J Log Lang Inf (2014) 23:141–168
DOI 10.1007/s10849-014-9192-1

Generation and Selection of Abductive Explanations for Non-Omniscient Agents

Fernando Soler-Toscano · Fernando R. Velázquez-Quesada

Published online: 25 March 2014

Abstract Among the non-monotonic reasoning processes, abduction is one of the most important. Usually described as the process of looking for explanations, it has been recognized as one of the most commonly used in our daily activities. Still, the traditional definitions of an abductive problem and an abductive solution mention only theories and formulas, leaving agency out of the picture. Our work proposes a study of abductive reasoning from an epistemic and dynamic perspective. In the first part we explore syntactic definitions of both an abductive problem in terms of an agent's information and an abductive solution in terms of the actions that modify the agent's information. We look at diverse kinds of agents, including not only omniscient ones but also those whose information is not closed under logical consequence and those whose reasoning abilities are not complete. In the second part, we look at an existing logical framework whose semantic model allows us to interpret the previously stated formulas, and we define two actions that represent forms of abductive reasoning.

Keywords Abductive reasoning · Non-omniscient agents · Dynamic epistemic logic

1 Abductive Reasoning

Among the non-monotonic reasoning processes, abduction (Aliseda 2006) is one of the most important. The concept, introduced into modern logic by Charles S. Peirce, is usually described as the process of *looking for an explanation*, and it has been recognized as one of the most commonly used in our daily activities. Observing that

F. Soler-Toscano (✉) · F. R. Velázquez-Quesada
Grupo de Lógica, Lenguaje e Información, Universidad de Sevilla, Seville, Spain
e-mail: fsoler@us.es

F. R. Velázquez-Quesada
e-mail: FRVelazquezQuesada@us.es

Mr. Wilson's right cuff is very shiny for five inches and that his left one has a smooth patch near the elbow, Holmes assumes that he (Mr. Wilson) has done a considerable amount of writing lately. Karen knows that when it rains, the grass gets wet, and that the grass is wet right now; then, she suspects that it has rained. There are also cases in science: in classical mechanics, rules are introduced in order to explain macroscopical movements. In Peirce's own words (Hartshorne and Weiss 1934), abduction can be described in the following way:

> The surprising fact χ is observed.
> But if ψ were true, χ would be a matter of course.
> Hence, there is reason to suspect that ψ is true.

But though traditional examples of abductive reasoning are given in terms of an agent's information and its changes, classic definitions of an abductive problem and its solutions are given in terms of theories and formulas, without mentioning the agent's information and how it is modified.

The present work proposes a study of abductive reasoning from an epistemic and dynamic perspective, and it is divided into two parts. The first one follows recent developments in *Epistemic Logic* [*EL*; Hintikka (1962)] and *Dynamic Epistemic Logic* [*DEL*; Ditmarsch et al. (2007), Benthem (2011)], and focuses on providing formulas defining what an abductive problem is in terms of an agent's information as well as what an abductive solution is in terms of the actions that modify it. We explore not only the case of omniscient agents, but also the cases of agents whose information is not closed under logical consequence and whose reasoning abilities are not complete. The second part focuses on a framework in which the provided formulas can be interpreted. We extend the ideas presented in Velázquez-Quesada et al. (2013), Nepomuceno-Fernández et al. (2013) in order to deal with non-omniscient agents, and we define two actions that allow us to represent certain forms of abductive reasoning. We finish with a summary and further research points.

2 From a Classical to a Dynamic Epistemic Approach

Traditionally, it is said that an abductive problem arises when there is a formula χ that does not follow from the theory Φ. The intuitive idea is that Φ represents the current information about the world, and observing a situation χ that is not entailed by Φ shows that the theory is incomplete. Then we should look for extra information that 'fills the gap'.

But even if the theory Φ does not entail the observed χ, it may as well be the case that it entails *the negation* of χ. This case, extensively studied in *belief revision* (Gärdenfors 1992; Williams and Rott 2001), has been recently incorporated as an abductive case.

All in all, the key ingredient for an abductive problem is the existence of a χ that is not entailed by the current information Φ. This situation generates two forms of abductive problems, according to what the information predicts about $\neg\chi$ (Aliseda 2006).

