whole model (Definition 4 in van Benthem and Minică (2009)). Then, the following formula states that ψ will be settled as an issue across the whole model *after asking* χ:

$$[\chi?]\, U\Big([\approx]\psi \vee [\approx]\neg\psi\Big)$$

But, from our non-omniscient perspective, the fact that the truth-value of ψ is uniform in all the agent's $\approx$-partitions is not enough to settle it as an issue *explicitly*. Consider, for example, our setting of Chapter 3 in which, in order for φ to be explicit information, we needed for the agent to be aware of it, defining $\mathrm{Ex}\,\varphi$ as $\Box(\varphi \wedge \mathrm{A}\,\varphi)$. Following the same methodology, we can interpret $[\approx]$ as *implicit* issue, and then define its *explicit* version as

$$\mathrm{Iss}\,\varphi := [\approx](\varphi \wedge \mathrm{A}\,\varphi)$$

Then, according to the mentioned formula, the following one expresses that after asking χ, ψ will be settled as an explicit issue across the whole model:

$$[\chi?]\, U\Big(\mathrm{Iss}\,\psi \vee \mathrm{Iss}\,\neg\psi\Big)$$

But now we can look for other versions of the act of asking a question. In van Benthem and Minică (2009)'s omniscient setting, "φ?" raises an issue not only about φ, but also about all formulas that are logically equivalent to it. But from our non-omniscient perspective, only the issue about φ should be raised *explicitly*, and the rest of the formulas logically equivalent to φ should be indeed an issue, but only in an *implicit* way.

Following the spirit of the act of *explicit observation* (Definition 3.7), an *explicit question* operation that follows the given intuition, denoted by "φ^+?", can be defined as the former "φ?" plus the additional effect of making the agent aware of φ. Given the definition of awareness, the latter requirement boils down to adding φ to the A-set of all possible worlds. Then we can build formulas like the following, expressing that ψ will be settled as an *explicit* issue across the whole model after asking χ *explicitly*:

$$[\chi^+?]\, U\Big(\mathrm{Iss}\,\psi \vee \mathrm{Iss}\,\neg\psi\Big)$$

This new setting, together with the actions defined in Section 3.5, allow us to express combinations of questions and changes in awareness. For example, the following formula expresses that, after asking χ explicitly, ψ will become an issue settled *implicitly* across the whole model, and it will be an issue settled *explicitly* as soon as the agent considers it:

$$[\chi^+?]\Big(\; U([\approx]\psi \vee [\approx]\neg\psi) \;\wedge\; [+\psi]\, U(\mathrm{Iss}\,\psi \vee \mathrm{Iss}\,\neg\psi) \;\Big)$$

These examples show how a question can be understood as mechanism that raises issues, and therefore creates awareness. Such changes can affect now only the agent's explicit knowledge (and beliefs), but also other attitudes, like preferences (Guo and Xiong 2010).

Questions and inferences The second strong connection is not with acts of awareness change, but with acts of inference. Most of the scientific inquiry can be described as a combination of questions and inferences, like Hintikka emphasizes in his *Interrogative Model of Inquiry* (*IMI*: Hintikka (1999); Hintikka et al. (2002); Hintikka (2007)). In the *IMI*, inquiry is represented as an information-seeking process in which the inquirer, based on some premises, tries to establish certain conclusion. At each stage, she has a choice between performing a deductive step in which a logical conclusion is derived from the information she has acquired so far, or performing an interrogative move in which she test a fact that she cannot justify or discard with her current information.

In the search for formalizations of the *IMI*, combinations of frameworks for questions and inference have already produced fruitful results. The master's dissertation Hamami (2010a) combines a logic for questions with a logic for tableau-based inference, and the inference part shares some similarities with our approach for rule based inference of Chapter 2, like the definition of explicit knowledge (called *local* knowledge) and the restriction for explicit information about only propositional formulas. On top of that, the system has the important advantage of providing to the agent a *complete* reasoning system.

For a simple example of a useful combination of questions an inference, recall the described *DEL* approach to questions (van Benthem and Minică 2009). Another important action defined there is the action of *resolution*, "!", in which the indistinguishability relation $\sim$ is restricted within the issue partitions (that is, is redefined as $\sim \cap \approx$). As indicated in the mentioned work, this operation is a natural generalization of an announcement that need not have natural language correspondent. With such operation we can build formulas like

$$[\chi?]\,[!]\,U\left([\sim]\,\psi \vee [\sim]\,\neg\psi\right)$$

expressing that a question about χ followed by a resolution will produce knowledge everywhere about whether ψ.

In a non-omniscient setting, a question and a resolution may not be enough to produce *explicit* knowledge. Even if an *implicit* question of χ followed by resolution produce indeed *implicit* knowledge about ψ, there is no guarantee that ψ will be also *explicitly* known. But then, all the agent needs is a further *inference step*. So suppose that, indeed, $[\chi?]\,[!]\,U\,([\sim]\,\psi \vee [\sim]\,\neg\psi)$ is the case. Then, we expect the following formula to be true too:

$$[\chi^{+}?]\,[!]\,[\hookrightarrow_{(\chi \Rightarrow \psi)}]\;U\,(K_{\mathrm{Ex}}\psi \vee K_{\mathrm{Ex}}\psi)$$

The formula expresses that an explicit question about χ followed by a resolution *and then an inference step* will produce *explicit* knowledge about ψ.

The combination of questions and inferences become even more appealing when we look not only at the inquirer's knowledge and her deductive inferences, but also at her beliefs and her inferences in general.

7.1.3 Pragmatics

Suppose your partner tells you truthfully "I'm cooking meat tonight". What information is conveyed by this announcement?

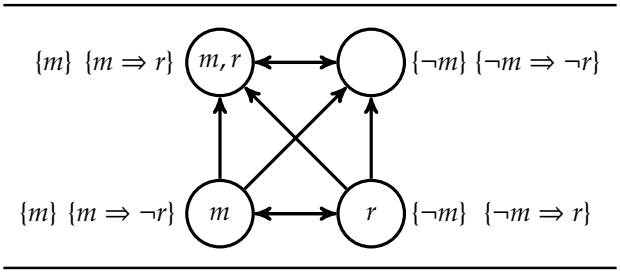
Besides the plain fact that your partner indeed will be cooking meat tonight, the message usually provides more information. Depending on the specific circumstances, it may also indicate "bring red wine", "do not be late" or, in some extreme cases, "do not show up at all". These pieces of information, despite being beyond the proper meaning of the announcement, are usually understood and acknowledged in our conversations.

Where does this extra information (implicatures) comes from? Why is it communicated? What is the role of the proper semantic meaning of the announced sentence in the extra information it provides? These questions are the concern of linguistic *Pragmatics*.

One of the most influential pragmatic theories is the one introduced by Paul Grice (see Grice (1989)). The main idea of his proposal is that, based on the assumption that the speaker obey certain 'maxims' about the informative purpose of a conversation, the hearer can extract additional information that is not covered by the semantic meaning of the statement. In other words, Grice proposed that conversational implicatures can be seen as further *inferences*, and that they can be justified by a reasoning process that takes into account not only the semantic meaning of the announced sentence, but also some aspects of the conversational context.

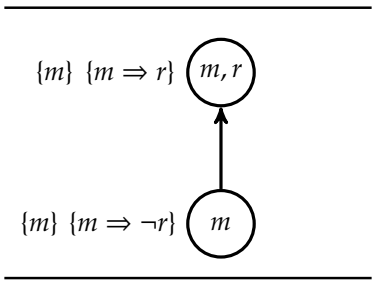
With this idea in mind, and following the methodology of our approach, implicatures can be seen as the result of further inference steps based on beliefs the hearer has about the speaker's intentions. Consider the mentioned example: a speaker announces "I'm cooking meat tonight", and from this the hearer infers that she needs to get red wine. How can this be represented in our setting?

First, note that the assumptions the hearer made about the drinking preferences of the speaker (white wine when cooking fish, red whine when cooking meat) are already encoded in the hearer's beliefs before the announcement takes place; it is in this sense that the hearer makes assumptions about the speaker's intentions during a conversation. This situation corresponds to the following model, in which m stands for *meat* (hence $\neg m$ stands for *fish*) and r stands for *red wine* ($\neg r$ stands for *white wine*). The arrows represent the plausibility relation, with the reflexive arcs omitted.

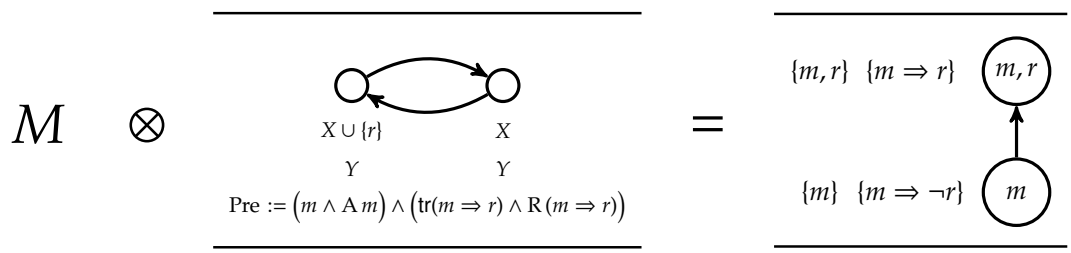


The hearer's plausibility order puts on top the *meat-red wine* and *fish-white wine* situations. Nevertheless, this is only implicit; the only explicit information the hearer has is about the food's choice in each possible situation, and about what drink it would imply in each one of them.

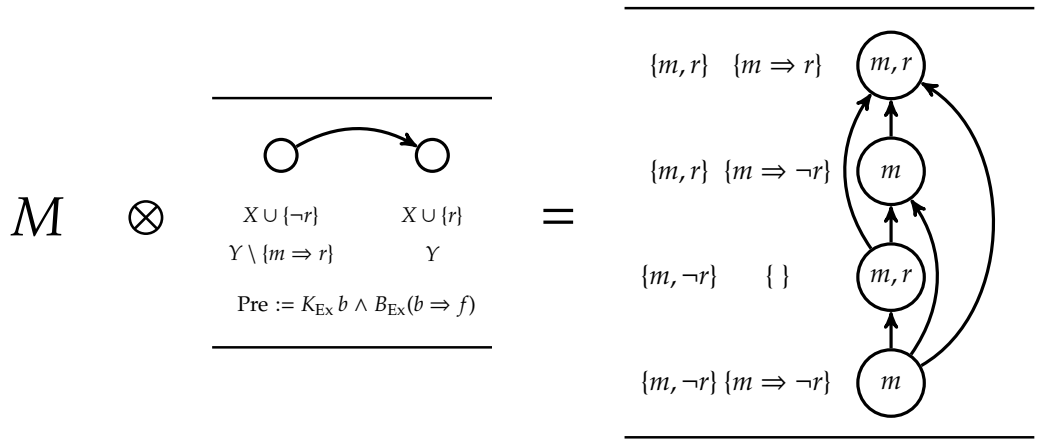
Then the speaker says 'I'm cooking meat tonight". The immediate effect of the utterance is that of a public announcement: the hearer will discard those situations she recognizes as $\neg m$-ones, i.e., the situations on the right column (see Section 4.4.2). This gives us the following model, which we call M:



Here is where the further reasoning takes place. The agent believes explicitly that meat corresponds to red wine, so she can perform an inference that will make her implicit belief about the red wine explicit. In our setting, she has at least two ways of doing it. The first one, a *strong local inference* with $m \Rightarrow r$ (Definition 5.18) will only make explicit the red wine belief ($B_{\mathrm{Ex}}\, r$):



The second possibility, an inference with known premises m and believed rule $m \Rightarrow r$ (Definition 5.15), will additionally acknowledge explicitly a (not plausible but still possible) situation with white wine ($B_{\mathrm{Ex}}\, r \wedge \widehat{K}_{\mathrm{Ex}}\, \neg r$):



Pragmatics as iterated best response The *iterated best response* mechanism proposed in Franke (2009) explains pragmatic phenomena as the result of a sequence of iterated best responses that start from the literal semantic meaning of the announcement and continue for as long as it is reasonable and the agents can do it. We have shown how our setting can capture the small steps that makes explicit the assumptions in each response, and our non-omniscient belief revision (Section 5.3.2) can represent how the agents 'correct' their beliefs for future responses. To describe the long-term behaviour, an extension that deal with iterations is needed, as we will discuss in Chapter 8.

7.2 Cognitive Science

Our framework deals with the representation of several notions of information and the way they are affected by diverse actions. Thus, it also has connections with Cognitive Science, the study of mind and how information (perception, language, reasoning, and emotion) is represented and transformed in the brain.

7.2.1 Learning Theory

Leaving aside for a moment the particular frameworks and tools that we have explored so far, the main subject of this dissertation is the study and representation of changes in information. Besides Logic (and, in particular, Epistemic Logic and its dynamic extensions), there are other approaches that deal with epistemic changes. One of these frameworks is *Learning Theory* (*LT*; see e.g. Jain et al. (1999)), the study of functions that attempt to identify the correct hypothesis from a collection of possibilities based on inductively given streams of data. Though it evolved from the study of language acquisition, learning theory focuses on various properties of the process of conjecture-change over time, and therefore it is applicable also in other fields, like philosophy of science, where it can be interpreted as a theory of empirical inquiry (Kelly 1996).

Using the language learning terminology, the basic premises in Learning Theory are that the agent considers several languages as the possible ones. Then she receives an infinite sequence of data that contains all words of the actual language. Based on this information, the agent tries to identify the actual language, in some cases by making a single conjecture after a finite number of data (*finite identification*; Mukouchi (1992)), and in some others by making a conjecture every certain time and waiting for the conjecture to stabilize to the correct one (*inductive inference*; Gold (1967); Angluin and Smith (1983)).

Recent works have looked at connections between Learning Theory and *DEL* (Gierasimczuk 2009; Ma 2009; Baltag and Smets 2009; Dégremont and Gierasimczuk 2009). The main idea is that, by representing the languages the agent considers possible in a possible worlds style, learning can be seen as a mechanism that, at each stage, decides how the just received data will change the agent's knowledge/beliefs.

Our fine-grained setting allows us to represent this process from the perspective of non-ideal agents. First, a 'real' agent may not have at hand the full language each possibility represents. More precisely, our A-sets that so far have contained the formulas the agent has acknowledged as true in that world, can now contain the words the agent has recognized as part of the language represented by that world. For example, if the agent considers only two possibilities, one standing for the language $\mathbf{a}(\mathbf{a}+\mathbf{b})^*$ and another standing for the language $\mathbf{a}(\mathbf{a}+\mathbf{b})^*\mathbf{a}$, then she knows *implicitly* that the word *aba* is in the language because it is in the two languages she considers possible. But, following our definition of explicit knowledge of Chapters 4 and 5, the knowledge is not explicit if she has not recognized the word in the two possibilities. This can be expressed with the following formula:

$$K_{\mathrm{Im}}(aba) \ \wedge \ \neg K_{\mathrm{Ex}}(aba)$$

Second. Though a 'real' agent does not need to have the full language each possibility represents, she may as well be able to construct words of it. The agent can have information not only about the words of each language, but also about how to generate more words from current ones. Following the ideas of *Formal Grammar*, one way to do this is by using *production rules*: then we can express situations in which, thanks to her current knowledge, the agent can derive another word of the language:

$$\Big(K_{\mathrm{Ex}}(abX) \wedge K_{\mathrm{Ex}}(X \Rightarrow a)\Big) \ \rightarrow \ [\hookrightarrow_{(X \Rightarrow a)}] \, K_{\mathrm{Ex}}(aba)$$

Third. Adding beliefs to the picture enriches the setting, allowing us to representing the agent's implicit and explicit hypothesis about the actual language at each stage and how it changes due to the received piece of information (Baltag and Smets 2009; Dégremont and Gierasimczuk 2009). In our non-omniscient

case, the following formula expresses that after receiving the explicit information that *abaa* is a word in the language, the agent will believe that *abaaa* is a word too:

$$\neg B_{\mathrm{Ex}}(abaaa) \wedge [abaa^{+} \Uparrow] B_{\mathrm{Ex}}(abaaa)$$

Finally, our setting for inferences involving beliefs allows us to represent small belief changes that are not direct consequence of the received data (just like conversational implicatures). For example, if the agent believes explicitly that $X \Rightarrow b$ is a production rule of the actual language, then she can use it to generate a new explicit belief. This is expressed by the following formula in which the *PA* action model $(C_{KB}^{X \Rightarrow b}, e)$ is the one of Definition 5.15.

$$\Big(K_{\mathrm{Ex}}(abX) \wedge B_{\mathrm{Ex}}(X \Rightarrow b)\Big) \rightarrow \langle C_{KB}^{X \Rightarrow b}, e\rangle \, (B_{\mathrm{Ex}}(abb) \wedge \neg K_{\mathrm{Ex}}(abb))$$

7.2.2 The notion of surprise

Surprising observations Many belief dynamics, like abduction and belief revision, are related to the notion of *surprise*, and there are some formal approaches to this concept. In particular, the framework of Lorini and Castelfranchi (2007) investigates the role of surprise in triggering the process of belief change, and distinguishes two main forms of surprise: *mismatch-based surprise* and *astonishment*. While the first one appears when the agent perceives information that contradicts the beliefs she is currently focusing on, the second one appears when the agent perceives information that is not in the focus of the agent. The latter has two variants: after the agent brings the topic into focus, she recognizes that she did not expect the observation, or even worst, that she expected the opposite of the observation.

