Small steps in dynamics of information

Velazquez Quesada, F.R.

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: https://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.
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 \( \sigma \) is for the agent to know explicitly \( \sigma \) 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 \( \sigma \) 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)

\[
\left( \bigwedge_{\psi \in \text{pm}(\sigma)} K_{Ex} \psi \land K_{Ex} \sigma \right) \rightarrow \langle \leftarrow, \rightarrow \rangle K_{Ex} \text{cn}(\sigma)
\]

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 “\( \varphi \) is true”, we usually have information of the form “\( \varphi \) is plausible”. A classical situation is the one used in the running example of Chapter 5: birds typically fly.
6.2. Default reasoning

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 Willy 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 $\psi$ is the prerequisite of the rule, each $\phi_i$ is a justification, and $\chi$ is the conclusion.

$\psi : \phi_1, \ldots, \phi_n \Rightarrow \chi$

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 $\chi$, so the rule is read as “if $\psi$ is the case and $\chi$ 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 φ₁, ..., φₙ are consistent with her knowledge (i.e., she considers φ₁, ..., φₙ 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 φᵢ 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 $\langle \text{Def}_χ^ψ : φ₁, ..., φₙ \rangle ϕ$, defined as

$$\langle \text{Def}_χ^ψ : φ₁, ..., φₙ \rangle ϕ := K_{E_X}ψ \land (\overline{K_{E_X}φ₁} \land \cdots \land \overline{K_{E_X}φₙ}) \land \langle χ^+ \rangle ϕ$$

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).
6.2. Default reasoning

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 $\psi$ is explicitly known and $\psi \Rightarrow \chi$ is explicitly believed, that is,

$$\langle \text{Def}_{\chi}^{\psi;\phi_1,...,\phi_n} \varphi \rangle := K_{\text{Ex}} \psi \land B_{\text{Ex}}(\psi \Rightarrow \chi) \land \langle C_{\text{KB}}^{\psi \Rightarrow \chi}, e_1 \rangle \varphi$$

where $C_{\text{KB}}^{\psi \Rightarrow \chi}$ 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 $\phi_i$ fails, Chilly Willy does not fly, that is, she should believe, at least implicitly, that for every justification $\phi_i$, formulas of the form $\neg \phi_i \rightarrow \neg \chi$ hold. With our notation, this is

$$B_{\text{im}}(\neg \phi_i \rightarrow \neg \chi)$$

stating that every $\neg \phi_i \rightarrow \neg \chi$ is true the agent’s most plausible worlds.

Suppose that indeed some justification $\phi_k$ fails and the agent knows it explicitly, that is,

$$K_{\text{ex}} \neg \phi_k$$

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,

$$K_{\text{ex}} \psi$$

This makes $\psi$ true in all the worlds she considers possible, and then the most plausible ones satisfy $\psi$, but also $\neg \chi$. But now she cannot believe $\psi \Rightarrow \chi$ neither implicitly nor explicitly, and therefore she cannot apply the inference step.

$$\neg B_{\text{im}}(\psi \Rightarrow \chi)$$

Even if she does not know that some justification fails and simply believes it explicitly, $B_{\text{ex}} \neg \phi_k$, this would still make $\psi$ 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 $\chi$ is observed.
But if $\psi$ were true, $\chi$ would be a matter of course.
Hence, there is reason to suspect that $\psi$ 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
6.3. Abductive reasoning

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 \( \sigma \) be a rule, and recall that for a given formula \( \varphi \), the worlds the agent recognizes as \( \varphi \)-worlds are those satisfying \( G_\varphi := \varphi \land A \varphi \). The action of knowledge-based abduction (that is, abduction with known rule and known conclusion) is given by the PA action model \( C_{Abl(\sigma)}^{KK} \) whose definition is the following.

\[
\begin{align*}
E & := \{ e_1, e_2 \} \\
\leq & := \{(e_1, e_1), (e_1, e_2), (e_2, e_2)\} \\
\text{Pre}(e_1) & := (K_{Ex} \sigma \land K_{Ex}cn(\sigma)) \land \neg\left( \bigwedge_{\psi \in pm(\sigma)} G_\psi \right) \\
\text{Pre}(e_2) & := (K_{Ex} \sigma \land K_{Ex}cn(\sigma)) \land \left( \bigwedge_{\psi \in pm(\sigma)} G_\psi \right)
\end{align*}
\]

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.