Definition 1 (*Abductive problem*) Let Φ and χ be a theory and a formula, respectively, in some language $\mathcal{L}$. Let $\vdash$ be a consequence relation on $\mathcal{L}$.

– The pair (Φ, χ) forms a *novel abductive problem* when neither χ nor $\neg\chi$ are consequences of Φ, i.e., when

$$\Phi \nvdash \chi \quad \text{and} \quad \Phi \nvdash \neg\chi$$

– The pair (Φ, χ) forms an *anomalous abductive problem* when, though χ is not a consequence of Φ, $\neg\chi$ is, i.e., when

$$\Phi \nvdash \chi \quad \text{and} \quad \Phi \vdash \neg\chi$$

It is typically assumed that the theory Φ is a set of formulas closed under logical consequence, and that $\vdash$ is a truth-preserving consequence relation.

Consider a novel abductive problem. The observation of a χ about which the theory does not have any opinion shows that the theory is incomplete, and then we should look for information that 'completes' it, making χ a consequence of it. This solves the problem because now the theory is strong enough to *explain* χ. Consider now an anomalous abductive problem. The observation of a χ whose negation is entailed by the theory shows that the theory contains a mistake, and now we need two steps to solve this. First, perform a *theory revision* that stops $\neg\chi$ from being a consequence of Φ; after this we are now in a novel abductive case, and we can look for information that 'completes' the theory, making χ a consequence of it. Here are the formal definitions.

Definition 2 (*Abductive solution*)

– Given a *novel* abductive problem (Φ, χ), the formula ψ is said to be an *abductive solution* if

$$\Phi \cup \{\psi\} \vdash \chi$$

– Given an *anomalous* abductive problem (Φ, χ), the formula ψ is an *abductive solution* if it is possible to perform a theory revision to get a *novel* problem (Φ', χ) for which ψ is a solution.

This definition of an abductive solution is often considered as too weak, since ψ can take many 'trivial' forms, like anything that contradicts Φ (then everything, including χ, follows from $\Phi \cup \{\psi\}$) and even χ itself (we clearly have $\Phi \cup \{\chi\} \vdash \chi$). Further conditions can be imposed in order to define more satisfactory solutions; here are some of them (Aliseda 2006).

Definition 3 (*Classification of abductive solutions*) Let (Φ, χ) be an abductive problem. An abductive solution ψ is

consistent iff	$\Phi, \psi \nvdash \bot$	
explanatory iff	$\psi \nvdash \chi$	
minimal iff, for every other solution φ,	$\psi \vdash \varphi$	implies $\varphi \vdash \psi$

The *consistency* requirement discards those ψ inconsistent with Φ, since a reasonable explanation should not be in contradiction with the theory. In a similar way, the *explanatory* requirement discards those explanations that would justify χ by themselves, since it is preferred that the explanation only complements the current theory. Finally, the *minimality* requirement works as the Occam's razor, looking for the simplest explanation: a solution ψ is minimal if it is in fact logically equivalent to any other solution it implies.

2.1 From an Agent's Perspective

Abduction, like other forms of non-monotonic inference, is classified as commonsense reasoning rather than mathematical one, and most of its classic examples involve 'real' agents and their information. It is Holmes who observes that Mr. Wilson's right cuff is very shiny; it is Karen who observes that the grass is wet; it is the scientific community who wants to explain the movement of objects. Nevertheless, the classic definitions of an abductive problem and an abductive solution do not mention agents at all. More important, an abductive solution is typically defined as a piece of information that will produce certain results *when incorporated into the current information*. But, again, the definition of abductive solutions focuses on the properties a formula should satisfy in order to be a solution, but not in the different forms it can be incorporated into an agent's information.

Epistemic logic deals with agents, their information, and properties of this information. Its dynamic counterpart, Dynamic Epistemic Logic, allows us to reason about the way the agent's information changes as consequence of different informational actions. The following three ideas are, from our perspective, the most important of this methodology. First, the notion of information should be closely related to the notion of agency. When we talk about information, there is always an abstract agent (a human being, a computer program, etc.) involved: the one that owns and uses the information. Second, though in mathematical reasoning there is only one notion of information, *true* information, in commonsense 'human' reasoning there are several of them. We usually deal not only with information that is true (what we *know*), but also with information that though not certain is very plausible (what we *believe*), information that we entertain (what we are *aware of*) and so on. And even inside each one of these notions we can make further refinements, like distinctions between *implicit* and *explicit* forms. As interesting as it is to study each notion in isolation, it is more interesting to study them together to observe the relations between them and the way they interact with each other. And third, notions of information are not static: we are continuously performing informational actions that modify them. Actions are important, and deserve to appear explicitly in the analysis.