These two notions of surprise can be represented in our framework. Note that the key difference between the two notions is the focus of the agent, so to get a proper representation, we need to incorporate the *awareness of* notion (we will use that of Chapter 4) to the beliefs framework of Chapter 5. Assume an extended definition of implicit and explicit beliefs of the following form:

The agent believes *implicitly* the *formula* φ	$B_{\mathrm{Im}}\varphi := \langle \leq \rangle [\leq] ([]\varphi \wedge \varphi)$
The agent believes *explicitly* the *formula* φ	$B_{\mathrm{Ex}}\varphi := \langle \leq \rangle [\leq] \Big([]\varphi \wedge \varphi \wedge \mathrm{A}\, \varphi \Big)$

We will use the modality $\langle \varphi^{+}! \rangle$ for the finer non-omniscient observation (i.e., an announcement with unspecified announcer) sketched in Section 4.4.2.

The first form of surprise, *mismatch-based surprise*, occurs when the agent faces an observation that contradicts the beliefs she is currently focusing on. Our notion of explicit beliefs already requires the agent's attention, so in our

setting this situation corresponds roughly to the following formula

$$B_{\mathrm{Ex}}\neg\chi \ \wedge\ \langle\chi^{+}!\rangle\top \qquad [1]$$

The second form, *astonishment*, occurs when the agent faces an observation that is not in her current focus and, after bringing into focus the related information, she recognizes that either she did not expected the observation, or else she expected exactly the opposite. This situation corresponds to the formula

$$\neg\mathrm{Aw}\,\chi \wedge \Big(\neg\langle\leq\rangle[\leq](\chi\wedge \mathrm{A}\,\chi)\vee\langle\leq\rangle[\leq](\neg\chi\wedge \mathrm{A}\,\neg\chi)\Big)\wedge\langle\chi^{+}!\rangle\top \qquad [2]$$

where $\langle\leq\rangle[\leq](\varphi\wedge \mathrm{A}\,\varphi)$ stands for beliefs that just need the agent's attention (i.e., awareness) to become explicit (cf. Subsection 4.3.2).

But our system can also express situations in which the agent can perceive surprises that happen not only at the *explicit* level, but also at the *implicit* one. Such surprises are stronger because what fails is not the agent's ability to make explicit her implicit beliefs (that is, the surprise does not arise because of lack of reasoning), but rather her plausibility order.

Other actions producing surprise We have discussed surprises that occur as a result of *observations*. But, are there other actions that can produce surprise?

For simplicity, we go back to the awareness-less definitions of implicit and explicit beliefs of Chapter 5. We will say that the formula χ is a *weak explicit* surprise if and only if the agent does not believe it explicitly: $\neg B_{\mathrm{Ex}}\chi$; we will say that χ is a *strong explicit* surprise if and only if the agent believes $\neg\chi$ explicitly: $B_{\mathrm{Ex}}\neg\chi$. Correspondent notions of *implicit* weak and strong surprise can be obtained by replacing B_{Ex} by B_{Im} in the previous definitions. The forms of surprise that can be produced by each one of the actions in our setting depend on what each action needs to take place.

Consider first our knowledge-related actions. In order for the agent to observe some χ she does not need any previous information, so an observation can produce not only the forms of surprise we just defined, but also many others. In the case of knowledge-based inference with a rule σ, a weak explicit surprise can be produced, since the agent does not need to believe explicitly the rule's conclusion before applying the rule. But none of the other forms of surprise is possible, and the reason is that in order for the inference to take place, $\mathsf{cn}(\sigma)$ should be already implicit knowledge, and therefore implicit belief. Then weak implicit surprise is not possible because it asks for $\mathsf{cn}(\sigma)$ not to be

[1] In fact, $\langle\chi^{+}!\rangle\top$ does not express that χ is *actually* observed; just that *it can be* observed. Then, the whole formula does not state that the agent is surprised, but rather than *she can be* surprised. An alternative definition would take as the evaluation point not the stage *before* the observation, but the stage *after* it. Nevertheless, it would need a *past-looking* modality to express what the agent believed *before* the observation (cf. Yap (2007)).

[2] Again, the formula only states that the agent can be surprised.

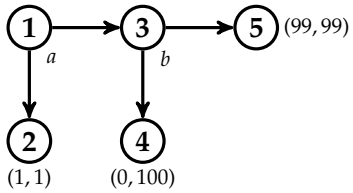
implicitly believed, and the two strong cases are also not possible because in both the agent would need to believe $\neg \mathsf{cn}(\sigma)$ implicitly, and our setting does not allow inconsistent beliefs.

Now consider our belief-related actions. Arbitrary acts of revision (i.e., arbitrary rearrangement of beliefs) can definitely produce surprises, again because there is no precondition attached to them. Whether a non-arbitrary rearrangement can produce surprising information or not depends on what the agent needs to perform the action. For example, our inference with known premises and believed rule aims to produce a situation in which the agent believes explicitly the rule's conclusion (and yet considers explicitly a possibility in which it fails). Then, the requirements imply that the rule's conclusion is already implicitly believed, and therefore, though the action can produce weak explicit surprise, it cannot produce the weak implicit or any of the strong versions. But this is reasonable, because as we discuss, this kind of inference resembles default reasoning, an inference that works on what is most likely to be the case.

Other forms of inferences involving beliefs can produce surprises. The already discussed *abductive reasoning* requires from the agent some information about the rule and its conclusion, but since this does not imply any attitude about the premises, surprises are possible. Nevertheless, recall that the goal of abductive reasoning is to find the 'best' an explanation of a given χ. Though an explanation can be definitely surprising, there are usually several candidates, and choosing the 'best' is generally understood as choosing the one that will produce the less amount of changes or, in other words, the one that is less surprising (and even not surprising at all). In fact, the notion of 'best' explanation could be tied to how surprising this explanation would be.

7.3 Game Theory

Game Theory analyzes competitive situations in which several agents have to make a (sequence of) choice(s). The outcome of the situation is then determined by the individual choices of each one of them. A typical example is the so called centipede game, whose diagram is shown below.



This game takes place between two players, a and b, with **1** the starting point and **2**, **4** and **5** the final ones. At points **1** and **3**, an agent has to take a decision (a and b, respectively), and the final payoffs of the game, indicated in the form (*player a*, *player b*), are determined by the players' choices.

One of the main goals of *Game Theory* is the study of *solution concepts*: formal rules that indicate optimal strategies for the game, and therefore predict the (not necessarily unique) final outcome. For example, *backward induction* predicts that in the centipede game of above, if point **3** were reached, b would choose **4** instead of **5**, getting 100 instead of 99, and leaving a with 0 instead of 99. But if a recognizes this, then she will realize that choosing between **2** and **3** actually means to choose between a payoff of 1 and a payoff of 0, respectively. Then she will choose **2**, and the final payoff of the game will be (1,1).

In order to indicate what each player will do, a solution concept needs to make assumptions not only about the nature of the game (perfect/imperfect information, strategic/extensive game, etc.), but also about the nature of the involved players. Recent literature (Aumann (1995); Stalnaker (1996); Aumann and Brandenburger (1995); Polak (1999); Chen et al. (2007) among others) has looked at the epistemic conditions that players need to satisfy in order to follow the solution concept's specification, and some of them have used *Epistemic Logic* and *Dynamic Epistemic Logic* tools to make formal these epistemic requirements. It turns out that the assumption of rationality that some of the most important solution concepts in the literature make (the mentioned *backward induction*, *Nash equilibrium*, *iterated elimination of strictly dominated strategies* among others) implies not only that every player will always make the choice that will give her the best possible outcome, but also that the players are always able to perfectly calculate every single consequence of every action. Not surprisingly, the predictions of solution concepts based on such strong assumptions usually do not coincide with the choices real agents do when facing these situations. Even in approaches that model games with incomplete information (e.g., Feinberg (2004)), it is implicitly assume that the players can derive all logical consequences of the information they have, and this is not necessarily the case.

Our framework allows us to model situations in which the involved agents are non-omniscient, and therefore allows us to explain why non-ideal players do not necessarily behave in an optimal way. Consider again the presented centipede game, and suppose a knows explicitly not only the structure of the game but also her preferences about the final state. Moreover, suppose that she believes explicitly that, if the game reaches point **3**, b will choose **4**. These assumptions can be expressed with the following formulas:

Game's structure	**Final state's preferences**	**Rationality**
$K^a_{\mathrm{Ex}}\big(\mathbf{1} \to [\mathrm{Next}](\mathbf{2} \vee \mathbf{3})\big)$	$K^a_{\mathrm{Ex}}\big(\mathbf{2} \to \mathrm{Neutral}_a\big)$ $K^a_{\mathrm{Ex}}\big(\mathbf{4} \to \mathrm{Worst}_a\big)$ $K^a_{\mathrm{Ex}}\big(\mathbf{5} \to \mathrm{Best}_a\big)$	$B^a_{\mathrm{Ex}}(\mathbf{3} \to [\mathrm{Next}]\mathbf{4})$

The game starts, so $K^a_{\mathrm{Ex}}\mathbf{1}$ is the case. By deduction, a knows explicitly that she has a choice between **2** and **3**, that is, $K^a_{\mathrm{Ex}}[\mathrm{Next}](\mathbf{2} \vee \mathbf{3})$. What will she do?

The important point here is that a non-omniscient agent may have not noticed the connection between her current choice, her explicit belief about b's choice at **3**, and the payoffs in each case. In other words, she may have not realized that if she chooses **3**, the game is very likely to end with **4** as the final state. In order to link those pieces of information, she needs to apply the following truth-preserving rule

$$\left\{[\mathrm{Next}]\,(\mathbf{2} \vee \mathbf{3})\ ,\ \mathbf{3} \rightarrow [\mathrm{Next}]\,\mathbf{4}\right\} \quad \Rightarrow \quad [\mathrm{Next}]\,(\mathbf{2} \vee [\mathrm{Next}]\mathbf{4})$$

While the first premise, $[\mathrm{Next}]\,(\mathbf{2} \vee \mathbf{3})$, is something a knows (explicitly), the second one, $\mathbf{3} \rightarrow [\mathrm{Next}]\,\mathbf{4}$, is something she only believes (explicitly). Then, after the inference step, a will only believe (explicitly) the conclusion, that is,

$$B^a_{\mathrm{Ex}}[\mathrm{Next}]\left(\mathbf{2} \vee [\mathrm{Next}]\mathbf{4}\right)$$

But again, being non-omniscient, she may still need to link the final states with her preference about them, that is, she may need to apply

$$\left\{[\mathrm{Next}]\,(\mathbf{2} \vee [\mathrm{Next}]\,\mathbf{4})\ ,\ (\mathbf{2} \rightarrow \mathrm{Neutral}_a)\ ,\ (\mathbf{4} \rightarrow \mathrm{Worst}_a)\right\}$$
$$\Rightarrow \quad [\mathrm{Next}]\left(\mathrm{Neutral}_a \vee [\mathrm{Next}]\,\mathrm{Worst}_a\right)$$

Again, the conclusion of this rule will be only believed (explicitly), that is,

$$B^a_{\mathrm{Ex}}[\mathrm{Next}]\left(\mathrm{Neutral}_a \ \vee \ [\mathrm{Next}]\,\mathrm{Worst}_a\right)$$

Only after these two inference steps a will realize that, according to her beliefs, her choice between **2** and **3** actually boils down to a choice between $\mathrm{Neutral}_a$ and a future Worst_a.

Though the example makes some simplifications, it definitely highlights one of the main reasons[3] why real agents might not choose the solution *backward induction* proposes: even if they have full knowledge about the structure and payoffs of the game and even if they believe they all will pick the highest payoff when having the choice, they might fail in establishing a direct relation between early moves in the game and later outcomes. In these cases, our non-omniscient analysis allows us to model not only the information these non-ideal agents have, but also the reasoning steps they need in order to reach an information state in which the strategy proposed by the solution concept will actually be played.

[3]Other explanation is a common agreement to reach an outcome that is better for *both* agents (**5** in our example).

Chapter 8

Conclusions

By focussing on finer notions of information, the framework developed in this dissertation allows us to represent small steps in dynamics of information.

More precisely, by zooming in on the omniscient notions of knowledge and belief, we have identified the notions of awareness, implicit/explicit knowledge, and implicit/ explicit beliefs that have been discussed through this work, as well as many others that have been just sketched. Technically, this has been achieved by extending the possible worlds model with functions that associate to each possible world a set of formulas, a set of atomic propositions and a set of rules. This merge of semantic and syntactic machineries has allowed us to represent finer notions of information and therefore non-omniscient agents. Table 8.1 shows the most important discussed notions.

Then we have studied different informational acts that transform these finer notions, focusing not only on non-omniscient versions of acts already studied, like observation and upgrade, but also on the actions that become meaningful in this non-omniscient setting: changes in awareness and different kinds of inferences. Technically, these actions have been defined as operations that modify not only the semantic part but also the syntactic component of our models. Table 8.2 shows the most important defined actions.

Let us review in more detail what each particular chapter has achieved.

8.1 Summary of the chapters

In Chapter 2 we have put together ideas from frameworks representing syntactic inference in a modal style (Duc 1997; Jago 2009) with the key ideas of the semantic-based Epistemic Logic (Hintikka 1962; Fagin et al. 1995). The result is a setting in which we can represent an agent's implicit and explicit information. Thus, the agent does not need to be omniscient anymore, because her implicit information does not need to be explicit. An important observation here is that our agent has explicit information not only about the way the world can be (i.e.,

Notion	**Definition**	**Model requirements**
Awareness of formulas	Chap. 3: $\mathrm{A}\,\varphi$	——
	Chap. 4: $\Box\,{}^{[]}\varphi$	——
Awareness of rules	Chap. 4: $\Box\,{}^{[]}\mathsf{tr}(\rho)$	——
Implicit *information* about formulas	Chap. 2: $\Box\,\varphi$	——
	Chap. 3: $\Box\,\varphi$	——
	Chap. 4: $\Box\,({}^{[]}\varphi \wedge \varphi)$	——
Implicit *information* about rules	Chap. 4: $\Box\,({}^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho))$	——
Explicit *information* about formulas	Chap. 2: $\mathrm{A}\,\gamma$	Coherence
	Chap. 3: $\Box\,(\varphi \wedge \mathrm{A}\,\varphi)$	——
	Chap. 4: $\Box\,({}^{[]}\varphi \wedge \varphi \wedge \mathrm{A}\,\varphi)$	——
Explicit *information* about rules	Chap. 4: $\Box\,({}^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho)$	——
Implicit *knowledge* about formulas	Chap. 2: $\Box\,\varphi$	Equivalence relation
	Chap. 4: $\Box\,({}^{[]}\varphi \wedge \varphi)$	Equivalence relation
	Chap. 5: $[\sim]\,\varphi$	Equivalence relation
Implicit *knowledge* about rules	Chap. 4: $\Box\,({}^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho))$	Equivalence relation
	Chap. 5: $[\sim]\,\mathsf{tr}(\rho)$	Equivalence relation
Explicit *knowledge* about formulas	Chap. 2: $\mathrm{A}\,\gamma$	Coherence and truth
	Chap. 4: $\Box\,({}^{[]}\varphi \wedge \varphi \wedge \mathrm{A}\,\varphi)$	Equivalence relation
	Chap. 5: $[\sim]\,\big(\varphi \wedge \mathrm{A}\,\varphi\big)$	Equivalence relation
Explicit *knowledge* about rules	Chap. 4: $\Box\,({}^{[]}\mathsf{tr}(\rho) \wedge \mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho)$	Equivalence relation
	Chap. 5: $[\sim]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$	Equivalence relation
Implicit *belief* about formulas	Chap. 5: $\langle\leq\rangle\,[\leq]\,\varphi$	Locally well-preorder
Implicit *belief* about rules	Chap. 5: $\langle\leq\rangle\,[\leq]\,\mathsf{tr}(\rho)$	Locally well-preorder
Explicit *belief* about formulas	Chap. 5: $\langle\leq\rangle\,[\leq]\,\big(\varphi \wedge \mathrm{A}\,\varphi\big)$	Locally well-preorder
Explicit *belief* about rules	Chap. 5: $\langle\leq\rangle\,[\leq]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$	Locally well-preorder

Table 8.1: Static notions of information.

Action	Description
Increasing awareness	The agent increases her awareness. Studied in Chapters 3 (public and private versions) and 4.
Dropping awareness	The agent increases her awareness. Studied in Chapter 3 (public and private versions).
Knowledge-based (i.e., truth-preserving) inference	Inference with explicitly known premises and explicitly known rule. Studied in Chapters 2, 4 and 5.
Belief-based inference	Inference that involve believed premises and/or believed rules. Studied in Chapter 5.
Structural operation	Extend the rules the agent knows. Studied in Chapter 2.
Implicit observation	An observation that does not affect the agent's explicit information. Studied in Chapter 3.
Explicit observation	An observation that affects the agent's explicit information. Studied in Chapters 2, 3 and 4.
Upgrade (revision)	Reordering (revision) of the agent's beliefs (omniscient and non-omniscient versions). Studied in Chapter 3.

Table 8.2: Actions and their effects.

her explicit information is not only about *formulas*), but also about procedures that allow her to extract more explicit information from what she already has (i.e., she also has information about *rules*).

