Small steps in dynamics of information
Velazquez Quesada, F.R.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
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);
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 com-
mon semantic model, the *possible worlds model*, is formally defined as follows.

**Definition 1.1 (Possible worlds model)** Let $P$ 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);
- $V : W \rightarrow \wp(P)$ is an *atomic valuation function*, indicating the atomic pro-
  positions in $P$ 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 distin-
guished 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 inform-
ation. 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* $\varphi$”. The formal definition is the following.
**Definition 1.2 (Epistemic Logic language)** Let \( P \) 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 \( p \in P \) is a formula in the language.
2. If \( \varphi \) and \( \psi \) are formulas in the language, so are \( \neg \varphi \), \( \varphi \lor \psi \) and \( \Box \varphi \).
3. Nothing else is a formula in the language.

The definition can be abbreviated with the following statement. Formulas \( \varphi, \psi \) of the EL language are built according to the following rule, where \( p \) is in \( P \):

\[
\varphi ::= p \mid \neg \varphi \mid \varphi \lor \psi \mid \Box \varphi
\]

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

\[
\varphi \land \psi := \neg (\neg \varphi \lor \neg \psi), \quad \varphi \rightarrow \psi := \neg \varphi \lor \psi, \quad \varphi \leftrightarrow \psi := (\varphi \rightarrow \psi) \land (\psi \rightarrow \varphi).
\]

Similarly, the constants \( \top \) and \( \bot \) can be defined as \( p \lor \neg p \) and \( p \land \neg p \), 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 \( \varphi \) 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,

\[
\begin{align*}
(M, w) \Vdash p & \iff p \in V(w) \\
(M, w) \Vdash \neg \varphi & \iff \text{it is not the case that } (M, w) \Vdash \varphi \\
(M, w) \Vdash \varphi \lor \psi & \iff (M, w) \Vdash \varphi \text{ or } (M, w) \Vdash \psi \\
(M, w) \Vdash \Box \varphi & \iff \text{for all } u \in W, Rwu \text{ implies } (M, u) \Vdash \varphi
\end{align*}
\]

When \( (M, w) \Vdash \varphi \), we say that \( \varphi \) 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 \( \varphi \) at \( w \) in \( M \), \( (M, w) \Vdash \Box \varphi \), if and only if \( \varphi \) is true in all the worlds that are
Chapter 1. Introduction

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

We can use the semantic interpretation of our primitive connectives and modalities \( \neg, \lor \) and \( \Box \) in order to get that of the defined ones \( \land, \rightarrow, \leftrightarrow \) 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:

\[
(M, w) \models \Diamond \varphi \iff \text{there is a } u \in W \text{ such that } Rwu \text{ and } (M, u) \models \varphi
\]

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 \( \neg q \).

\[\begin{array}{c}
(M, w_1) \models \Box p \\
(M, w_1) \models \neg \Box q \land \neg \Box \neg q
\end{array}\]

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 \( p \lor q \) (\( \Box (p \lor q) \) is true at \( w_1 \)), \( q \rightarrow p \) (\( \Box (q \rightarrow p) \) is true at \( w_1 \)) 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 \( p \lor q \), but also about the logically equivalent \( \neg (\neg p \land \neg q) \), \( \neg p \rightarrow q \), \( \neg q \rightarrow p \) and so on.
1.1. Structure of information

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 \( p \rightarrow p \) are, from the agent’s perspective, identical to more ‘illuminating’ ones like \( (p \land (p \rightarrow q)) \rightarrow q \).

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 \( \varphi \) at \( w \) if and only if \( \varphi \) 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 \( \varphi \rightarrow \psi \) and \( \varphi \) are in the set, then consistency gives us the right to add \( \psi \) 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 \( \varphi \rightarrow \psi \) and \( \varphi \) at \( w \) (if \( \Box (\varphi \rightarrow \psi) \) and \( \Box \varphi \) are true at \( w \)), then both \( \varphi \rightarrow \psi \) and \( \varphi \) are true in all worlds R-reachable from \( w \). But then \( \psi \) also holds in each one of such worlds, and therefore the agent is also informed about \( \psi \) 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
Chapter 1. Introduction

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 \( p \land q \) and \( q \land p \) 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 \( p \land p \) 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
1.1. Structure of information

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 $N : W \rightarrow \varphi(\varphi(W))$ that assigns, to each possible world, a set of *sets of worlds*. Then it is said that the agent is informed about some formula $\varphi$ at a world $w$ if and only if the set of worlds in which $\varphi$ is true in $M$ (denoted by $\llbracket \varphi \rrbracket_M$) 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 $\llbracket \varphi \rrbracket_M$ and $\llbracket \varphi \rightarrow \psi \rrbracket_M$ does not imply to have $\llbracket \psi \rrbracket_M$. 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 $p \rightarrow p$ and $(p \land (p \rightarrow q)) \rightarrow q$ 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 $\varphi$ 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”. Gardenfors (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.
1.2. Dynamics

When looking for an 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 \( \chi \) is a formula of the language. More precisely,
Chapter 1. Introduction

**Definition 1.4 (Observation Logic language)** Let $P$ be a set of atomic propositions. Formulas $\varphi, \psi, \chi$ of the language of Observation Logic are given by:

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

with $p$ an atomic proposition in $P$. Formulas of the form $\langle \chi! \rangle \varphi$ are read as "\( \chi \) can be observed and after that $\varphi$ 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 $\chi$, $\varphi$ 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 $\chi$, the agent realizes that $\chi$ is true, and therefore she can discard those possibilities in which $\chi$ does not hold, hence keeping just those in which $\chi$ 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 $\chi$ be a formula in the Observation Logic language. The possible worlds model $M_\chi = \langle W', R', V' \rangle$ is given by

- $W' := \{ w \in W \mid (M, w) \models \chi \}$;
- $R' := R \cap (W' \times W')$;
- 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 $\chi$ 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 $\chi$ a formula of the observation language.

$$(M, w) \models \langle \chi! \rangle \varphi \iff (M, w) \models \chi \text{ and } (M_\chi, w) \not\models \varphi$$

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

$$(M, w) \models [\chi!] \varphi \iff (M, w) \models \chi \text{ implies } (M_\chi, w) \not\models \varphi$$
Note how the observation modality comes with a precondition. Indeed, \( \chi \) can be observed and after doing it \( \varphi \) will be the case, \((M, w) \models (\chi!) \varphi\), if and only if \( \chi \) can be observed, \((M, w) \models \chi\), and after observing it, \( \varphi \) is the case, \((M_{\chi!}, w) \models \varphi\). This precondition has a technical reason. We need for the evaluation point to satisfy \( \chi \); otherwise, it will not survive the observation operation. But there is also an intuitive and very natural reason: in order for \( \chi \) to be observed, \( \chi \) 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 \((M, w_1)\)).

\[
\begin{align*}
(M, w_1) &\not\models \neg \Box q \land (q!) \Box q \\
(M, w_1) &\not\models \Box p \land (q!) \Box p \\
(M, w_1) &\not\models \Box \neg \Box q \land (q!) (\neg \Box \neg \Box q \land \Box \Box q)
\end{align*}
\]

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 \( p \land q \) 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 everyday 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).
1.3. A summary and a choice

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); Agotnes (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 $r \lor \neg r$ 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 □ 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).