Based on these ideas, our work proposes a dynamic epistemic approach to abductive reasoning, proposing definitions of abductive problem and abductive solution in terms of an agent's information and the way it changes [cf. Velázquez-Quesada et al. (2013)]. The first step is, then, to answer the question: when does an agent have an abductive problem?

When we put an agent in the picture, the theory Φ becomes *the agent's information*. Then the key ingredient for an abductive problem, a formula χ that does not follow from the theory Φ, becomes a formula *that is not part of the agent's information*. The following definitions use formulas in *EL* style, with Inf φ read as "*φ is part of the agent's information*".

Definition 4 (*Subjective abductive problems*) Let χ be a formula.

– An agent has a *novel χ-abductive problem* when neither χ nor $\neg\chi$ are part of her information, i.e., when the following formula holds:

$$\neg \text{Inf}\,\chi \;\wedge\; \neg \text{Inf}\,\neg\chi$$

– An agent has an *anomalous χ-abductive problem* when χ is not part of her information but $\neg\chi$ is, i.e., when the following formula holds:

$$\neg \text{Inf}\,\chi \;\wedge\; \text{Inf}\,\neg\chi$$

We have identified Φ with the agent's information. Then, a solution for the subjective *novel* case is a formula ψ that, when added to the agent's information, makes the agent informed about χ. This highlights the fact that the requisites of a solution involve an *action*; an action that changes the agent's information by adding ψ to it.

We will express changes in the agent's information by using formulas in *DEL* style. In particular, we will use formulas of the form $\langle \text{Add}_\psi \rangle\, \varphi$ ("*ψ can be added to the agent's information and, after that, φ is the case*") and formulas of the form $\langle \text{Rem}_\psi \rangle\, \varphi$ ("*ψ can be removed from the agent's information and, after that, φ is the case*").[1]

Definition 5 (*Subjective abductive solution*) Suppose an agent has a *novel* χ-abductive problem, that is, $\neg \text{Inf}\,\chi \wedge \neg \text{Inf}\,\neg\chi$ holds. A formula ψ is an *abductive solution* to this problem iff, when added to the agent's information, the agent becomes informed about χ. In a formula,

$$\langle \text{Add}_\psi \rangle\, \text{Inf}\,\chi$$

Now suppose the agent has an *anomalous* χ-abductive problem, that is, $\neg \text{Inf}\,\chi \wedge \text{Inf}\,\neg\chi$ holds. A formula ψ is an *abductive solution* to this problem iff the agent can revise her information to remove $\neg\chi$ from it, producing in this way a novel χ-abductive problem for which ψ is a solution. In a formula,

$$\langle \text{Rem}_{\neg\chi} \rangle \left(\neg \text{Inf}\,\neg\chi \;\wedge\; \langle \text{Add}_\psi \rangle\, \text{Inf}\,\chi \right)$$

We can also provide formulas that characterize properties of abductive solutions.

[1] Note how the operations of adding and removing information cannot be fully specified until we fix a specific notion and a semantic model. For example, in the possible worlds semantics in which knowledge is understood as what is true in all epistemically possible situations, change in knowledge amounts to expand or shrink this set of possible worlds. But if in the same setting we work with the notion of beliefs, usually understood as what is true in only the most plausible worlds, then change in beliefs amounts only to change in the plausibility ordering.

Definition 6 (*Classification of subjective abductive solutions*) Suppose an agent has a χ-abductive problem. A solution ψ is

– *consistent* iff it is a solution and can be added to the agent's information without making it inconsistent:

$$\langle \text{Add}_\psi \rangle \left(\text{Inf}\ \chi \wedge \neg \text{Inf}\ \bot \right)$$

– *explanatory* iff it is a solution and it does not entail χ, that is, it only *complements* the agent's information to produce χ:

$$\psi \rightarrow \chi \ \text{ is not valid,} \quad \text{and} \ \langle \text{Add}_\psi \rangle\, \text{Inf}\ \chi$$