Then we have turned our attention to the actions that modify the agent's information. We have presented an *explicit* version of the known *observation* (public announcement) action from Plaza (1989); Gerbrandy (1999). More interestingly, we have provided a model operation representing the act of inference, an act not considered in standard *DEL* due to the omniscient nature of the represented agents. This action allows the agent to extend her explicit information by making explicit the information that was only implicit before. We have also observed that, just like the agent can increase the formulas she explicitly knows, she can also extend the rules she explicitly have. We have provided model operations representing the application of *structural rules* (reflexivity, monotonicity, cut); their effect is to increase the rules the agent can apply.

Chapter 3 is devoted to the notion of awareness. We have observed that another reason for which an agent may not have explicit information about something that is true in all the worlds she considers possible is because she may not be aware of it. Then, we have looked at the *Awareness Logic* of Fagin and Halpern (1988), and discussed the different possibilities it offer us for defining a notion of explicit information.

On the dynamic side, we have examined different actions that modify the primitive notions of the framework, awareness and implicit information, and therefore modify the defined notion of explicit information. For the notion of awareness, we have reviewed actions that increase and decrease what the agent is aware of; for the notion of implicit information, we have recalled the notion of observation (public announcement). In all cases we reviewed the effect of these actions in the agent's *explicit* information. When going to a multi-agent scenario, we have observed that the actions we defined are 'public'; then, we have presented an extension of the action models of Baltag et al. (1999) that can deal with the syntactic component of our model, and therefore allows us to represent private and even unconscious versions of the mentioned actions.

In Chapter 4 we have put together the ingredients that have helped us to defined explicit information in the two previous chapters: in order to have explicit information about a certain formula, the agent needs to be aware of it, have implicit information about it, and acknowledge it as true. In particular, we have worked with a language-based notion of awareness that is given not by an arbitrary formulas as in Chapter 3, but by those generated from a set of atomic propositions. We have reviewed properties of these notions, focusing on the particular case of true information, that is, implicit and explicit knowledge.

We have then reviewed the actions of awareness raising, inference and explicit observations defined in the previous chapters, adapting them to the new richer setting. The new awareness raising operation works now by adding atomic propositions to the proper set (and the related act of awareness dropping can be represented by removing atoms from it); the inference action takes advantage of the language-based definition of the awareness notion; the explicit observation action becomes now an explicit *announcement* action that extends its previous behaviour by producing not only implicit and explicit information, but also awareness about the announced formula.

Chapter 5 has focused on a fine representation of the notion of beliefs. By combining ideas for representing beliefs in a possible worlds framework (van Benthem 2007) with ideas from the previous chapters for representing non-omniscient information, we have introduced a semantic model that allows us to represents an agent's implicit/explicit knowledge/beliefs about formulas/rules.

Several actions can be defined over our new model. We first explored a notion of belief revision already existing in the *DEL* literature, and we adapted it to our non-omniscient setting. But just like our agent can perform inference based on knowledge (that is, deduction), she can also perform inference that involve beliefs. By combining the plausibility action models of Baltag and Smets (2008) with the action models introduced in Chapter 3, we provided a setting that allows us to express a large variety of inferences. In particular, we defined inference with known premises and believed rule, believed premises and known rule, and even weak and strong forms of local inference.

Finally, Chapters 6 and 7 provide connections and applications of the developed framework in other fields. The first discusses the relations with diverse known forms of reasoning, focusing on deductive reasoning, default reasoning, abductive reasoning, the relation of belief bases with our explicit beliefs and the relation with purely inferential belief revision. The second discusses on connections with *Linguistics*, *Cognitive Science* and *Game Theory*.

While looking for answers for our original questions, the present work has also shed some light on some other areas.

First, we have shown how a definition of information that merges semantic and syntactic ideas allows us to get the best of both worlds. We still have some level of the abstract structure a semantic approach give us, but we also have the fine granularity that syntactic approaches provide us. More importantly, in our framework we can define the 'external' semantic actions that represent the agent's interaction with her environment as well as the 'internal' introspective acts that represent the agent's own reasoning.

Second, we have shown that there is a harmony between the external and the internal actions that change our information. Just as in standard *DEL* we have acts of 'hard' information that produce knowledge (observations) and 'soft' acts that produce beliefs (upgrade), in our non-omniscient setting we have 'hard' acts of knowledge-based inference that produce explicit knowledge as well as 'soft' acts of belief-based inference that produce explicit beliefs.

Third, though the presented framework for belief-based inference, *PA* action models, was developed for representing inferences that combine known/believed premises with known/believed rules, these ideas have produced a rich framework in which we can represent deductive reasoning as well as certain forms of default and abductive reasoning.

Finally, our whole approach has been based in defining explicit information in terms of other notions, like awareness, acceptance of formulas and implicit information. Thus, we have shown that, in a setting with multiple notions of information, the reductionist approach that defines some of them in terms of combinations of the others is feasible and interesting in its own right.

8.2 Further work

Like most research works, ours has provided some answers, but has also raised interesting questions. Here are the ones that we consider most appealing.

Long-term We have defined the effect of a single execution of several informational actions, but the result of their iterative application is also important. More precisely, fixed-point operators would allow us express the effect of iterative application of the defined actions, analogous to the Kleene star operator in *Propositional Dynamic Logic*.

How are fixed-point operators useful? Consider first the case of knowledge. Restricting ourselves to the agent's purely propositional knowledge, one would expect that, when provided with a 'complete' set of rules, the result of an agent's iterative applications of truth-preserving inferences over her explicit knowledge would be her implicit knowledge. This situation can be expressed with a formula of the form

$$K_{\mathrm{Im}}\gamma \rightarrow [(\hookrightarrow_{\cup})^{*}]\, K_{\mathrm{Ex}}\gamma$$

where the modality $[(\hookrightarrow_{\cup})^{*}]$ stands for the reflexive transitive closure of the application of the union of all truth-preserving inferences the agent can perform. But recall that explicit knowledge is already implicit knowledge, and that in the case of purely propositional facts, this explicit knowledge is not affected by inference operations (that is, if the agent knows explicitly some purely propositional fact, then she will still know it after any truth-preserving inference). Then what we have in fact is the full equivalence

$$K_{\mathrm{Im}}\gamma \leftrightarrow [(\hookrightarrow_{\cup})^{*}]\, K_{\mathrm{Ex}}\gamma$$

This formula shows how the omniscient epistemic notion of knowledge can be seen as the fixed point of a sequence of actions over a non-omniscient but dynamic notion. More generally, the use of fixed points suggest that ideal states can be seen not as a static property, but as the fixed point of the application of finer actions, thus highlighting the actions that are needed to reach such optimal point.

Now, the equivalence does not need to extend to the case of epistemic knowledge, since part of the agent's explicit knowledge can be invalidated by further inferences (consider Moore-type sentences "I do not know φ explicitly"). But then fixed points would provide us a way to study which pieces of epistemic information will be eventually overthrown and which ones will not.

Expressing the result of long-term iterative application of actions is also interesting when we look at the agent's beliefs. It would allow us to talk about information that, though it might never become proper knowledge, can nevertheless become a 'stable' belief that will not be affected by further steps. There are already some results on the effect of iterated belief revision (e.g., Baltag and Smets (2009)), and a fixed-point extension of our setting would allow us to include the described forms of default and abductive reasoning in the iteration. Some specific fields in which fixed points of our finer informational acts can be useful are

- *Learning Theory*, in particular, when dealing with the *identification in the limit* paradigm in which learning is seen as an infinite process that is successful if there is a finite number of steps after which the agent's hypothesis about the real language becomes stable,

- *Game Theory*, in particular, in the *Evolutionary Game Theory* approach that focuses not on properties of an ideal strategy but rather on the way strategies arise and evolve until they become optimal, and hence stable.

Multi-agent notions of information We have dealt with the single-agent notions of awareness, implicit/explicit knowledge, and implicit/explicit beliefs. And though we have dealt with multi-agent situations, we have not dealt multi-agent notions of information, like group knowledge/beliefs and, more interestingly, common knowledge/beliefs. In our fine-notions-of-information setting, this amounts to the study of implicit and explicit forms of group and common knowledge/beliefs.

The combination of implicit/implicit forms information gives us several cases. For example, in the case of group knowledge, while the very implicit form can be defined as "everybody in the group knows φ *implicitly*" (coinciding with the standard omniscient notion) and the very explicit form can be defined as "everybody in the group knows φ *explicitly*", there are now intermediate points in which while some agents know φ explicitly, the rest know it only implicitly.

The case of common knowledge is more interesting. Different from group knowledge, the omniscient version of common knowledge is not defined by a finite conjunction (assuming the group of agents is finite), but as an infinite one. There is more room for intermediate forms: explicit and full common knowledge corresponds to "everybody knows it explicitly and everybody knows that everybody knows it explicitly and ...", and we can find several situations in which some levels of the knowledge of some agent is only implicit.

Now for the dynamics. Once that implicit group knowledge has been reached (everybody knows implicitly that φ holds), then just actions of awareness raising (in the case of Chapter 3) or even acts of inference (in the case of Chapter 4) are needed to reach explicit group knowledge. But for reaching common explicit knowledge from its implicit form we need *group* introspective acts in which the agent recognizes that everybody in the group knows something explicitly. Moreover, being an infinitary notion, we do need fixed-points operations to define explicit common knowledge, since our language can only deal with finite formulas. But then we can look at the finite versions and verify how many levels of group introspection are actually needed for real agents to behave in a proper way (cf. Flobbe et al. (2008)).

Justifications By comparing our framework with the Logic of Justifications of Artemov and Nogina (2005), we can observe that our static framework can be seen as a special case in which each formula can have one and only one justification: the formula itself. Under this interpretation, our definitions of explicit information get a different reading. For example, our awareness-less definition of explicit knowledge, $\Box(\varphi \wedge \mathrm{A}\,\varphi)$, read as *"$\varphi$ is true and the agent has*

acknowledged it in all the worlds she considers possible", can be read now as "φ *is true and the agent has the justification for it in all the worlds she considers possible*". But clearly there may be more than one justification (evidence) for every true formula. In fact, some authors (e.g., van Benthem and Martínez (2008)) have proposed a definition of explicit knowledge that involve a universal quantification over the possible situations and existential quantification over evidence: the agent knows explicitly φ if and only if she has *a* justification for it *in all* the worlds she considers possible.

This suggest a further classification of justification. For example, while some facts are justified by deductive reasoning, some others are justified by their observation. But then we should look more formally at a calculus of justifications that involve our inferential steps. With this idea in mind, the act of deductive inference can be seen as an act of building a justification for the conclusion from the justifications of the premises.

Further dimension of syntactic structure When dealing with knowledge, existing semantic and syntactic approaches consider plain sets: of possible worlds in the first case and of formulas in the second. Our approach for pure knowledge of Chapters 2 and 4 simply combines them by assigning a set of formulas (and of rules) to each possible world.

When dealing with beliefs, existing semantic and syntactic approaches consider ordered sets: of possible worlds in the first case and of formulas in the second. Our approach for beliefs of Chapter 5 provides only an ordering for the worlds the agent considers possible, and the collection of formulas attached to each world does not have any further structure. The proposed framework can be extended in order to be fair to both semantic and syntactic approaches by considering, besides an ordering among the possible worlds, an ordering among the formulas (and rules) accepted in each one of them. This would allow us to deal with internal syntactic contradictions, as we briefly discussed in Section 6.5, but would also provide us with tools for more refined inference processes. In particular, it would give us tools to choose 'the best' explanation in abductive reasoning (cf. Section 6.3).

Appendix A

Technical appendix

A.1 Completeness of IE_K w.r.t. IE-models

This section provides the completeness proof for the implicit/explicit language with respect to **IE**-models (Theorem 2.1). Non-defined concepts like satisfiability of a formula and a set of formulas, a Λ-consistent (inconsistent) set and a maximal Λ-consistent set (for Λ an axiom system) are completely standard, and can be found in chapter 4 of Blackburn et al. (2001).

For a completeness proof, the key observation is that an axiom system Λ is strongly complete with respect to a class of models if and only if every Λ-consistent set is satisfiable in a structure of the given class (Proposition 4.12 of Blackburn et al. (2001)). Then, we will use the canonical model technique to show that every IE-consistent set is satisfiable in a pointed **IE**-model. Proofs of Lindenbaum's Lemma and Existence Lemma are standard. For the Truth Lemma, we will prove the case for our access and rule set formulas.

Lemma A.1 (Lindenbaum's Lemma) *For any IE-consistent set of formulas Σ, there is a maximal IE-consistent set Σ^+ such that $\Sigma \subseteq \Sigma^+$.* ■

Definition A.1 (Canonical model for IE) Recall that $\mathcal{IE}$ denotes the implicit/explicit language. The canonical model of the axiom system IE is the model

$$M^{\mathsf{IE}} = \langle W^{\mathsf{IE}}, R^{\mathsf{IE}}, V^{\mathsf{IE}}, \mathsf{A}^{\mathsf{IE}}, \mathsf{R}^{\mathsf{IE}} \rangle$$

where:

- W^{IE} is the set of all maximal IE-consistent set of formulas;
- $R^{\mathsf{IE}}wu$ iff for all φ in $\mathcal{IE}$, $\varphi \in u$ implies $\Diamond \varphi \in w$ (or, equivalently, $R^{\mathsf{IE}}wu$ iff for all φ in $\mathcal{IE}$, $\Box \varphi \in w$ implies $\varphi \in u$);

and, for all $w \in W^{\mathsf{IE}}$

- $V^{\mathsf{IE}}(w) := \{\, p \in \mathtt{P} \mid p \in w \,\}$;
- $\mathsf{A}^{\mathsf{IE}}(w) := \{\, \gamma \in \mathcal{L}_P \mid \mathrm{A}\,\gamma \in w \,\}$;
- $\mathsf{R}^{\mathsf{IE}}(w) := \{\, \rho \in \mathcal{R} \mid \mathrm{R}\,\rho \in w \,\}$; ◂

Lemma A.2 (Existence Lemma) *For every world $w \in W^{\mathsf{IE}}$, if $\Diamond\,\varphi \in w$, then there is a world $u \in W^{\mathsf{IE}}$ such that $R^{\mathsf{IE}}wu$ and $\varphi \in u$.* ■

Lemma A.3 (Truth Lemma) *For all $w \in W^{\mathsf{IE}}$, we have $(M^{\mathsf{IE}}, w) \Vdash \varphi$ iff $\varphi \in w$.*

Proof. We prove the case of access and rule set formulas. For the first,

$$\begin{array}{lll} (M^{\mathsf{IE}}, w) \Vdash \mathrm{A}\,\gamma \quad \text{iff} \quad \gamma \in \mathsf{A}^{\mathsf{IE}}(w) & & \text{by semantic interpretation} \\ \qquad\qquad\quad \text{iff} \quad \mathrm{A}\,\gamma \in w & & \text{by definition of } \mathsf{A}^{\mathsf{IE}} \end{array}$$

The case of rule set formulas is similar. ■

By the mentioned Proposition 4.12 of Blackburn et al. (2001), all we have to do is show that every IE-consistent set is satisfiable in a pointed **IE**-model, so take any such set Σ. By Lindenbaum's Lemma, we can extend it to a maximal IE-consistent set Σ^+; by the Truth Lemma, we have $(M^{\mathsf{IE}}, \Sigma^+) \Vdash \Sigma$, so Σ is satisfiable in the canonical model of IE at Σ^+. It is only left to show that M^{IE} is indeed a model in **IE**, that is, we have to show that it satisfies coherence for formulas and rules.

Remember that any maximal IE-consistent set Φ is closed under logical consequence, that is, if φ and $\varphi \to \psi$ are in Φ, so is ψ.

- *Coherence for formulas*. Suppose $\gamma \in \mathsf{A}^{\mathsf{IE}}(w)$; we want to show that for all u such that $R^{\mathsf{IE}}wu$ we have $\gamma \in \mathsf{A}^{\mathsf{IE}}(u)$. Note that $\mathrm{A}\,\gamma \to \Box\,\mathrm{A}\,\gamma$ (axiom $Coh_{\mathcal{L}_P}$) is in w.

  By definition, $\gamma \in \mathsf{A}^{\mathsf{IE}}(w)$ implies $\mathrm{A}\,\gamma \in w$; by the logical consequence closure, we have $\Box\,\mathrm{A}\,\gamma \in w$. Take any u such that $R^{\mathsf{IE}}wu$; by definition of R^{IE} we have $\mathrm{A}\,\gamma \in u$, and therefore $\gamma \in \mathsf{A}^{\mathsf{IE}}(u)$.

- *Coherence for rules*. Similar to the case of formulas, using the $Coh_{\mathcal{R}}$ axiom.

A.2 Completeness of IE_K w.r.t. $\mathbf{IE}_K$-models

This section provides the completeness proof for the implicit/explicit language with respect to $\mathbf{IE}_K$-models (Theorem 2.2).

We know already that IE is complete with respect to models in **IE** (Theorem 2.1). In order to show that IE_K is complete with respect to $\mathbf{IE}_K$, we just have to show that the canonical model for IE_K satisfy *equivalence*, *truth for formulas* and *truth for rules*.

Definition A.2 (Canonical model for IE_K) The canonical model for IE_K, $M^{\mathsf{IE}_K} = (W^{\mathsf{IE}_K}, R^{\mathsf{IE}_K}, V^{\mathsf{IE}_K}, \mathsf{A}^{\mathsf{IE}_K}, \mathsf{R}^{\mathsf{IE}_K})$, is defined just as the canonical model for IE (Definition A.1), but the worlds are maximal IE_K-consistent sets of formulas instead of maximal IE-consistent ones. ◄

Here is the proof for the three properties.