![Diagram](image)

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.
Chapter 6. Connections with other forms of reasoning

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 been 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.

The strong abductive precondition, $K_{Ex}(t \Rightarrow c) \wedge K_{Ex} c$, 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.
6.3. Abductive reasoning

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 \( \sigma \) be a rule, and recall that for any formula \( \varphi \), the worlds the agent recognize as \( \varphi \)-worlds are those satisfying \( G_\varphi := \varphi \land A \varphi \). The action of belief-based abduction is given by the PA action model \( C^{\text{Abd}(\sigma)}_{KB} \), differing from its knowledge-based counterpart \( C^{\text{Abd}(\sigma)}_{KK} \) (Definition 6.1) only in the abductive precondition, which now becomes \( K_{Ex}\sigma \land B_{Ex}\text{cn}(\sigma) \). More precisely, the precondition of events \( e_1 \) and \( e_2 \) are now given by

- \( \text{Pre}(e_1) := (K_{Ex}\sigma \land B_{Ex}\text{cn}(\sigma)) \land \neg \left( \bigwedge_{\psi \in \text{pm}(\sigma)} G_\psi \right) \)
- \( \text{Pre}(e_2) := (K_{Ex}\sigma \land B_{Ex}\text{cn}(\sigma)) \land \left( \bigwedge_{\psi \in \text{pm}(\sigma)} G_\psi \right) \)

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
Belief bases in belief revision

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

\[
\text{BelSet}_{(M,w)} := \{ \varphi \in \mathcal{L} | (M, w) \models B_{\text{Im}} \varphi \}
\]

In other words, the formula \( \varphi \) 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:

\[
\text{BelBas}_{(M,w)} := \{ \varphi \in \mathcal{L} | (M, w) \models B_{\text{Ex}} \varphi \}
\]

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:

\[
\begin{array}{c}
\text{p}
\end{array}
\]

\[
\begin{array}{c}
\rightarrow \text{A := \{\}} \\
\end{array}
\]

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 \( p \in \mathcal{P} \) and for all \( w \in \mathcal{W} \),

\[
p \in V(w) \text{ implies } p \in A(w) \quad \text{and} \quad p \notin V(w) \text{ implies } \neg p \in A(w)
\]

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

\[ B_{\text{im}}\varphi \rightarrow \langle \ast \rangle B_{\text{ex}}\varphi \]

where the modality \( \langle \ast \rangle \) 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 non-monotonic 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 \( \chi \) 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 \( \chi \) (known or just believed) and how reliable is the external observation (infallible or just plausible). In our setting, the case in which the agent knows \( \chi \) and gets informed that \( \neg \chi \) certainly holds cannot happen, because \( \chi \) 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 \( \chi \), then being suggested that \( \neg \chi \) is the case will not affect neither her knowledge
Dealing with contradictions

nor her beliefs. On the other hand, if the agent believes $\chi$, 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 $\chi$ 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 $\text{DEL}$ 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 $B_{\text{Im}} \phi \land B_{\text{Im}} \neg \phi$ is not satisfiable,
and therefore neither is $B_{\text{Ex}} \phi \land B_{\text{Ex}} \neg \phi$. For the case of knowledge, the indistin-
guishability relation is reflexive (because the plausibility relation is reflexive);
hence $K_{\text{Im}} \phi \land K_{\text{Im}} \neg \phi$ is not satisfiable and therefore neither is $K_{\text{Ex}} \phi \land K_{\text{Ex}} \neg \phi$.

Note how implicit/explicit knowledge/beliefs cannot face internal contra-
dictions 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 ac-
cess sets. 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 $(\neg) (A \phi \land A \neg \phi)$ 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 $\chi$ 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.

---

2In 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 $\chi$ and then she truthfully and explicitly observes $\neg \chi$, there should be a copy of the world in which she has not acknowledged $\chi$ (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.