– *minimal* iff it is a solution and, for any other solution φ, if φ is incorporated into the agent's information after ψ is added, then ψ is also incorporated into the agent's information after φ is added.

$$\langle \text{Add}_\psi \rangle\, \text{Inf}\ \chi \wedge \Big(\big(\langle \text{Add}_\varphi \rangle\, \text{Inf}\ \chi \wedge \langle \text{Add}_\psi \rangle\, \text{Inf}\ \varphi \big) \rightarrow \langle \text{Add}_\varphi \rangle\, \text{Inf}\ \psi \Big)$$

Our just provided formulas are straightforward translations of the classic definitions of an abductive problem and an abductive solution. But in these classic definitions the set of formulas Φ is understood as closed under logical consequence: we have provided definitions just for *omniscient* agents. Of course, non-omniscient agents can also face abductive problems.

3 Different Problems for Different Kinds of Agents

Though examples of abductive reasoning involve agents and their information, not all abductive problems have the same form. For example, while in the Holmes' example what Sherlock is missing is a premise that would allow him to derive Mr. Wilson cuff's status, in the mechanics' one what is missing is a relation between facts (that is, a *rule*). And we can even think on situations in which what is missing is not some piece of information, but rather a reasoning step (see Subsect. 3.1).

Pierce himself did not remain quite convinced that one logical form covers all cases of abductive reasoning (Peirce 1911). In fact, different forms of abductive problems arise when we consider agents with different abilities, and even more appear when we combine different notions of information. In this section we will discuss some examples, presenting formulas stating that certain χ poses an abductive problem, and that certain ψ is a solution to it.[2]

[2] A systematic revision of the different cases that arise can be found in Soler-Toscano and Velázquez-Quesada (2010).

3.1 Adding Reasoning to the Picture

Suppose that Karl is in his dining room and sees smoke going out of the kitchen. Karl does not understand why there would be smoke, but then he realizes that the chicken he put on the fire has been there for a long time, and it should be burnt by now. Though initially Karl did not have any explanation about the smoke, he did not need any additional information in order to find a reason for the fire: a simple reasoning step was more than enough.

This case does not correspond to any of the abductive problems described before (Definition 4), and the reason is that Karl is not an omniscient agent: he does not have all logical consequences of his information, and therefore he did not realize that the information he had before seeing the smoke was enough to predict it (i.e, to infer that there would be smoke). This shows that non-omniscient agents can face at least a new kind of abductive problem.

In order to provide formal definitions for abductive problems and solutions involving non-omniscient agents, we need to distinguish between the information the agent actually has, her *explicit* information ($\mathrm{Inf}_{\mathrm{Ex}}$), and what follows logically from it, her *implicit information* ($\mathrm{Inf}_{\mathrm{Im}}$) [see, e.g., Levesque (1984), Vardi (1986)]. Based on this distinction we can say that, for a non-omniscient agent to have an χ-abductive problem, χ should not be part of her *explicit* information. Then, just as there are two abductive problems according to whether the agent is informed about $\neg\chi$ or not, a non-omniscient agent can face eight different abductive problems, according to whether the agent has *explicit* information about $\neg\chi$ or not, and whether she has *implicit* information about χ and $\neg\chi$ or not (Table 1). However, not all these cases are possible: according to our intuition, explicit information should be also implicit information, that is, we should have

$$\mathrm{Inf}_{\mathrm{Ex}}\,\varphi \rightarrow \mathrm{Inf}_{\mathrm{Im}}\,\varphi$$

Then we can discard the cases in which this formula does not hold.[3]

Definition 7 (*Non-omniscient abductive problems*) A non-omniscient agent can face six different abductive problems, each one of them characterized by a formula in Table 1.

The smoke example corresponds to the case (1.2). Though the observation is surprising before it takes place (that is, the agent is not explicitly informed about the smoke in the kitchen), she had this information implicitly, and she could have predicted it by the proper reasoning step(s).

Abductive solutions. A solution for a χ-abductive problem has now as a goal to make the agent *explicitly* informed about χ, without having neither implicit nor explicit information about $\neg\chi$. Each one of the different cases admits different kinds of solutions; let us briefly review the possibilities for the consistent cases, leaving (1.4) and (2.4) for future work.

[3] More cases can be eliminated with further assumptions about the agent's information, like truth or consistency; see Subsect. 3.4.