- *Equivalence*. Axioms *T*, *4* and *5* are canonical for reflexivity, transitivity and euclideanity, respectively, so R^{IE_K} is an equivalence relation.
- *Truth for formulas*. We want to show that $\gamma \in \mathsf{A}^{\mathsf{IE}_K}(w)$ implies $(M^{\mathsf{IE}_K}, w) \Vdash \gamma$. Suppose $\gamma \in \mathsf{A}^{\mathsf{IE}_K}(w)$; then we get $\mathrm{A}\,\gamma \in w$. By axiom $Tth_{\mathcal{L}_p}$ we have $\gamma \in w$; by the Truth Lemma, $(M^{\mathsf{IE}_K}, w) \Vdash \gamma$.
- *Truth for rules*. Similar to the case of formulas, with axiom $Tth_{\mathcal{R}}$.

A.3 Closure of deduction operation

This section proves that $\mathbf{IE}_K$-models are closed under the deduction operation (Proposition 2.1).

Let M be a model in $\mathbf{IE}_K$. To show that $M_{\hookrightarrow_\rho}$ (Definition 2.16) is also in $\mathbf{IE}_K$, we will show that it satisfies coherence and truth for formulas, coherence and truth for rules, and equivalence. Equivalence and both properties of rules are immediate since neither the accessibility relation nor the rule set function are modified. For the properties of formulas, we have the following.

- *Coherence for formulas*. Suppose $\gamma \in \mathsf{A}'(w)$ and pick any $u \in W$ such that Rwu in $M_{\hookrightarrow_\rho}$; we will show that $\gamma \in \mathsf{A}'(u)$.

  From the definition of A', we know that γ was added by the operation or was already in $\mathsf{A}(w)$. In the first case, γ should be $\mathsf{cn}(\sigma)$ and therefore $\mathsf{pm}(\sigma) \subseteq \mathsf{A}(w)$ and $\sigma \in \mathsf{R}(w)$. But then, by coherence for formulas and rules of M and the fact that Rwu, we have $\mathsf{pm}(\sigma) \subseteq \mathsf{A}(u)$ and $\sigma \in \mathsf{R}(u)$; therefore, the operation also adds $\mathsf{cn}(\sigma)$ (our γ) to the access set of u, that is, $\gamma \in \mathsf{A}'(u)$. In the second case, by coherence for formulas of M and Rwu, we have $\gamma \in \mathsf{A}(u)$ and therefore $\gamma \in \mathsf{A}'(u)$.

- *Truth for formulas*. Suppose $\gamma \in \mathsf{A}'(w)$; we will show that $(M_{\hookrightarrow_\sigma}, w) \Vdash \gamma$.

  Again, from the definition of A', we know that γ was added by the operation or was already in $\mathsf{A}(w)$. In the first case, γ should be $\mathsf{cn}(\sigma)$ and therefore $\mathsf{pm}(\sigma) \subseteq \mathsf{A}(w)$ and $\sigma \in \mathsf{R}(w)$. By truth for formulas of M we have $(M, w) \Vdash \bigwedge_{\gamma \in \mathsf{pm}(\sigma)} \gamma$; by truth for rules of M we have $(M, w) \Vdash (\bigwedge_{\gamma \in \mathsf{pm}(\sigma)} \gamma) \rightarrow \mathsf{cn}(\sigma)$. Therefore, we have $(M, w) \Vdash \mathsf{cn}(\sigma)$, i.e., $(M, w) \Vdash \gamma$. In the second case, by truth for formulas of M we also get $(M, w) \Vdash \gamma$.

Now, γ is a propositional formula so its truth value depends only on the valuation at w. But since the valuations at M and $M_{\hookrightarrow_\sigma}$ are the same, we get $(M_{\hookrightarrow_\sigma}, w) \Vdash \gamma$, as required.

A.4 Closure of structural operations

This section proves that $\mathbf{IE}_K$-models are closed under the three structural operations of Definition 2.18, *reflexivity*, *monotonicity* and *cut* (Proposition 2.2).

The proposition already argues for equivalence and the two properties of formulas, coherence and truth. It is only left to prove coherence and truth for rules for each one of the three operations.

Coherence In the case of the *reflexivity* operation, the coherence property follows immediately, since the original model M already has the property and the new rule ς_δ is added to the rule set of all worlds in M_{Ref_δ}. For the *monotonicity* operation, we just need to check coherence for the new rule ς'. Recall that it is added only to worlds that already have ς. But if a world w has ς in M, then, by coherence for rules of M, every world R-reachable from w also has ς in M, and therefore it will have ς' in $M_{\mathrm{Mon}_{\delta,\varsigma}}$. The case of *cut* is similar: ς' is added to those worlds that have ς_1 and ς_2, but then every world R-accessible from w also has ς_1 and ς_2 in M, so they will have ς' in $M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}$.

Truth Note that for this property, it is enough in all three cases to show that the added rules are truth-preserving in M because the truth-value of the translation, a purely propositional formula, depends just on the valuation, which is preserved by the operations. So let R be the rule set function of the original model M, R' be the rule set function of the corresponding new model, and pick any world $w \in W$.

- *Reflexivity*. Recall that ς_δ is given by $\{\delta\} \Rightarrow \delta$, and pick any rule $\rho \in \mathsf{R}'(w)$. If ρ is already in $\mathsf{R}(w)$, we have $(M, w) \Vdash \mathsf{tr}(\rho)$ since M is in $\mathbf{IE}_K$. Otherwise, ρ is ς_δ, and we obviously have $(M, w) \Vdash \delta \to \delta$.

- *Monotonicity*. Recall that ς' is given by $\mathsf{pm}(\varsigma) \cup \{\delta\} \Rightarrow \mathsf{cn}(\varsigma)$, and pick any $\rho \in \mathsf{R}'(w)$. If ρ is already in $\mathsf{R}(w)$, we have $(M, w) \Vdash \mathsf{tr}(\rho)$. Otherwise, ρ is ς', and then $\varsigma \in \mathsf{R}(w)$. Since M is in $\mathbf{IE}_K$, we have $(M, w) \Vdash (\bigwedge_{\gamma \in \mathsf{pm}(\varsigma)} \gamma) \to \mathsf{cn}(\varsigma)$ and therefore $(M, w) \Vdash ((\bigwedge_{\gamma \in \mathsf{pm}(\varsigma)} \gamma) \wedge \delta) \to \mathsf{cn}(\varsigma)$.

- *Cut*. Recall that ς' is given by $(\mathsf{pm}(\varsigma_2) \setminus \{\mathsf{cn}(\varsigma_1)\}) \cup \mathsf{pm}(\varsigma_1) \Rightarrow \mathsf{cn}(\varsigma_2)$, and pick any $\rho \in \mathsf{R}'(w)$. If $\rho \in \mathsf{R}(w)$, we have $(M, w) \Vdash \mathsf{tr}(\rho)$ since M is in $\mathbf{IE}_K$. Otherwise, ρ is ς' and we have $\{\varsigma_1, \varsigma_2\} \subseteq \mathsf{R}(w)$.

  Suppose $\bigwedge_{\gamma \in \mathsf{pm}(\varsigma')} \gamma$ is true at w in M; then, every premise of ς' is true at w in M. This includes every premise of ς_1 and every premise of ς_2 except

$\mathsf{cn}(\varsigma_1)$. But since every premise of ς_1 is true at w in M and ς_1 is in $\mathsf{R}(w)$, truth for rules of M tells us that $\mathsf{cn}(\varsigma_1)$ is true at w in M and hence every premise of ς_2 is true at w in M. Now, since ς_2 is in $\mathsf{R}(w)$, truth for rules of M tell us that $\mathsf{cn}(\varsigma_2)$, that is, $\mathsf{cn}(\varsigma')$, is true at w in M. Then we have $(M, w) \Vdash \mathsf{tr}(\varsigma')$.

A.5 Structural operations and deduction

This section provides a sketch for the proof of the validities of Table 2.8.

Take any pointed $\mathbf{IE}_K$-model (M, w). The main idea of the proof is that, under the appropriate circumstances, different sequences of operations produce models that are exactly the same *from w's point of view*, and therefore satisfy the same formulas of our language. For example, we will argue that if we have $\delta \in \mathsf{A}(w)$ and $\varsigma \in \mathsf{R}(w)$, then the pointed models $((M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}, w)$ and $((M_{\hookrightarrow_{\varsigma}})_{\mathrm{Mon}_{\delta,\varsigma}}, w)$ are the same from w's point of view (third entry for monotonicity in Table A.1 below). Note how a stronger identity between models does not hold because we cannot verify what happens in worlds that are not reachable from w. Therefore, we will state this "identity from w's perspective" in terms of an extended notion of *bisimulation* that asks for related worlds to have the same access and rule set.

Definition A.3 (Bisimulation) Take two **IE**-models $M_1 = \langle W_1, R_1, V_1, \mathsf{A}_1, \mathsf{R}_1\rangle$ and $M_2 = \langle W_2, R_2, V_2, \mathsf{A}_2, \mathsf{R}_2\rangle$. A non empty relation $B \subseteq (W_1 \times W_2)$ is a *bisimulation* between M_1 and M_2 (in symbols, $M_1 \underline{\leftrightarrow}_B M_2$) if and only if B is a standard bisimulation between $\langle W_1, R_1, V_1\rangle$ and $\langle W_2, R_2, V_2\rangle$ and, if Bw_1w_2, then $\mathsf{A}_1(w_1) = \mathsf{A}_2(w_2)$ and $\mathsf{R}_1(w_1) = \mathsf{R}_2(w_2)$.

We will write $(M_1, w_1) \underline{\leftrightarrow}_B (M_2, w_2)$ when $M_1 \underline{\leftrightarrow}_B M_2$ and Bw_1w_2. ◀

The validity of the formulas stated in Table 2.8 follows from the bisimilarities between models stated in Table A.1, where models of the form $(M_{\mathrm{STR}})_{\hookrightarrow_\sigma}$ are the result of applying the structural operation STR and then the deduction operation with rule σ, and similar for models of the form $(M_{\hookrightarrow_\sigma})_{\mathrm{STR}}$. In all cases, the bisimulation is the identity relation over worlds reachable from w.

Now for the proof. The involved operations (structural ones and deduction) preserve worlds, accessibility relations and valuations. Then, in order to show that the identity relation over worlds reachable from w is indeed a bisimulation, we just need to show that such worlds have the same access and rule set in both models.

Consider as an example the third bisimilarity for monotonicity; we will work with w first. For access sets, take any γ in the access set of w in $(M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}$; by definition, either it was already in that of w in $M_{\mathrm{Mon}_{\delta,\varsigma}}$ or else it was added by

Reflexivity with ς_δ the rule $\{\delta\} \Rightarrow \delta$	
If $\sigma \neq \varsigma_\delta$, then	$\left((M_{\mathrm{Ref}_\delta})_{\hookrightarrow_\sigma}\ ,\ w\right) \underline{\leftrightarrow} \left((M_{\hookrightarrow_\sigma})_{\mathrm{Ref}_\delta}\ ,\ w\right)$
If $\varsigma_\delta \in \mathsf{R}(w)$, then	$\left((M_{\mathrm{Ref}_\delta})_{\hookrightarrow_{\varsigma_\delta}}\ ,\ w\right) \underline{\leftrightarrow} \left(M_{\hookrightarrow_{\varsigma_\delta}}\ ,\ w\right)$
If $\delta \in \mathsf{A}(w)$, then	$\left((M_{\mathrm{Ref}_\delta})_{\hookrightarrow_{\varsigma_\delta}}\ ,\ w\right) \underline{\leftrightarrow} \left((M_{\hookrightarrow_{\varsigma_\delta}})_{\mathrm{Ref}_\delta}\ ,\ w\right)$
Monotonicity with ς' the rule $\mathsf{pm}(\varsigma) \cup \{\delta\} \Rightarrow \mathsf{cn}(\varsigma)$	
If $\sigma \neq \varsigma'$, then	$\left((M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_\sigma}\ ,\ w\right) \underline{\leftrightarrow} \left((M_{\hookrightarrow_\sigma})_{\mathrm{Mon}_{\delta,\varsigma}}\ ,\ w\right)$
If $\varsigma' \in \mathsf{R}(w)$, then	$\left((M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}\ ,\ w\right) \underline{\leftrightarrow} \left(M_{\hookrightarrow_{\varsigma'}}\ ,\ w\right)$
If $\delta \in \mathsf{A}(w)$ and $\varsigma \in \mathsf{R}(w)$, then	$\left((M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}\ ,\ w\right) \underline{\leftrightarrow} \left((M_{\hookrightarrow_\varsigma})_{\mathrm{Mon}_{\delta,\varsigma}}\ ,\ w\right)$
Cut with ς' the rule $(\mathsf{pm}(\varsigma_2) \setminus \{\mathsf{cn}(\varsigma_1)\}) \cup \mathsf{pm}(\varsigma_1) \Rightarrow \mathsf{cn}(\varsigma_2)$	
If $\sigma \neq \varsigma'$, then	$\left((M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}})_{\hookrightarrow_\sigma}\ ,\ w\right) \underline{\leftrightarrow} \left((M_{\hookrightarrow_\sigma})_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}}\ ,\ w\right)$
If $\varsigma' \in \mathsf{R}(w)$, then	$\left((M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}})_{\hookrightarrow_{\varsigma'}}\ ,\ w\right) \underline{\leftrightarrow} \left(M_{\hookrightarrow_{\varsigma'}}\ ,\ w\right)$.
If $(\mathsf{pm}(\varsigma_1) \cup \{\mathsf{cn}(\varsigma_1)\}) \in \mathsf{A}(w)$ and $\varsigma_1 \in \mathsf{R}(w)$, then	$\left(M_{\mathrm{Cut}_{\varsigma_1,\varsigma_2}\hookrightarrow_{\varsigma'}}\ ,\ w\right) \underline{\leftrightarrow} \left(M_{\hookrightarrow_{\varsigma_2}\mathrm{Cut}_{\varsigma_1,\varsigma_2}}\ ,\ w\right)$.

Table A.1: Bisimilarities for deduction and structural operations

the deduction operation with ς'. In the first case, γ is in the access set of w in M, since structural operations do not modify access sets; then it is also in the access set of w in $M_{\hookrightarrow_\varsigma}$ and in that of w in $(M_{\hookrightarrow_\varsigma})_{\mathrm{Mon}_{\delta,\varsigma}}$. In the second case, γ should be $\mathsf{cn}(\varsigma')$, but then we have the premises of ς' (and hence those of ς) in the access set of w in $M_{\mathrm{Mon}_{\delta,\varsigma}}$. Then, they are already in that of w in M and, by hypothesis, we have ς in the rule set of w in M, so $\mathsf{cn}(\varsigma)$, which is nothing but $\mathsf{cn}(\varsigma')$, is in the access set of w in $M_{\hookrightarrow_\varsigma}$ and hence it is in that of w in $(M_{\hookrightarrow_\varsigma})_{\mathrm{Mon}_{\delta,\varsigma}}$.

For the other direction, take γ in the access set of w in $(M_{\hookrightarrow_\varsigma})_{\mathrm{Mon}_{\delta,\varsigma}}$. Then it is in that of w in $M_{\hookrightarrow_\varsigma}$ and therefore either it was already in that of w in M, or else it was added by the deduction operation. In the first case, γ is preserved through the monotonicity and the deduction operations, and therefore it is in the access set of w at $(M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}$. In the second case, γ should be $\mathsf{cn}(\varsigma)$, and then we should have $\mathsf{pm}(\varsigma)$ and ς in the corresponding sets of w in M. By hypothesis, we have δ in the access set of w in M, so we have all the premises of ς' in the access set of w in M; therefore they are also in that of w at $M_{\mathrm{Mon}_{\delta,\varsigma}}$. Since we have ς in the rule set of w in M, we have ς' in that of w in $M_{\mathrm{Mon}_{\delta,\varsigma}}$ too. Hence, we have $\mathsf{cn}(\varsigma')$, which is nothing but $\mathsf{cn}(\varsigma)$, in the access set of w in $(M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}$.

The argument for rules is similar, and hence w has the same access and rule sets in $(M_{\mathrm{Mon}_{\delta,\varsigma}})_{\hookrightarrow_{\varsigma'}}$ and $(M_{\hookrightarrow_\varsigma})_{\mathrm{Mon}_{\delta,\varsigma}}$.

Now suppose a world u is reachable from w through the accessibility relation at $M_{\mathrm{Mon}_{\delta,\varsigma}\hookrightarrow_{\varsigma'}}$. Since R is not modified by the operations, u is reachable from w at M and therefore u is reachable from w at $M_{\hookrightarrow_{\varsigma}\mathrm{Mon}_{\delta,\varsigma}}$ too. Now we use the coherence properties: since $\delta \in \mathsf{A}(w)$ and $\varsigma \in \mathsf{R}(w)$, we have δ and ς in the corresponding sets of u, and then we can apply the argument used for w to show that u has the same information and rule set on both models.