Table 1 Abductive problems for non-omniscient agents

$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \;\wedge\; \neg\mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \;\wedge$	$\neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\neg\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.1)**
	$\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\neg\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.2)**
	$\neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.3)**
	$\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.4)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \;\wedge\; \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \;\wedge$	$\neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(2.3)**
	$\mathrm{Inf}_{\mathrm{Im}}\,\chi \;\wedge$	$\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(2.4)**

In case (1.1) the agent needs extra external information; this is because χ is not even part of her implicit information. If the extra information ψ makes χ already explicit, nothing else is needed; otherwise, if it makes it only implicit, then the agent will need to perform a further reasoning step. In case (1.2), the one of our example, reasoning is enough and extra information is not essential. Cases (1.3) and (2.3) contain an anomaly: in the first, the anomaly is implicit and the agent needs to make it explicit before solving it; in the second, the anomaly is already explicit, and after solving it the agent will be in case (1.1). In each case there is more than one strategy (i.e., more than one sequence of actions) that solves the abductive problem; for simplicity we will focus on the most representative one for each one of them.

What is interesting here is how, though cases (1.1) and (2.3) are essentially the novel and anomalous abductive cases from before (see Definition 4), cases (1.2) and (1.3) are truly new, and their solutions involve an action that typically is not considered for solving abductive problems: a reasoning step that makes explicit information that was only implicit before. For example, though external information can solve case (1.2), the agent does not really need that: a reasoning step is enough. And for case (1.3), though the agent only sees a novel abductive problem, she has in fact an anomalous one that should be made explicit via reasoning before it can be properly solved.

In the following definition, formulas of the form $\langle\alpha\rangle\,\varphi$ indicate that the agent can perform some reasoning step α after which φ is the case.

Definition 8 (*Non-omniscient abductive solutions*) Representative solutions for consistent non-omniscient abductive problems appear in Table 2.

Classification of abductive solutions. The extra requisites of Definition 6 can be adapted to this non-omniscient case. The *consistency* and the *explanatory* requirements do not have important changes (for the first we ask for the agent's *implicit* information to be consistent at the end of the sequence of actions: $\neg\mathrm{Inf}_{\mathrm{Im}}\,\bot$). The minimality requirement now gives us more options. We can define it over the action $\langle \mathrm{Add}_{\psi}\rangle$, looking for the weakest formula ψ, but it can also be defined over the action $\langle \mathrm{Rem}_{\neg\chi}\rangle$, looking for the revision that removes the smallest amount of information. It can even be defined over the action $\langle\alpha\rangle$, looking for the shortest reasoning chain.

Table 2 Solutions for consistent non-omniscient abductive problems

Case	Solution
(1.1)	A formula ψ such that $\langle \mathrm{Add}_\psi \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\, \chi$
(1.2)	A reasoning α such that $\langle \alpha \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\, \chi$
(1.3)	A reasoning α and a formula ψ such that $\langle \alpha \rangle \big(\mathrm{Inf}_{\mathrm{Ex}}\, \neg\chi \wedge \langle \mathrm{Rem}_{\neg\chi} \rangle\, (\neg \mathrm{Inf}_{\mathrm{Im}}\, \neg\chi \wedge \langle \mathrm{Add}_\psi \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\, \chi)\big)$
(2.3)	A formula ψ such that $\langle \mathrm{Rem}_{\neg\chi} \rangle \big(\neg \mathrm{Inf}_{\mathrm{Im}}\, \neg\chi \wedge \langle \mathrm{Add}_\psi \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\, \chi\big)$

3.2 Not Only Formulas But Also Rules

Consider now the mechanics example. At some stage in history, the scientific community, our ‘agent’, observed the trajectory described by cannonballs, and become interested in explaining this and other related phenomena. But rather than a plain piece of information, the found explanation was an equation that relates initial speed, initial angle and gravity with the described trajectory. In other words, the found explanation was a *rule* that, given the initial conditions, returns us the movement of the projectile.[4]

Again, this case does not fit any of the abductive problems described before. The difference is that our ‘agent’ is not only non-omniscient in the sense that not all her *implicit* information is also *explicit*; she also lacks of the necessary reasoning tools that would allow her to make explicit her implicit information. To put it in other words, besides not having all the logical consequences of her explicit information automatically, the agent might not be able to even derive them. This gives us another kind of agent, and therefore a new and finer classification of abductive problems and abductive solutions.

In order to classify the new abductive problems, we need to make a further distinction: we need to distinguish between what follows logically from the agent’s explicit information, the *objectively* implicit information ($\mathrm{Inf}_{\mathrm{Im}}$), and what the agent can actually derive, the *subjectively* implicit information ($\mathrm{Inf}_{\mathrm{Der}}$). With this refinement, each one of the six abductive problems of Table 1 turns into four cases, according to whether the agent can derive or not what follows logically from her explicit information, that is, according to whether $\mathrm{Inf}_{\mathrm{Der}}\, \chi$ and $\mathrm{Inf}_{\mathrm{Der}}\, \neg\chi$ hold or not. However, we can also make further reasonable assumptions: explicit information is derivable (by the *do-nothing* action) and derivable information is also implicit (this assumes that the agent’s inferential tools are sound). Then we have

$$\mathrm{Inf}_{\mathrm{Ex}}\, \varphi \rightarrow \mathrm{Inf}_{\mathrm{Der}}\, \varphi \qquad \text{and} \qquad \mathrm{Inf}_{\mathrm{Der}}\, \varphi \rightarrow \mathrm{Inf}_{\mathrm{Im}}\, \varphi$$

[4] Physical laws can be properly considered logical formulas, but the action that we define below for abductive reasoning (Definition 17) uses a formula and a rule as different kinds of information. In this setting, it is more appropriate to consider that physical laws are rules.

Table 3 Abductive problems with subjectively/objectively implicit information

$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\left\{ \neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \right\}$	$\wedge \neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.1.a)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$	$\wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.2.a)**
	$\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$		**(1.2.b)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$	$\wedge \neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.3.a)**
	$\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$		**(1.3.c)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$	$\wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(1.4.a)**
	$\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg\mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$		**(1.4.b)**
	$\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$		**(1.4.c)**
	$\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$		**(1.4.d)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\left\{ \neg\mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \right\}$	$\wedge \neg\mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(2.3.c)**
$\neg\mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge$	$\left\{ \mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \right\}$	$\wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge\ \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	**(2.4.d)**

Table 4 Solutions for consistent extended abductive problems

Case	(1.1.a)	(1.2.a)	(1.2.b)	(1.3.a)	(1.3.c)	(2.3.c)
$\langle \mathrm{Add}_{\psi} \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\,\chi$	Solved	–	–	–	–	–
$\langle \mathrm{Add}_{\psi/\alpha} \rangle\, \mathrm{Inf}_{\mathrm{Der}}\,\chi$	–	(1.2.b)	–	–	–	–
$\langle \alpha \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\,\chi$	–	–	Solved	–	–	–
$\langle \mathrm{Add}_{\psi/\alpha} \rangle\, \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi$	–	–	–	(1.3.c)	–	–
$\langle \alpha \rangle\, \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi$	–	–	–	–	(2.3.c)	–
$\langle \mathrm{Rem}_{\neg\chi} \rangle\, \neg\mathrm{Inf}_{\mathrm{Im}}\,\neg\chi$	–	–	–	–	–	(1.1.a)

Definition 9 (*Extended abductive problems*) A non-omniscient agent without complete reasoning abilities can face eleven different abductive problems, each one of them characterized by a formula in Table 3.

The mechanics example corresponds to the case (1.2.a). The trajectory of a projectile is fixed (that is, it is implicit information) once the initial conditions are given; nevertheless, the scientific community could not predict (i.e., derive) the trajectory without knowing the relevant equations.

Abductive solutions. As in Subsect. 3.1, introducing agents with new limitations allows us to explore other actions for solving abductive problems. For example, though case (1.2.b) can be solved by reasoning steps (χ is subjective implicit information so the agent can derive it), this is not possible in case (1.2.a): χ is objectively but not subjectively implicit information, so the agent cannot derive it. This case can be solved with a formula that, when added to the agent's explicit information, makes χ explicit. More interestingly, it can be also solved by *extending the agent's reasoning abilities* with a formula/rule that allows her to derive χ. The same happens with other cases.