These bisimulations allow us to prove the validities of Table 2.8. For example, recall the two formulas for monotonicity:

Monotonicity with ς' the rule $\mathsf{pm}(\varsigma) \cup \{\delta\} \Rightarrow \mathsf{cn}(\varsigma)$
$\langle \mathrm{Mon}_{\delta,\varsigma}\rangle \langle \hookrightarrow_{\sigma}\rangle \varphi \;\leftrightarrow\; \langle \hookrightarrow_{\sigma}\rangle \langle \mathrm{Mon}_{\delta,\varsigma}\rangle \varphi \quad \text{for } \sigma \neq \varsigma'$
$\langle \mathrm{Mon}_{\delta,\varsigma}\rangle \langle \hookrightarrow_{\varsigma'}\rangle \varphi \;\leftrightarrow\; \big(\langle \hookrightarrow_{\varsigma'}\rangle \varphi \vee (\mathrm{A}\,\delta \wedge \mathrm{R}\,\varsigma \wedge \langle \hookrightarrow_{\varsigma}\rangle \langle \mathrm{Mon}_{\delta,\varsigma}\rangle \varphi)\big)$

As mentioned in the text, the first formula indicates that the operation does not affect deduction with a rule different from the new one, and it follows from the first bisimilarity for monotonicity:

$$\text{if } \sigma \neq \varsigma', \text{ then } (M_{\mathrm{Mon}_{\delta,\varsigma}\hookrightarrow_{\sigma}}\ ,\ w) \underline{\leftrightarrow} (M_{\hookrightarrow_{\sigma}\mathrm{Mon}_{\delta,\varsigma}}\ ,\ w)$$

The second formula indicates how deduction with the generated rule changes after the structural operation, and it expresses the disjunction of two cases. If the rule created by the monotonicity operation was already in the original rule set, then the monotonicity operation is irrelevant, and just deduction is needed. This follows from the second bisimilarity for this structural operation:

$$\text{If } \varsigma' \in \mathsf{R}(w), \text{ then } (M_{\mathrm{Mon}_{\delta,\varsigma}\hookrightarrow_{\varsigma'}}\ ,\ w) \underline{\leftrightarrow} (M_{\hookrightarrow_{\varsigma'}}\ ,\ w)$$

But if the created rule was not in the original set, then we need some requirements, as the third bisimilarity for the operation shows:

$$\text{If } \delta \in \mathsf{A}(w) \text{ and } \varsigma \in \mathsf{R}(w), \text{ then } (M_{\mathrm{Mon}_{\delta,\varsigma}\hookrightarrow_{\varsigma'}}\ ,\ w) \underline{\leftrightarrow} (M_{\hookrightarrow_{\varsigma}\mathrm{Mon}_{\delta,\varsigma}}\ ,\ w)$$

A.6 Closure of explicit observation operation

This section proves that $\mathbf{IE}_K$-models are closed under the *explicit observation operation* of Definition 2.20 (Proposition 2.3).

Let $M = \langle W, R, V, \mathsf{A}, \mathsf{R}\rangle$ be a model in $\mathbf{IE}_K$. To show that $M_{\chi!^+}$ is also in $\mathbf{IE}_K$, we will show that it satisfies equivalence, coherence and truth for formulas and coherence and truth for rules.

Equivalence is immediate since the new model is a sub-model of the original one. For the other properties, suppose χ is a formula. *Coherence for formulas*

of $M_{\chi!^+}$ follows from that of M and the fact that χ is added uniformly to all preserved worlds. *Coherence for rules* of $M_{\chi!^+}$ follows simply from that of M. *Truth for formulas* of $M_{\chi!^+}$ follows from that of M since the truth of formulas in A-sets depends only on the atomic valuation of each world, which is not modified by the operation, and from the fact that the preserved worlds are precisely those in which χ is true. *Truth for rules* of $M_{\chi!^+}$ simply relies on that of M, again because the truth-value of the translation of a rule depends only on the unmodified atomic valuation of each world. The argument for the case in which χ is a rule is similar.

A.7 Explicit observation and deduction

This section provides a sketch for the proof of the validities of Table 2.10.

Just as the case of structural operations and deduction, the validity of the formulas follows from bisimilarities, this time the ones stated in Table A.2.

If χ is a formula:	
If $\chi \notin \mathsf{pm}(\sigma)$, then	$(M_{\chi!^+\hookrightarrow_\sigma}, w) \leftrightarrows (M_{\hookrightarrow_\sigma \chi!^+}, w)$
If $\chi \in \mathsf{pm}(\sigma)$ and $\chi \in \mathsf{A}(w)$, then	$(M_{\chi!^+\hookrightarrow_\sigma}, w) \leftrightarrows (M_{\hookrightarrow_\sigma \chi!^+}, w)$
If χ is a rule:	
If $\chi \neq \sigma$, then	$(M_{\chi!^+\hookrightarrow_\sigma}, w) \leftrightarrows (M_{\hookrightarrow_\sigma \chi!^+}, w)$
If $\chi = \sigma$ and $\chi \in \mathsf{R}(w)$, then	$(M_{\chi!^+\hookrightarrow_\sigma}, w) \leftrightarrows (M_{\hookrightarrow_\sigma \chi!^+}, w)$.

Table A.2: Bisimilarities for deduction and explicit observation operations

The proof is also similar to the case of structural operations and deduction, keeping in mind that explicit observations remove worlds, therefore modifying accessibility relations.

A.8 Awareness as a full language

This section provides the proofs of Lemmas 4.1 and 4.2.

Lemma 4.1 states that if an agent i has the formula φ at her disposal, that is, if $(M, w) \Vdash {}^{[i]}\varphi$, then she has at her disposal all atoms in it, that is, $(M, w) \Vdash {}^{[i]}p$ for every $p \in \mathsf{atm}(\varphi)$. In other words, the formula

$$ {}^{[i]}\varphi \to {}^{[i]}p $$

is valid for every $p \in \mathsf{atm}(\varphi)$.

We will prove the equivalent statement:

$$(M, w) \models {}^{[i]}\varphi \quad \text{implies} \quad \mathsf{atm}(\varphi) \subseteq \mathsf{PA}_i(w)$$

The equivalence of the statements follows from the semantic interpretation of formulas of the form ${}^{[i]}p$ (Definition 4.5).

The proof is by induction on φ. The base case is immediate: if φ is an atom p and $(M, w) \Vdash {}^{[i]}p$, then the semantic interpretation gives us $p \in \mathsf{PA}_i(w)$, hence $\mathsf{atm}(p) \subseteq \mathsf{PA}_i(w)$. For the inductive cases we have the following.

φ as ${}^{[j]}p$. Suppose $(M, w) \Vdash {}^{[i]}({}^{[j]}p)$. From Definition 4.3 we have the validity ${}^{[i]}({}^{[j]}p) \leftrightarrow {}^{[i]}p$ so $(M, w) \Vdash {}^{[i]}p$; then $p \in \mathsf{PA}_i(w)$ and hence $\mathsf{atm}({}^{[j]}p) \subseteq \mathsf{PA}_i(w)$.

φ as $\mathrm{A}_j\,\psi$. Suppose $(M, w) \Vdash {}^{[i]}(\mathrm{A}_j\,\psi)$. Definition 4.3 gives us ${}^{[i]}(\mathrm{A}_j\,\psi) \leftrightarrow {}^{[i]}\psi$ so $(M, w) \Vdash {}^{[i]}\psi$. But ψ is a sub-formula of $\mathrm{A}_j\,\psi$; then, by inductive hypothesis, $\mathsf{atm}(\psi) \subseteq \mathsf{PA}_i(w)$, and hence $\mathsf{atm}(\mathrm{A}_j\,\psi) \subseteq \mathsf{PA}_i(w)$.

φ as $\mathrm{R}_j\,\rho$. Suppose $(M, w) \Vdash {}^{[i]}(\mathrm{R}_j\,\rho)$. Definition 4.3 gives us ${}^{[i]}(\mathrm{R}_j\,\rho) \leftrightarrow {}^{[i]}\rho$ so $(M, w) \Vdash {}^{[i]}\rho$. But ρ is a sub-formula of $\mathrm{R}_j\,\rho$; then, by inductive hypothesis, $\mathsf{atm}(\rho) \subseteq \mathsf{PA}_i(w)$, and hence $\mathsf{atm}(\mathrm{R}_j\,\rho) \subseteq \mathsf{PA}_i(w)$.

φ as $\neg\psi$. Suppose $(M, w) \Vdash {}^{[i]}(\neg\psi)$. Definition 4.3 gives us ${}^{[i]}(\neg\psi) \leftrightarrow {}^{[i]}\psi$ so $(M, w) \Vdash {}^{[i]}\psi$. Since ψ is a sub-formula of $\neg\psi$, inductive hypothesis gives us $\mathsf{atm}(\psi) \subseteq \mathsf{PA}_i(w)$ and hence $\mathsf{atm}(\neg\psi) \subseteq \mathsf{PA}_i(w)$.

φ as $\psi \vee \psi'$. Suppose $(M, w) \Vdash {}^{[i]}(\psi \vee \psi')$. Definition 4.3 gives us ${}^{[i]}(\psi \vee \psi') \leftrightarrow ({}^{[i]}\psi \wedge {}^{[i]}\psi')$ so $(M, w) \Vdash ({}^{[i]}\psi \wedge {}^{[i]}\psi')$, that is, $(M, w) \Vdash {}^{[i]}\psi$ and $(M, w) \Vdash {}^{[i]}\psi'$. Since ψ and ψ' are sub-formulas of $\psi \vee \psi'$, inductive hypothesis gives us $(\mathsf{atm}(\psi) \cup \mathsf{atm}(\psi')) \subseteq \mathsf{PA}_i(w)$ and hence $\mathsf{atm}(\psi \vee \psi') \subseteq \mathsf{PA}_i(w)$.

φ as $\Box_j\psi$. Suppose $(M, w) \Vdash {}^{[i]}(\Box_j\psi)$. Definition 4.3 gives us ${}^{[i]}(\Box_j\psi) \leftrightarrow {}^{[i]}\psi$ so $(M, w) \Vdash {}^{[i]}\psi$. But ψ is a sub-formula of $\Box_j\psi$; then, by inductive hypothesis, $\mathsf{atm}(\psi) \subseteq \mathsf{PA}_i(w)$, and hence $\mathsf{atm}(\Box_j\psi) \subseteq \mathsf{PA}_i(w)$.

This completes the proof.

Lemma 4.2 states that if agent i has all atoms in $\{p_1, \ldots, p_n\}$ at her disposal, that is, if $(M, w) \Vdash {}^{[i]}p_k$ for every $k \in \{1, \ldots, n\}$, then she has at her disposal any formula built from such atoms, that is, $(M, w) \Vdash {}^{[i]}\varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. In other words, the formula

$$\Big(\bigwedge_{k \in \{1,\ldots,n\}} {}^{[i]}p_k \Big) \rightarrow {}^{[i]}\varphi$$

is valid for every φ built from $\{p_1, \ldots, p_n\}$.

Again, we will prove an equivalent statement, this time:

$$\{p_1, \ldots, p_n\} \subseteq \mathsf{PA}_i(w) \quad \text{implies} \quad (M, w) \Vdash {}^{[i]}\varphi$$

for every φ built from $\{p_1, \ldots, p_n\}$. Again, the equivalence of the statements follows from the semantic interpretation of formulas of the form ${}^{[i]}p$.

So suppose $\{p_1, \ldots, p_n\} \subseteq \mathsf{PA}_i(w)$; we will proceed by induction on φ. The base case is immediate: if φ is any p_i in $\{p_1, \ldots, p_n\}$, then we obviously have $(M, w) \Vdash {}^{[i]}p_i$. For the inductive cases we have the following.

φ as ${}^{[j]}p$. In this case p should be in $\{p_1, \ldots, p_n\}$ so $(M, w) \Vdash {}^{[i]}p$. But by Definition 4.3 we have ${}^{[i]}({}^{[j]}p) \leftrightarrow {}^{[i]}p$, so $(M, w) \Vdash {}^{[i]}({}^{[j]}p)$.

φ as $\mathbf{A}_j\,\psi$. Since $\mathrm{A}_j\,\psi$ is built from atoms in $\{p_1, \ldots, p_n\}$, we have $\mathsf{atm}(\mathrm{A}_j\,\psi) \subseteq \{p_1, \ldots, p_n\}$, that is, $\mathsf{atm}(\psi) \subseteq \{p_1, \ldots, p_n\}$. Since ψ is a sub-formula of $\mathrm{A}_j\,\psi$, inductive hypothesis gives us $(M, w) \Vdash {}^{[i]}\psi$. But by Definition 4.3 we have ${}^{[i]}(\mathrm{A}_j\,\psi) \leftrightarrow {}^{[i]}\psi$, so $(M, w) \Vdash {}^{[i]}(\mathrm{A}_j\,\psi)$.

φ as $\mathbf{R}_j\,\rho$. We have $\mathsf{atm}(\mathrm{R}_j\,\rho) \subseteq \{p_1, \ldots, p_n\}$, that is, $\mathsf{atm}(\rho) \subseteq \{p_1, \ldots, p_n\}$. Since ρ is a sub-formula of $\mathrm{R}_j\,\rho$, inductive hypothesis gives us $(M, w) \Vdash {}^{[i]}\rho$. But by Definition 4.3 we have ${}^{[i]}(\mathrm{R}_j\,\rho) \leftrightarrow {}^{[i]}\rho$, so $(M, w) \Vdash {}^{[i]}(\mathrm{R}_j\,\rho)$.

φ as $\neg\psi$. We have $\mathsf{atm}(\psi) \subseteq \{p_1, \ldots, p_n\}$. Since ψ is a sub-formula of $\neg\psi$, inductive hypothesis gives us $(M, w) \Vdash {}^{[i]}\psi$. But by Definition 4.3 we have ${}^{[i]}(\neg\psi) \leftrightarrow {}^{[i]}\psi$, so $(M, w) \Vdash {}^{[i]}(\neg\psi)$.

φ as $\psi \vee \psi'$. We have $(\mathsf{atm}(\psi) \cup \mathsf{atm}(\psi')) \subseteq \{p_1, \ldots, p_n\}$. Since ψ and ψ' are both sub-formulas of $\psi \vee \psi'$, inductive hypothesis yields $(M, w) \Vdash {}^{[i]}\psi \wedge {}^{[i]}\psi'$. Definition 4.3 gives us ${}^{[i]}(\psi \vee \psi') \leftrightarrow ({}^{[i]}\psi \wedge {}^{[i]}\psi')$ so we have $(M, w) \Vdash {}^{[i]}(\psi \vee \psi')$.

φ as $\Box_j\psi$. We have $\mathsf{atm}(\psi) \subseteq \{p_1, \ldots, p_n\}$. Since ψ is a sub-formula of $\Box_j\psi$, inductive hypothesis gives us $(M, w) \Vdash {}^{[i]}\psi$. But by Definition 4.3 we have ${}^{[i]}(\Box_j\psi) \leftrightarrow {}^{[i]}\psi$, so $(M, w) \Vdash {}^{[i]}(\Box_j\psi)$.

Note how this case does not involve availability at worlds other than w. It simply says that if $\Box_j\psi$ is a formula built from atoms the agent has *locally* available at w, then all atoms in ψ should be *locally* available. By inductive hypothesis, ψ should be *locally* available, ${}^{[i]}\psi$, and therefore by Definition 4.3, $\Box_j\psi$ is also *locally* available, ${}^{[i]}(\Box_j\psi)$.

This completes the proof.

A.9 Upgrade and locally well-preorders

This section provides the proof of Proposition 5.11.

We need to show that if the relation $\leq$ is a locally well-preorder, so is the relation $\leq'$ defined as

$$\leq' := \underbrace{(\leq; \chi?)}_{(1)} \cup \underbrace{(\neg\chi?; \leq)}_{(2)} \cup \underbrace{(\neg\chi?; \sim; \chi?)}_{(3)}$$

In words, we have $w \leq' u$ if and only if in M (1) $w \leq u$ and u is a χ-world, or (2) w is a $\neg\chi$-world and $w \leq u$, or (3) w is a $\neg\chi$-world, u is a χ-world and the two worlds are in the same comparability class. Note that the only case in which we do not have $w \leq' u$ is when, in M, w is a χ-world and u is a $\neg\chi$-world.

The key observation is that a locally well-preorder is a locally connected and conversely well-founded preorder. We will prove that if $\leq$ satisfies such properties, so does $\leq'$ defined as above.

For *reflexivity*, pick any $w \in W$. Since $\leq$ is reflexive, we have $w \leq w$. Now, w is either a χ-world or a $\neg\chi$-one. In the first case we get $w \leq' w$ from part (1) of the definition of $\leq'$; in the second case we get it from part (2).

For *transitivity*, suppose $w \leq' u$ and $u \leq' v$ and consider w. If it is a χ-world, then so is u (otherwise there would not be a link from w to u) and hence so is v too; therefore, by part (1) of the definition, $w \leq' v$. If it is a $\neg\chi$-world, part (2) of the definition gives us $w \leq' v$. Hence, $\leq'$ is transitive.

For *local connectedness*, first we will show that, for every u_1, u_2 in W, we have $u_1 \sim u_2$ if and only if $u_1 \sim' u_2$.
($\Longrightarrow$) Suppose $u_1 \sim u_2$. If we have $u_1 \leq' u_2$, then we have $u_1 \sim' u_2$ and we are done. Otherwise, u_1 should be a χ-world in M and u_2 should be a $\neg\chi$-world in M; this together with $u_1 \sim u_2$ gives us $u_2 \leq' u_1$ by part (3) of the definition, and hence we have $u_1 \sim' u_2$.
($\Longleftarrow$) If $u_1 \sim' u_2$, then we have $u_1 \leq' u_2$ or $u_2 \leq' u_1$. Consider the first case, and let us review the three possibilities. If we have $u_1 \leq' u_2$ because of part (1) of the definition of $\leq'$, then we have $(u_1, u_2) \in (\leq; \chi?)$; hence $u_1 \leq u_2$ and therefore $u_1 \sim u_2$. If it is because of part (2), then we have $(u_1, u_2) \in (\neg\chi?; \leq)$; hence $u_1 \leq u_2$ and therefore $u_1 \sim u_2$. If it is because of part (3), then we have $(u_1, u_2) \in (\neg\chi?; \sim; \chi?)$; hence $u_1 \sim u_2$. In the three possibilities we get the required $u_1 \sim u_2$. The second case is analogous.

Now, to show local connectedness, take any $w \in W$ and pick u_1, u_2 in V_w under $\leq'$. By definition of V_w we have $w \sim' u_1$ and $w \sim' u_2$; by the just proved property we get $w \sim u_1$ and $w \sim u_2$; by local connectedness of $\leq$ we have $u_1 \sim u_2$ and then by the just proved property again we get the required $u_1 \sim' u_2$.

For *converse well-foundedness* we proceed by contradiction. Suppose that there is an infinite ascending chain $u_1 <' u_2 <' \cdots$. These worlds are either χ or

$\neg\chi$-worlds in the original model. Since the chain is infinite, there must be an infinite sub-chain of either χ or $\neg\chi$-worlds (we cannot have an alternation from a χ-world to a $\neg\chi$-one because of the definition of $\leq'$). But inside these areas, the new relation is the old one, contradicting the converse well-foundedness of $\leq$. Then, such infinite chain cannot exists, and therefore $\leq'$ is conversely well-founded.

This completes the proof.

A.10 Product update and locally well-preorders

This section provides the proofs of Proposition 5.12.

We will show that if $\leq$ and $\preccurlyeq$ are two locally well-preorders over W and E respectively, so is the relation $\leq'$ over $W \times E$ given by

$$(w_1, e_1) \leq' (w_2, e_2) \quad \text{iff} \quad \underbrace{\Big(e_1 \prec e_2 \text{ and } w_1 \sim w_2\Big)}_{(1)} \text{ or } \underbrace{\Big(e_1 \cong e_2 \text{ and } w_1 \leq w_2\Big)}_{(2)}$$

Recall that a locally well-preorder is a locally connected and a conversely well-founded preorder

For *reflexivity*, take any $(w, e) \in W \times E$. By reflexivity of $\leq$ and $\preccurlyeq$, we have $w \leq w$ and $e \preccurlyeq e$. Then $w \leq w$ and $e \cong e$ and hence $(w, e) \leq' (w, e)$ from (2) of the definition of $\leq'$.

For *transitivity*, suppose $(w_1, e_1) \leq' (w_2, e_2)$ and $(w_2, e_2) \leq' (w_3, e_3)$. According to the definition of $\leq'$, each one of these two inequalities has two possible reasons, and this gives us four cases. We will prove two of them in detail; the other two can be proved in a similar way.

1. Suppose that both $(w_1, e_1) \leq' (w_2, e_2)$ and $(w_2, e_2) \leq' (w_3, e_3)$ hold because of part (1) of the definition of $\leq'$. Then we have

   $$e_1 \prec e_2, \qquad w_1 \sim w_2, \qquad e_2 \prec e_3, \qquad w_2 \sim w_3.$$

   By unfolding the definitions of $\prec$ and $\sim$ we get

   $$e_1 \preccurlyeq e_2, e_1 \not\preccurlyeq e_2, \quad \begin{cases} w_1 \leq w_2 \\ w_2 \leq w_1 \end{cases}, \quad e_2 \preccurlyeq e_3, e_3 \not\preccurlyeq e_2, \quad \begin{cases} w_2 \leq w_3 \\ w_3 \leq w_2 \end{cases}.$$

   For the action part, recall that $\preccurlyeq$ is transitive. Then we have $e_1 \preccurlyeq e_3$. We also have $e_3 \not\preccurlyeq e_1$ because otherwise from $e_1 \preccurlyeq e_2$ we will get $e_3 \preccurlyeq e_2$, contradicting part of the assumptions. Then we have $e_1 \prec e_3$.

For the static part we have again four possibilities. If we have $w_1 \leq w_2$ and $w_2 \leq w_3$, then we get $w_1 \leq w_3$ by transitivity of $\leq$, and hence $w_1 \sim w_3$. If we have $w_1 \leq w_2$ and $w_3 \leq w_2$, then we should have $w_1 \leq w_3$ or $w_3 \leq w_1$ because $\leq$ is locally connected; hence $w_1 \sim w_3$. In the other two cases, a similar reasoning shows that $w_1 \sim w_3$ holds too.

Then, part (1) of the definition of $\leq'$ gives us $(w_1, e_1) \leq' (w_3, e_3)$.

2. Suppose that while $(w_1, e_1) \leq' (w_2, e_2)$ holds because of part (1) of the definition of $\leq'$, $(w_2, e_2) \leq' (w_3, e_3)$ holds because of part (2). Then

$$e_1 \prec e_2, \qquad w_1 \sim w_2, \qquad e_2 \cong e_3, \qquad w_2 \leq w_3.$$

By unfolding the definitions we get

$$e_1 \preccurlyeq e_2, e_1 \not\preccurlyeq e_2, \quad \begin{cases} w_1 \leq w_2 \\ w_2 \leq w_1 \end{cases}, \quad e_2 \preccurlyeq e_3, e_3 \preccurlyeq e_2, \quad w_2 \leq w_3,$$

For the action part, recall that $\preccurlyeq$ is transitive. Then we have $e_1 \preccurlyeq e_3$. We also have $e_3 \not\preccurlyeq e_1$ because otherwise from $e_2 \preccurlyeq e_3$ we will get $e_2 \preccurlyeq e_1$, contradicting part of the assumptions. Then we have $e_1 \prec e_3$.

For the static part we have two possibilities. If we have $w_1 \leq w_2$, then together with $w_2 \leq w_3$ we get $w_1 \leq w_3$; hence $w_1 \sim w_3$. If we have $w_2 \leq w_1$ then from $w_2 \leq w_3$ we should have $w_1 \leq w_3$ or $w_3 \leq w_1$ because $\leq$ is locally connected; hence $w_1 \sim w_3$.

Then, part (1) of the definition gives us $(w_1, e_1) \leq' (w_3, e_3)$.

For *local connectedness*, first we will show that, for every w_1, w_2 in W and e_1, e_2 in E, we have $w_1 \sim w_2$ and $e_1 \approx e_2$ if and only if $(w_1, e_1) \sim' (w_2, e_2)$.
($\Longrightarrow$) If $w_1 \sim w_2$ and $e_1 \approx e_2$, then we have $w_1 \leq w_2$ or $w_2 \leq w_1$, and $e_1 \preccurlyeq e_2$ or $e_2 \preccurlyeq e_1$. This gives us four cases. For example, suppose $w_1 \leq w_2$ and $e_2 \preccurlyeq e_1$. If we have $e_1 \preccurlyeq e_2$ we get $e_1 \cong e_2$; then by part (2) of the definition we get $(w_1, e_1) \leq' (w_2, e_2)$ and hence $(w_1, e_1) \sim' (w_2, e_2)$. If we do not have $e_1 \preccurlyeq e_2$, then we have $e_2 \prec e_1$ and from $w_1 \leq w_2$ we already have $w_1 \sim w_2$; then by part (1) of the definition we have $(w_2, e_2) \leq' (w_1, e_1)$ and hence $(w_1, e_1) \sim' (w_2, e_2)$. The other three cases can be proved in a similar way.
($\Longleftarrow$) If $(w_1, e_1) \sim' (w_2, e_2)$, then $(w_1, e_1) \leq' (w_2, e_2)$ or $(w_2, e_2) \leq' (w_1, e_1)$. In the first case, we have either possibility (1), which gives us $e_1 \prec e_2$ and $w_1 \sim w_2$, hence $e_1 \approx e_2$ and $w_1 \sim w_2$, or else possibility (2), which gives us $e_1 \cong e_2$ and $w_1 \leq w_2$, hence $e_1 \approx e_2$ and $w_1 \sim w_2$. The second case is similar.

Now, to show local connectedness, take any $(w, e) \in (W \times E)$ and pick $(w_1, e_1), (w_2, e_2)$ in $V_{(w,e)}$ under $\leq'$. By definition of $V_{(w,e)}$ we have $(w, e) \sim' (w_1, e_1)$ and $(w, e) \sim' (w_2, e_2)$; by the just proved property we get $w \sim w_1$, $e \approx e_1$, $w \sim w_2$

and $e \approx e_2$; by local connectedness of $\leq$ and $\preccurlyeq$ we have $w_1 \sim w_2$ and $e_1 \approx e_2$ and then by the just proved property again we get the required $(w_1, e_1) \sim' (w_2, e_2)$.

For *converse well-foundedness* we proceed by contradiction. Suppose that there is an infinite ascending chain $(w_1, e_1) <' (w_2, e_2) <' \cdots$. Consider the infinite chain $e_1, e_2, \ldots$: if there is an infinite number of pairs e_i and e_{i+1} for which the plausibility order is strict, that is, if $e_i \prec e_{i+1}$ happens infinitely often, then we have an infinite ascending chain in E, contradicting the converse well-foundedness of $\preccurlyeq$. On the other hand, if $e_i \prec e_{i+1}$ only happens finitely often, then from some moment on we have only equal plausibility, that is, from some moment on we have $e_i \cong e_{i+1}$. But then, from that moment on, we have $w_i < w_{i+1}$, which is an infinite ascending chain in W, contradicting the converse well-foundedness of $\leq$. Then, the infinite chain in $W \times E$ cannot exists, and hence $\leq'$ is conversely well-founded.

This completes the proof.

Bibliography

T. Ågotnes. *A Logic of Finite Syntactic Epistemic States*. PhD thesis, Department of Informatics, University of Bergen, Bergen, Norway, 2004. Cited on page **17**.

T. Ågotnes and N. Alechina. Full and relative awareness: a decidable logic for reasoning about knowledge of unawareness. In Samet (2007), pages 6–14. Cited on page **77**.

T. Ågotnes and N. Alechina, editors. *Special issue on Logics for Resource Bounded Agents*, 2009. *Journal of Logic, Language and Information*, 18(1). Cited on page **2**.

C. E. Alchourrón and D. Makinson. The logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982. Cited on page **160**.

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2):510–530, 1985. Cited on pages **16, 124, and 125**.

A. Aliseda. *Abductive Reasoning. Logical Investigations into Discovery and Explanation*, volume 330 of *Synthese Library Series*. Springer, 2006. Cited on page **156**.

D. Angluin and C. H. Smith. Inductive inference: Theory and methods. *ACM Computing Surveys*, 15(3):237–269, 1983. ISSN 0360-0300. DOI: 10.1145/356914.356918. Cited on page **173**.

C. Areces and D. Figueira. Which semantics for neighbourhood semantics? In C. Boutilier, editor, *IJCAI 2009*, pages 671–676, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc. Cited on page **11**.

S. N. Artemov and E. Nogina. Introducing justification to epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, 2005. Cited on pages **12 and 185**.

G. Aucher. A combined system for update logic and belief revision. Master's thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands, 2003. URL: `http://www.illc.uva.nl/Publications/ResearchReports/MoL-2003-03.text.pdf`. ILLC Master of Logic Thesis Series MoL-2003-03. Cited on page **131**.

R. J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8(1):6–19, 1995. ISSN 0899-8256. DOI: 10.1016/S0899-8256(05)80015-6. Cited on page **177**.

R. J. Aumann and A. Brandenburger. Epistemic conditions for nash equilibrium. *Econometrica*, 63(5):1161–1180, Sept. 1995. URL: `http://www.jstor.org/stable/2171725`. Cited on page **177**.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, Amsterdam, The Netherlands, 2008. ISBN 978-90 8964 026 0. Cited on pages **13, 112, 114, 117, 118, 125, 131, 139, 142, and 182**.

A. Baltag and S. Smets. Learning by questions and answers: From belief-revision cycles to doxastic fixed points. In H. Ono, M. Kanazawa, and R. J. G. B. de Queiroz, editors, *WoLLIC*, volume 5514 of *Lecture Notes in Computer Science*, pages 124–139. Springer, 2009. ISBN 978-3-642-02260-9. DOI: 10.1007/978-3-642-02261-6_11. Cited on pages **173 and 184**.

A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicious. Technical Report SEN-R9922, CWI, Amsterdam, 1999. Cited on pages **13, 70, 131, 134, and 182**.

J. Barwise. Three views of common knowledge. In Vardi (1988), pages 365–379. ISBN 0-934613-66-4. Cited on page **3**.

J. van Benthem. Reflections on epistemic logic. *Logique et Analyse*, 133–134(34): 5–14, 1993. Cited on page **12**.

J. van Benthem. Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128:49–76, Mar. 2006. DOI: 10.1007/s11098-005-4052-0. Cited on page **3**.

J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007. Cited on pages **13, 68, 112, 114, 125, 126, and 182**.

J. van Benthem. Logic and reasoning: Do the facts matter? *Studia Logica*, 88(1): 67–84, Feb. 2008a. ISSN 0039-3215. DOI: 10.1007/s11225-008-9101-1. Cited on page **23**.

J. van Benthem. Tell it like it is: Information flow in logic. *Journal of Peking University (Humanities and Social Science Edition)*, 1:80–90, 2008b. Cited on page **24**.

J. van Benthem. Merging observation and access in dynamic logic. *Journal of Logic Studies*, 1(1):1–17, 2008c. Cited on pages **101, 106, and 119**.

J. van Benthem. Tell it like it is: Information flow in logic. *Journal of Peking University (Humanities and Social Science Edition)*, 1:80–90, 2008d. Cited on page **12**.

J. van Benthem. Logic, mathematics, and general agency. In P. Bour, M. Rebuschi, and L. Rollet, editors, *Festschrift for Gerhard Heinzmann*. Laboratoire d'histoire des ceinces et de la philosophie, Nancy, 2009. Cited on page **164**.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2010. To appear. Cited on page **67**.

J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007. Cited on pages **13, 68, 117, and 128**.

J. van Benthem and M. Martínez. The stories of logic and information. In P. Adriaans and J. van Benthem, editors, *Philosophy Of Information*, Handbook of the Philosophy of Science, pages 217–280. North-Holland, Amsterdam, 2008. ISBN 978-0-444-51726-5. Cited on pages **3, 10, 12, and 186**.

J. van Benthem and S. Minică. Toward a dynamic logic of questions. In He et al. (2009), pages 27–41. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7_3. Cited on pages **167, 168, and 169**.

J. van Benthem and F. R. Velázquez-Quesada. The dynamics of awareness. *Synthese (Knowledge, Rationality and Action)*, 177(0):5–27, 2010. ISSN 0039-7857. DOI: 10.1007/s11229-010-9764-9. Cited on page **21**.

J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, Nov. 2006. ISSN 0890-5401. DOI: 10.1016/j.ic.2006.04.006. Cited on pages **70, 71, 75, 133, and 139**.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic*. Number 53 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, New York, USA, 2001. ISBN 0-521-80200-8. Cited on pages **118, 187, and 188**.

O. Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49 (1):49–80, Oct. 2004. DOI: 10.1016/j.geb.2003.10.006. Cited on page **114**.

O. Board and K.-S. Chung. Object-based unawareness. In Bonanno et al. (2006), pages 35–41. Cited on page **77**.

O. Boissier, A. E. F. Seghrouchni, S. Hassas, and N. Maudet, editors. *Proceedings of The Multi-Agent Logics, Languages, and Organisations Federated Workshops (MALLOW 2010)*, volume 627, Lyon, France, Aug. 2010. CEUR Workshop Proceedings. URL: `http://ceur-ws.org/Vol-627`. Cited on pages **213 and 214**.

G. Bonanno, W. van der Hoek, , and M. Wooldridge, editors. *Proceedings of The 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT)*, July 2006. Cited on pages **204 and 213**.

C. Boutilier. Toward a logic for qualitative decision theory. In J. Doyle, E. Sandewall, and P. Torasso, editors, *KR 94*, pages 75–86, Bonn, Germany, May 1994a. Morgan Kaufmann. ISBN 1-55860-328-X. Cited on page **117**.

C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68(1):87–154, 1994b. Cited on page **114**.

C. Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68(1):33–85, 1994c. Cited on page **154**.

J. P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II, chapter 2, pages 89–133. Reidel, 1984. Cited on page **115**.

R. Carnap. *The Continuum of Inductive Methods*. University of Chicago Press, Chicago, 1952. Cited on page **10**.

B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, Mass., 1980. Cited on page **11**.

Y.-C. Chen, N. V. Long, and X. Luo. Iterated strict dominance in general games. *Games and Economic Behavior*, 61(2):299–315, 2007. ISSN 0899-8256. DOI: 10.1016/j.geb.2007.02.002. Cited on page **177**.

M. D'Agostino and L. Floridi. The enduring scandal of deduction. is propositional logic really uninformative? *Synthese*, 167(2):271–315, Mar. 2009. DOI: 10.1007/s11229-008-9409-4. Cited on page **148**.

C. Dégremont and N. Gierasimczuk. Can doxastic agents learn? on the temporal structure of learning. In He et al. (2009), pages 90–104. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7_8. Cited on page **173**.

E. Dekel, B. Lipman, and A. Rustichini. Standard state-space models preclude unawareness. *Econometrica*, 66(1):159–174, 1998. Cited on page **77**.

H. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147(2):229–275, 2005. ISSN 0039-7857. DOI: 10.1007/s11229-005-1349-7. Cited on pages **125 and 131**.

H. van Ditmarsch and T. French. Awareness and forgetting of facts and agents. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT 2009)*, Milan, 2009. Cited on pages **60, 77, 85, 90, 99, 100, 101, and 119**.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007. Cited on pages **13 and 40**.

H. van Ditmarsch, A. Herzig, J. Lang, and P. Marquis. Introspective forgetting. *Synthese (Knowledge, Rationality and Action)*, 169(2):405–423, July 2009. ISSN 0039-7857. DOI: 10.1007/s11229-009-9554-4. Cited on pages **60 and 77**.

K. Doets. *From Logic to Logic Programming*. The MIT Press, Cambridge, Mass., USA, 1994. Cited on page **16**.

J. J. Drapkin and D. Perlis. Step-logics: An alternative approach to limited reasoning. In *Proceedings of the European Conf. on Artificial Intelligence*, pages 160–163, Brighton, England, 1986. Cited on page **24**.

F. I. Dretske. *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA., 1981. Cited on page **35**.

H. N. Duc. Logical omniscience vs. logical ignorance. on a dilemma of epistemic logic. In C. A. Pinto-Ferreira and N. J. Mamede, editors, *EPIA 1995*, volume 990 of *Lecture Notes in Computer Science*, pages 237–248, Heilderberg, 1995. Springer. ISBN 3-540-60428-6. Cited on pages **17, 24, 25, and 119**.

H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997. Cited on pages **25, 101, and 179**.

H. N. Duc. *Resource-Bounded Reasoning about Knowledge*. PhD thesis, Institut für Informatik, Universität Leipzig, Leipzig, Germany, 2001. Cited on page **25**.

R. A. Eberle. A logic of believing, knowing, and inferring. *Synthese*, 26(3-4), 1974. Cited on page **8**.

J. van Eijck and Y. Wang. Propositional dynamic logic as a logic of belief revision. In W. Hodges and R. J. G. B. de Queiroz, editors, *WoLLIC*, volume 5110 of *Lecture Notes in Computer Science*, pages 136–148. Springer, 2008. ISBN 978-3-540-69936-1. DOI: 10.1007/978-3-540-69937-8_13. Cited on pages **126 and 139**.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988. ISSN 0004-3702. DOI: 10.1016/0004-3702(87)90003-8. Cited on pages **2, 55, 57, 58, 59, 61, 67, 76, 83, 87, 92, 100, 101, 119, and 181**.

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. The MIT Press, Cambridge, Mass., 1995. Cited on pages **2 and 179**.

Y. Feinberg. Subjective reasoning - games with unawareness. Research Papers 1875, Graduate School of Business, Stanford University, Nov. 2004. URL: `https://gsbapps.stanford.edu/researchpapers/detail1.asp?Document_ID=2584`. Cited on page **177**.

L. Flobbe, R. Verbrugge, P. Hendriks, and I. Krämer. Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, 17(4):417–442, 2008. DOI: 10.1007/s10849-008-9064-7. Cited on page **185**.

L. Floridi. Consciousness, agents and the knowledge game. *Minds and Machines*, 15(3-4):415–444, Nov. 2005. ISSN 0924-6495. DOI: 10.1007/s11023-005-9005-z. Cited on page **35**.

M. Franke. *Signal to Act: Game Theory in Pragmatics*. PhD thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands, Dec. 2009. ILLC Dissertation series DS-2009-11. Cited on page **172**.

A. Fuhrmann. Theory contraction through base contraction. *Journal of Philosophical Logic*, 20(2):175–203, May 1991. ISSN 0022-3611 (Print) 1573-0433 (Online). DOI: 10.1007/BF00284974. Cited on page **160**.

D. M. Gabbay. *Labelled Deductive Systems. Volume 1*. Number 33 in Oxford Logic Guides. Oxford University Press, Oxford, UK, 1996. Cited on page **9**.

P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Bradford Books & MIT Press, Cambridge, MA, 1988. Cited on page **12**.

P. Gärdenfors, editor. *Belief Revision*. Number 29 in Cambridge Tracts in Theoretical Computer Science. Cambridge Press, 1992. Cited on pages **9, 16, and 124**.

P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In Vardi (1988), pages 83–95. ISBN 0-934613-66-4. Cited on pages **125 and 164**.

P. Gärdenfors and H. Rott. Belief revision. In D. M. Gabbay, C. J. Hoger, and J. A. Robinson, editors, *Epistemic and Temporal Reasoning*, volume IV of *Handbook of Logic in Artificial Intelligence and Logic Programming*. Oxford University Press, 1995. Cited on pages **16 and 124**.

J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands, 1999. ILLC Dissertation Series DS-1999-01. Cited on pages **13, 61, 126, and 181**.

N. Gierasimczuk. Learning by erasing in dynamic epistemic logic. In A. H. Dediu, A.-M. Ionescu, and C. Martín-Vide, editors, *LATA*, volume 5457 of *Lecture Notes in Computer Science*, pages 362–373. Springer, 2009. ISBN 978-3-642-00981-5. DOI: 10.1007/978-3-642-00982-2_31. Cited on page **173**.

J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–101, 1987. ISSN 0304-3975. DOI: 10.1016/0304-3975(87)90045-4. URL: `http://iml.univ-mrs.fr/~girard/linear.pdf`. Cited on page **32**.

P. Girard. *Modal Logic for Belief and Preference Change*. PhD thesis, Department of Philosophy, Stanford University, Stanford, CA, USA, Feb. 2008. ILLC Dissertation Series DS-2008-04. Cited on page **117**.

E. M. Gold. Language identification in the limit. *Information and Control*, 10 (5):447–474, 1967. URL: `http://www.isrl.uiuc.edu/~amag/langev/paper/gold67limit.html`. Cited on page **173**.

P. Grice. *Studies in the Ways of Words*. Harvard University Press, Cambridge, Mass., 1989. Cited on page **170**.

J. Groenendijk. Inquisitive semantics: Two possibilities for disjunction. In P. Bosch, D. Gabelaia, and J. Lang, editors, *TbiLLC*, volume 5422 of *Lecture Notes in Computer Science*, pages 80–94. Springer, 2007. ISBN 978-3-642-00664-7. DOI: 10.1007/978-3-642-00665-4_8. Cited on page **167**.

D. Grossi. A note on brute vs. institutional facts: Modal logic of equivalence up to a signature. In G. Boella, P. Noriega, G. Pigozzi, and H. Verhagen, editors,

Normative Multi-Agent Systems, number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. URL: `http://drops.dagstuhl.de/opus/volltexte/2009/1910`. Cited on page **99**.

D. Grossi and F. R. Velázquez-Quesada. *Twelve Angry Men*: A study on the fine-grain of announcements. In He et al. (2009), pages 147–160. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7_12. Cited on page **21**.

D. Grossi and F. R. Velázquez-Quesada. *Twelve Angry Men*: A dynamic-epistemic study of awareness, implicit and explicit information. In L. Kurzen, D. Grossi, and F. R. Velázquez-Quesada, editors, *Logic and Interactive RAtionality*, pages 42–68. Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2010. Cited on page **21**.

A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17 (2):157–170, May 1988. ISSN 0022-3611. DOI: 10.1007/BF00247909. Cited on pages **114, 125, and 137**.

M. Guo and Z. Xiong. A dynamic preference logic with issue-management. Manuscript, 2010. Cited on page **169**.

J. Y. Halpern, editor. *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge, Monterey, CA, March 1986*, San Francisco, CA, USA, 1986. Morgan Kaufmann Publishers Inc. ISBN 0-934613-04-4. Cited on pages **211 and 214**.

J. Y. Halpern. Alternative semantics for unawareness. *Games and Economic Behavior*, 37:321–339, 2001. Cited on pages **57 and 77**.

J. Y. Halpern and R. Pucella. Dealing with logical omniscience. In *TARK '07: Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 169–176, New York, NY, USA, 2007. ACM. DOI: 10.1145/1324249.1324273. Cited on page **2**.

J. Y. Halpern and L. C. Rêgo. Interactive unawareness revisited. In R. van der Meyden, editor, *TARK*, pages 78–91. National University of Singapore, 2005. ISBN 981-05-3412-4. Cited on page **77**.

J. Y. Halpern and L. C. Rêgo. Reasoning about knowledge of unawareness revisited. In A. Heifetz, editor, *TARK*, pages 166–173, 2009. Cited on page **77**.

Y. Hamami. The interrogative model of inquiry meets dynamic epistemic logics. Master's thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands,

2010a. URL: http://www.illc.uva.nl/Publications/ResearchReports/MoL-2010-04.text.pdf. ILLC Master of Logic Thesis Series MoL-2010-04. Cited on page **169**.

Y. Hamami, June 2010b. Private communication. Cited on page **108**.

S. O. Hansson. New operators for theory change. *Theoria*, 55:114–132, 1989. Cited on page **160**.

S. O. Hansson. In defense of base contraction. *Synthese*, 91(3):239–245, 1992. DOI: 10.1007/BF00413568. Cited on page **160**.

D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, MA, 2000. ISBN 0-262-08289-6. Cited on page **26**.

C. Hartshorne and P. Weiss, editors. *Collected Papers of Charles S. Peirce*, volume V: Pragmatism and Pramaticism. Harvard Universit Press, Cambridge, 1934. ISBN 9780674138001. Cited on page **156**.

X. He, J. F. Horty, and E. Pacuit, editors. *Logic, Rationality, and Interaction, Second International Workshop, LORI 2009, Chongqing, China, October 8-11, 2009. Proceedings*, volume 5834 of *Lecture Notes in Computer Science*, 2009. Springer. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7. Cited on pages **203, 205, 208, 211, and 214**.

A. Heifetz, M. Meier, and B. C. Schipper. Multi-person unawareness. In J. Y. Halpern and M. Tennenholtz, editors, *TARK*, pages 145–158. ACM, 2003. ISBN 1-58113-731-1. Cited on page **77**.

L. Henkin. Completeness in the theory of types. *The Journal of Symbolic Logic*, 15 (2):81–91, 1950. URL: http://projecteuclid.org/euclid.jsl/1183730860. Cited on page **7**.

B. Hill. Awareness dynamics. *Journal of Philosophical Logic*, 39(2):113–137, Apr. 2010. Cited on page **77**.

J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, N.Y., 1962. Cited on pages **2, 4, and 179**.

J. Hintikka. *Logic, language games and information. Kantian themes in the philosophy of logic*. Clarendon Press, Oxford, 1973. Cited on page **147**.

J. Hintikka. *Inquiry as inquiry: A logic of scientific discovery*. Kluwer Academic Publishers, 1999. Cited on page **169**.

J. Hintikka. *Socratic epistemology: explorations of knowledge-seeking by questioning*. Cambridge University Press, 2007. Cited on page **169**.

J. Hintikka and G. Sandu. What is logic? In D. Jacquette, editor, *Philosophy Of Logic*, Handbook of the Philosophy of Science, pages 13–39. North-Holland, Amsterdam, 2007. ISBN 978-0-444-51541-4. Cited on pages **8 and 148**.

J. Hintikka, I. Halonen, and A. Mutanen. Interrogative logic as a general theory of reasoning. In D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, and J. Woods, editors, *Handbook of the Logic of Argument and Inference: the turn towards the practical*, volume 1 of *Studies in Logic and Practical reasoning*, pages 295–337. Elsevier, Amsterdam, 2002. ISBN 9780444506504. DOI: 10.1016/S1570-2464(02)80009-X. Cited on page **169**.

W. H. Holliday, T. Hoshi, and T. F. Icard. Decidability of the *PAL* substitution core. Manuscript, 2010. Cited on page **77**.

T. de Jager. *"Now that you mention it, I wonder . . . ": Awareness, Attention, Assumption*. PhD thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands, Dec. 2009. ILLC Dissertation series DS-2009-10. Cited on page **166**.

M. Jago. Rule-based and resource-bounded: A new look at epistemic logic. In T. Ågotnes and N. Alechina, editors, *Proceedings of the Workshop on Logics for Resource-Bounded Agents, organized as part of the 18th European Summer School on Logic, Language and Information (ESSLLI)*, pages 63–77, Malaga, Spain, Aug. 2006a. Cited on pages **17, 27, and 119**.

M. Jago. *Logics for Resource-Bounded Agents*. PhD thesis, Department of Philosophy, University of Nottingham, Nottingham, UK, July 2006b. Cited on page **27**.

M. Jago. Epistemic logic for rule-based agents. *Journal of Logic, Language and Information*, 18(1):131–158, 2009. ISSN 0925-8531. DOI: 10.1007/s10849-008-9071-8. Cited on pages **27, 101, and 179**.

S. Jain, D. Osherson, J. S. Royer, and A. Sharma. *Systems that Learn*. The MIT Press, Cambridge, Mass., 1999. ISBN 0-262-10077-0. Cited on page **172**.

R. Kahle. Structured belief bases. *Logic and Logical Philosophy*, 10:45–58, 2002. URL: `http://www.logika.umk.pl/llp/10/kahle.pdf`. Cited on page **164**.

K. T. Kelly. *The logic of reliable inquiry*. Oxford University Press, 1996. ISBN 0195091957. Cited on page **172**.

K. Konolige. Belief and incompleteness. Technical Report 319, SRI International, 1984. Cited on page **2**.

L. A. Kornhauser and L. G. Sager. Unpacking the court. *Yale Law Journal*, 96(1): 82–117, 1986. Cited on page **80**.

R. A. Kowalski. *Logic for Problem Solving*. North Holland, New York, USA, 1979. Cited on page **16**.

G. Lakemeyer. Steps towards a first-order logic of explicit and implicit belief. In Halpern (1986), pages 325–340. ISBN 0-934613-04-4. Cited on page **2**.

P. Lamarre. S4 as the conditional logic of nonmonotonicity. In J. F. Allen, R. Fikes, and E. Sandewall, editors, *KR 91*, pages 357–367, Cambridge, MA, USA, Apr. 1991. Morgan Kaufmann. ISBN 1-55860-165-1. Cited on page **114**.

H. J. Levesque. A logic of implicit and explicit belief. In *Proc. of AAAI-84*, pages 198–202, Austin, TX, 1984. Cited on page **2**.

D. Lewis. General semantics. *Synthese*, 22:18–67, 1970. Cited on pages **10, 11, and 12**.

D. K. Lewis. *Counterfactuals*. Blackwell, Cambridge, Massachusetts, 1973. Cited on pages **114, 115, and 125**.

F. Liu. A two-level perspective on preference. Department of Philosophy, Tsinghua University, 2010. Available at `http://fenrong.net/archives/twolevel.pdf`, 2010. Cited on page **164**.

E. Lorini and C. Castelfranchi. The cognitive structure of surprise: looking for basic principles. *Topoi*, 26(1):133–149, Mar. 2007. ISSN 0167-7411. DOI: 10.1007/s11245-006-9000-x. URL: `http://www.istc.cnr.it/doc/83a_2007021612537t_extended_version.pdf`. Cited on page **174**.

M. Ma. Dynamic epistemic logic of finite identification. In He et al. (2009), pages 227–237. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7_18. Cited on page **173**.

D. Makinson. How to give up: A survey of some formal aspects of the logic of theory change. *Synthese*, 62:347–363, 1985. Cited on page **9**.

J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1-2):27–39, 1980. Cited on page **16**.

J.-J. C. Meyer and W. van der Hoek. A default logic based on epistemic states. *Fundamenta Informaticae*, 23(1):33–65, 1995. Cited on page **154**.

S. Modica and A. Rustichini. Awareness and partitional information structures. *Theory and Decision*, 37:107–124, May 1994. Cited on page **77**.

S. Modica and A. Rustichini. Unawareness and partitional information structures. *Games and Economic Behavior*, 27(2):265–298, May 1999. DOI: 10.1006/game.1998.0666. Cited on page **77**.

R. Montague. Universal grammar. *Theoria*, 36:373–398, 1970. Cited on page **11**.

R. C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94, 1985. ISSN 0004-3702. DOI: 10.1016/0004-3702(85)90042-6. Cited on page **16**.

R. C. Moore. Propositional attitudes and russellian propositions. In R. Bartsch, J. van Benthem, and P. van Embde Boas, editors, *Semantics and Contextual Expressions*, pages 147–174. Foris, Dordrecht, The Netherlands, 1989. Cited on page **10**.

R. C. Moore and G. G. Hendrix. Computational models of beliefs and the semantics of belief sentences. Technical Report 187, SRI International, Menlo Park, California, 1979. URL: `http://www.ai.sri.com/pubs/files/722.pdf`. Cited on page **8**.

M. Moortgat. Categorical type logics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 93–177. Elsevier Science Publishers, Amsterdam, 1997. Cited on page **32**.

A. Moreno. Avoiding logical omniscience and perfect reasoning: a survey. *AI Communications*, 11(2):101–122, 1998. ISSN 0921-7126. Cited on page **2**.

Y. Mukouchi. Characterization of finite identification. In K. P. Jantke, editor, *AII*, volume 642 of *Lecture Notes in Computer Science*, pages 260–267. Springer, 1992. ISBN 3-540-56004-1. DOI: 10.1007/3-540-56004-1_18. Cited on page **173**.

E. Pacuit. Neighborhood semantics for modal logic. an introduction, 2007. URL: `http://ai.stanford.edu/~epacuit/classes/esslli/nbhd_esslli.html`. Lecture notes for the ESSLLI's course *A Course on Neighborhood Structures for Modal Logic*. Cited on page **11**.

C. S. Peirce. A letter to J. H. Kehler, nem 3:203-204, 1911. Available at `http://www.helsinki.fi/science/commens/terms/retroduction.html`, 1911. Cited on page **156**.

J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, Tennessee, USA, 1989. Oak Ridge National Laboratory, ORNL/DSRD-24. Cited on pages **13, 61, 126, and 181**.

B. Polak. Epistemic conditions for nash equilibrium, and common knowledge of rationality. *Econometrica*, 67(3):673–676, May 1999. Cited on page **177**.

A. N. Prior. *Time and Modality*. Clarendon Press, Oxford, 1957. Cited on page **26**.

R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132, 1980. Cited on pages **16 and 153**.

H. Rott. *Change, Choice and Inference: a Study of Belief Revision and Nonmonotonic Reasoning*. Number 42 in Oxford Logic Guides. Oxford Science Publications, 2001. Cited on pages **16 and 124**.

B. Russell. *The Principles of Mathematics*. W. W. Norton and Company, Inc., New York, 1903. Cited on page **10**.

M. Ryan. *Ordered Presentations of Theories. A Hierarchical Approach to Default Reasoning*. PhD thesis, Department of Computing, Imperial Collegue, London, UK, 1992. Cited on pages **9 and 164**.

D. Samet, editor. *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2007), Brussels, Belgium, June 25-27, 2007*, 2007. Cited on pages **77 and 201**.

D. Scott. Advice in modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, Dordrecht, The Netherlands, 1970. Cited on page **11**.

K. Segerberg. The basic dynamic doxastic logic of AGM. In Williams and Rott (2001), pages 57–84. ISBN 978-0-7923-7021-5. Cited on page **114**.

S. Sequoiah-Grayson. The scandal of deduction. *Journal of Philosophical Logic*, 37(1):67–94, 2008. ISSN 0022-3611. DOI: 10.1007/s10992-007-9060-4. Cited on page **147**.

G. Sillari. Models of awareness. In Bonanno et al. (2006), pages 209–218. Cited on page **77**.

K. M. Sim. Epistemic logic and logical omniscience. A survey. *International Journal of Intelligent Systems*, 12(1):57–81, Jan. 1997. DOI: 10.1002/(SICI)1098-111X(199701)12:1⟨57::AID-INT3⟩3.0.CO;2-X. Cited on page **2**.

F. Soler-Toscano and F. R. Velázquez-Quesada. Abduction for (non-omniscient) agents. In Boissier et al. (2010). URL: `http://ceur-ws.org/Vol-627`. Cited on pages **21 and 156**.

R. Stalnaker. *Inquiry*. The MIT Press, Cambridge, MA, 1984. Cited on page **166**.

R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12(2):133–163, Oct. 1996. DOI: 10.1017/S0266267100004132. Cited on page **177**.

R. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128(1): 169–199, Mar. 2006. ISSN 0031-8116. DOI: 10.1007/s11098-005-4062-y. Cited on page **117**.

A. S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Number 43 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, U. K., second edition, July 2000. ISBN 0521779111. Cited on page **16**.

M. Y. Vardi. On epistemic logic and logical omniscience. In Halpern (1986), pages 293–305. ISBN 0-934613-04-9. Cited on page **2**.

M. Y. Vardi, editor. *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, CA, March 1988*, 1988. Morgan Kaufmann. ISBN 0-934613-66-4. Cited on pages **202 and 207**.

F. R. Velázquez-Quesada. Inference and update. In J. van Benthem and E. Pacuit, editors, *Proceedings of the Workshop on Logic and Intelligent Interaction, organized as part of the European Summer School on Logic, Language and Information (ESSLLI) 2008*, pages 12–20, Hamburg, Germany, Aug. 2008a. Cited on pages **21 and 42**.

F. R. Velázquez-Quesada. Inference and update. Technical Report PP-2008-48, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), 2008b. URL: `http://www.illc.uva.nl/Publications/ResearchReports/PP-2008-48.text.pdf`. Cited on page **21**.

F. R. Velázquez-Quesada. Inference and update. *Synthese (Knowledge, Rationality and Action)*, 169(2):283–300, July 2009a. ISSN 0039-7857. DOI: 10.1007/s11229-009-9556-2. Cited on page **21**.

F. R. Velázquez-Quesada. Dynamic logics for implicit and explicit information. Available at `http://staff.science.uva.nl/~fvelazqu/docs/InfBeliefs-06-25.pdf`, 2009b. Cited on pages **113 and 119**.

F. R. Velázquez-Quesada. Dynamic logics for explicit and implicit information. In He et al. (2009), pages 325–326. ISBN 978-3-642-04892-0. DOI: 10.1007/978-3-642-04893-7_31. Cited on page **21**.

F. R. Velázquez-Quesada. Dynamic epistemic logic for implicit and explicit beliefs. In Boissier et al. (2010). URL: `http://ceur-ws.org/Vol-627`. Cited on page **21**.

F. R. Velázquez-Quesada. Dynamics of implicit and explicit beliefs. In E. Lorini and L. Vieu, editors, *TIDIAD '10*, Copenhagen, Denmark, Aug. 2010b. Cited on page **21**.

F. Veltman. *Logics for Conditionals*. PhD thesis, Universiteit van Amsterdam, 1985. Cited on pages **114 and 115**.

F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25: 221–261, 1996. Cited on page **154**.

M.-A. Williams and H. Rott, editors. *Frontiers in Belief Revision*, number 22 in Applied Logic Series, 2001. Kluwer Academic Publishers. ISBN 978-0-7923-7021-5. Cited on pages **9, 16, 124, and 213**.

R. Z. Yang, Apr. 2009. Private communication. Cited on page **53**.

A. Yap. Product update and temporal modalities. Forthcoming in the proceedings for Dynamic Logic Montréal Workshop, 2007. Cited on page **175**.

Index

Symbols

A

B

C

D

E

I

K

L

M

N

O

P

R

S

T

U

V

W

Samenvatting

Dit proefschrift presenteert een logisch kader voor de weergave van kleine stappen in de dynamica van informatie.

De klassieke kennislogica, met de mogelijke-werelden semantiek, is een van de meest verbreide systemen voor de weergave van en het redeneren over informatie van actoren. Haar dynamisch tegenwicht, 'dynamic epistemic logic' (dynamische kennislogica), maakt het ons mogelijk om acties die informatie veranderen weer te geven en om daarover te redeneren. Een voorbeeld zijn de zogenaamde 'harde' aankondigingen die als gevolg hebben dat we alle mogelijkheden waarin de aangekondigde informatie niet waar is, compleet weggooien. Een ander voorbeeld zijn de 'zachte' aankondigingen waarin we eenvoudigweg het verkondigde aannemelijker vinden dan het tegendeel, maar waarin we niettemin de situaties waarin de aankondiging onwaar is niet elimineren.

In dit kennislogisch kader zijn de actoren echter alwetend: hun informatie is afgesloten onder de logische gevolgtrekkingsrelatie. Deze eigenschap, ook al is ze bruikbaar in sommige toepassingen, is een zeer sterke idealisering in andere toepassingen. Het wordt vaak beweerd dat kennislogica om die reden geen geschikt gereedschap is voor het redeneren over de informatie van 'echte' actoren met begrensde vermogens. En, wat nog belangrijker is, alwetendheid maakt de kleine stappen volstrekt irrelevant die wij in het dagelijks leven als niet-ideale actoren nemen, zoals bewustzijnsverandering, introspectie, en vooral: afleiding.

In dit proefschrift breiden we de klassieke kennislogica uit, opdat wij behalve de kennislogische notie van informatie, met dit aspect van alwetendheid, ook meer verfijnde noties van informatie kunnen weergeven. Deze verfijnde noties kunnen een aantal afsluitingseigenschappen missen en hoeven in het bijzonder niet afgesloten te zijn onder logisch gevolg. We onderzoeken verschillende zulke noties zoals bewustzijn ('awareness')(hoofdstukken 3 en 4),

expliciete kennis (hoofdstukken 2, 4, en 5), en expliciet geloof (hoofdstuk 5), en we bespreken de eigenschappen van deze noties.

Wat nog belangrijker is, we geven eveneens definities voor verfijnde vormen van acties die opereren op deze verfijnde noties van informatie. We introduceren acties voor de weergave van bewustzijnsverandering (hoofdstukken 3 en 4), voor de weergave van op kennis gebaseerde (waarheidsbehoudende) afleiding (hoofdstukken 2, 4, en 5) en voor de weergave van op geloof gebaseerde (niet waarheidsbehoudende) afleiding (hoofdstuk 5). We presenteren tevens niet-alwetende versies van de eerder genoemde acties van 'harde' en 'zachte' aankondiging (hoofdstukken 2 en 4 voor de eerste, en hoofdstuk 5 voor de tweede). In alle gevallen definiëren we de actie, presenteren haar basiseigenschappen, en geven een correct en volledig axiomatisch bewijssysteem.

Het door ons ontwikkeld systeem heeft verscheidene verbanden en toepassingen. In het bijzonder bespreken we de relatie van de diverse vormen van afleiding die we gedefiniëerd hebben tot standaard redeneervormen als deductie, verstekredeneren ('default reasoning'), en abductie (hoofdstuk 6). Wat betreft toepassingen maken we enige suggesties over de bruikbaarheid van ons systeem voor nieuwe perspectieven in taalkunde, cognitiewetenschap, en speltheorie (hoofdstuk 7).

We besluiten met het vermelden van nog verschillende andere vragen en uibreidingen die aanvullend onderzoek behoeven (hoofdstuk 8).

Abstract

This dissertation presents a logical framework for representing small steps in dynamics of information.

Classical *Epistemic Logic* with possible worlds models is one of the most widely used frameworks for representing and reasoning about agents' information. Its dynamic counterpart, *Dynamic Epistemic Logic*, allows us to represent and reason about actions that change this information, like 'hard' announcements that make us discard completely the possibilities where the announced proposition is not true, of 'soft' announcements where we simply consider the announced proposition very likely to be the case, but nevertheless we do not eliminate the situations in which it does not hold.

However, agents represented in the epistemic logic framework are *omniscient*: their information is closed under logical consequence. This property, useful in some applications, is a very strong idealization in some others: it is often argued that, because of it, epistemic logic is not an adequate tool for reasoning about the information of 'real' agents with bounded abilities. More importantly, omniscience makes irrelevant the small steps that we non-ideal agents perform every day in our life, like change in awareness, introspection and, especially, *inference*.

In this dissertation, we extend the classical epistemic logic framework in order to represent, besides the omniscient epistemic logic notion of information, other finer notions that do not *need* to have strong closure properties and, in particular, do not need to be closed under logical consequence. We explore different definitions for notions like *awareness* (Chapters 3 and 4), *explicit knowledge* (Chapters 2, 4 and 5) and *explicit beliefs* (Chapter 5), discussing some of their properties.

More importantly, we provide definitions for finer actions that affect these finer notions of information. We introduce actions representing changes in awareness (Chapters 3 and 4), *knowledge-based* (i.e., *truth-preserving*) *inference* (Chapters 2, 4 and 5) and *belief-based* (*non-truth-preserving*) *inference* (Chapter

5). We also present non-omniscient versions of the already studied acts of 'hard' and 'soft' announcement (Chapters 2 and 4 for the first, Chapter 5 for the second). In all cases we define the action, present its basic properties, and provide a sound and complete axiom system.

The developed framework has a wide range of connections and applications. In particular, we discuss the relation of the several acts of inference we define with known forms of reasoning, like deduction, default and abductive reasoning (Chapter 6). For applications, we make a few suggestions of how our framework might provide a useful tool that gives new perspective in fields like Linguistics, Cognitive Science and Game Theory (Chapter 7).

We conclude by mentioning further interesting questions and extensions that deserve additional investigation (Chapter 8).

Resumen

Este trabajo presenta un sistema lógico en el cual es posible representar pequeñas acciones que modifican la información de un agente.

El sistema de *Lógica Epistémica* con mundos posibles como modelo semántico es uno de los sistemas más usados para representar y razonar acerca de la información de un grupo de agentes. Su versión dinámica, *Lógica Dinámica Epistémica*, nos permite representar y razonar acerca de las acciones que modifican dicha información, tales como anuncios 'drásticos' que eliminan completamente las situaciones en las cuales la información anunciada es falsa, o anuncios 'débiles' que hacen que el agente considere como más probables las situaciones en las cuales la información anunciada es verdadera, sin descartar aquellas situaciones en las cuales el anuncio es falso.

Sin embargo, los agentes cuya información es representada en este sistema son *omniscientes*: la información que tienen es cerrada bajo consecuencia lógica. Esta propiedad, útil en algunas aplicaciones, es una idealización muy grande en otras, y de hecho una de las críticas más comunes a la lógica epistémica es que, debido a esta característica, no es una herramienta adecuada para representar y razonar acerca de la información de agentes 'reales' con capacidades limitadas. Aún más importante es el hecho de que la omnisciencia lógica vuelve irrelevantes las pequeñas acciones que modifican la información de agentes 'reales' en actividades comunes, tales como cambios en la información de la cual el agente es consciente, actos de introspección y, en particular, actos de inferencia.

El presente trabajo extiende el sistema de Lógica Epistémica con el fin de representar no solo la mencionada noción omnisciente de información, sino también nociones más refinadas que no tienen idealizaciones tan grandes y, en particular, pueden no ser cerradas bajo consecuencia lógica. En los capítulos 3 y 4 presentamos diferentes definiciones de *la información de la cual el agente es consciente*; en los capítulos 2, 4 y 5 presentamos diferentes definiciones de *conocimiento explícito*; finalmente, en el capítulo 5 presentamos una definición de *creencia explícita*. Además de presentar las definiciones de dichas nociones

de manera semántica y extender el lenguaje formal con modalidades que las representan, también discutimos varias de sus propiedades.

Una vez que tenemos nociones de información más finas, nuestro siguente paso es definir acciones que las modifican. En los capítulos 3 y 4 presentamos acciones que modifican la información de la cual el agente es consciente; en los capítulos 2, 4 y 5 definimos inferencia basada en conocimiento (es decir, deducción); finalmente, en el capítulo 5 presentamos la acción de inferencia basada en creencias. También presentamos versiones no-omniscientes de acciones estudiadas en Lógica Dinámica Epistémica: los ya mencionados anuncios 'drásticos' y 'débiles' (capítulos 2 y 4 en el primer caso, y capítulo 5 en el segundo). Para cada una de las acciones mencionadas, nuestro trabajo presenta su definición, analiza sus propiedades básicas y presenta un sistema de derivación correcto y completo.

El sistema que aquí se desarrolla tiene conexiones con diversos campos. En particular, la relación de las acciones de inferencia definidas con formas de razonamiento conocidas, tales como dedución y razonamientos por defecto y abductivo, es analizada en el capítulo 6. En cuanto a aplicaciones, el capítulo 7 describe como nuestro sistema proporciona una nueva perspectiva en áreas tales como Lingüística, Ciencias Cognitivas y Teoría de Juegos.

El trabajo concluye en el capítulo 8, donde presentamos nuestras conclusiones y discutimos preguntas y extensiones que merecen una investigación más detallada.

Titles in the ILLC Dissertation Series:

ILLC DS-2006-01: **Troy Lee**
Kolmogorov complexity and formula size lower bounds

ILLC DS-2006-02: **Nick Bezhanishvili**
Lattices of intermediate and cylindric modal logics

ILLC DS-2006-03: **Clemens Kupke**
Finitary coalgebraic logics

ILLC DS-2006-04: **Robert Špalek**
Quantum Algorithms, Lower Bounds, and Time-Space Tradeoffs

ILLC DS-2006-05: **Aline Honingh**
The Origin and Well-Formedness of Tonal Pitch Structures

ILLC DS-2006-06: **Merlijn Sevenster**
Branches of imperfect information: logic, games, and computation

ILLC DS-2006-07: **Marie Nilsenova**
Rises and Falls. Studies in the Semantics and Pragmatics of Intonation

ILLC DS-2006-08: **Darko Sarenac**
Products of Topological Modal Logics

ILLC DS-2007-01: **Rudi Cilibrasi**
Statistical Inference Through Data Compression

ILLC DS-2007-02: **Neta Spiro**
What contributes to the perception of musical phrases in western classical music?

ILLC DS-2007-03: **Darrin Hindsill**
It's a Process and an Event: Perspectives in Event Semantics

ILLC DS-2007-04: **Katrin Schulz**
Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals

ILLC DS-2007-05: **Yoav Seginer**
Learning Syntactic Structure

ILLC DS-2008-01: **Stephanie Wehner**
Cryptography in a Quantum World

ILLC DS-2008-02: **Fenrong Liu**
Changing for the Better: Preference Dynamics and Agent Diversity

ILLC DS-2008-03: **Olivier Roy**
Thinking before Acting: Intentions, Logic, Rational Choice

ILLC DS-2008-04: **Patrick Girard**
Modal Logic for Belief and Preference Change

ILLC DS-2008-05: **Erik Rietveld**
Unreflective Action: A Philosophical Contribution to Integrative Neuroscience

ILLC DS-2008-06: **Falk Unger**
Noise in Quantum and Classical Computation and Non-locality

ILLC DS-2008-07: **Steven de Rooij**
Minimum Description Length Model Selection: Problems and Extensions

ILLC DS-2008-08: **Fabrice Nauze**
Modality in Typological Perspective

ILLC DS-2008-09: **Floris Roelofsen**
Anaphora Resolved

ILLC DS-2008-10: **Marian Counihan**
Looking for logic in all the wrong places: an investigation of language, literacy and logic in reasoning

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder. . . ": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velázquez Quesada**
Small steps in dynamics of information