As before, different epistemic actions allow us to move between the abductive problems of Table 3. Again, we will focus on the consistent cases, discarding (1.4.$*$) and (2.4.d); again, though some abductive problems accept more than one sequence of steps as a solution, in each case we will focus on the most representative one. In Table 4, the action $\langle \text{Add}_{\psi/\alpha} \rangle$ extends the agent's explicit information by adding a formula ψ or some inference resource α (e.g., a rule) that increases the agent's subjectively implicit information.

Definition 10 (*Extended abductive solutions*) Solutions for consistent extended abductive problems are provided in Table 4, which should be read as a transition table that provides actions and conditions that should hold in order to move from one abductive problem to another.

Table 4 establishes a natural path to solve the new abductive problems. The longest path corresponds to case (1.3.a) in which the agent does not have explicit information about neither χ nor $\neg\chi$ and, though $\neg\chi$ follows logically from her explicit information, she cannot derive it. In this case, the agent should first get enough information to derive $\neg\chi$, thus going into case (1.3.c). Then, after reasoning to derive $\neg\chi$, she will have an explicit anomaly, case (2.3.c). Once there she needs to revise her information to remove $\neg\chi$ from it and, when done (case (1.1.a)), she needs to extend her information with some ψ that will make her be explicitly informed about χ.

Collapsing the cases. From a subjective point of view, the agent does not need to solve an anomaly that she cannot detect. What guides the process of solving an abductive problem is the explicit information and what she can derive from it. In other words, inaccessible anomalies should not matter.

Following this observation, we can notice that some problems in Table 3 are in fact indistinguishable for the agent. Without further external interaction, she can only access her explicit information and eventually what she can derive from it; the rest, the implicit information that is not derivable, is also not relevant. For example, abductive problems (1.{1,2,3,4}.a) are in fact the same from the agent's perspective. Then, by grouping indistinguishable problems, we get the following.

$\neg\text{Inf}_{\text{Ex}}\,\chi \ \wedge\ \neg\text{Inf}_{\text{Ex}}\,\neg\chi \ \wedge$	$\neg\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\neg\text{Inf}_{\text{Der}}\,\neg\chi$	**(1.{1,2,3,4}.a)**
	$\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\neg\text{Inf}_{\text{Der}}\,\neg\chi$	**(1.{2, 4}.b)**
	$\neg\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\text{Inf}_{\text{Der}}\,\neg\chi$	**(1.{3, 4}.c)**
	$\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\text{Inf}_{\text{Der}}\,\neg\chi$	**(1.4.d)**
$\neg\text{Inf}_{\text{Ex}}\,\chi \ \wedge\ \text{Inf}_{\text{Ex}}\,\neg\chi \ \wedge$	$\neg\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\text{Inf}_{\text{Der}}\,\neg\chi$	**(2.3.c)**
	$\text{Inf}_{\text{Der}}\,\chi$	$\wedge$	$\text{Inf}_{\text{Der}}\,\neg\chi$	**(2.4.d)**

Note how these classes correspond to abductive problems in Table 1 in which Inf_{Der} appears in place of Inf_{Im} .

3.3 Explaining Explicit Information

Abduction is usually defined as the problem of explaining a surprising observation. Novelty is an important characteristic of the fact to explain. So, what is the purpose of explaining the information the agent already has explicitly? This information (knowledge or belief) is not supposed to be surprising at all.

But let us look at this from another perspective. When we make a (truthful) observation, as surprising as it can be, it automatically becomes part of our explicit information. After observing Mr. Wilson's cuffs, Holmes knows that while one is very shiny on some area, the other one has a smooth patch; after observing the grass, Karen knows it is wet. What makes this explicit information special (that is, what makes the agent look for an explanation) is that it cannot be justified with the rest of the explicit information the agent possesses. For example, the *fifth postulate* is an obvious piece of explicit information in Euclidean geometry. But, can it be proved from the first four postulates?[5] This generates a variant of an abductive problem that does not depend of recent observations but rather from unjustified information. The agent identifies a piece of explicit information she has and, when she realizes that it cannot be supported by the rest of her information, she tries to find an explanation for it.

One way of stating abductive problems of this form is the following. We introduce the modality Dis_{φ}, representing the action through which the formula φ is discarded from the agent's *explicit* information. Note how this action differs from Rem_{φ}: while the latter removes φ from the agent's *implicit* information, making φ inaccessible without further external interaction, the former removes φ only from the agent's *explicit* information, and therefore φ will be derivable whenever the agent has the proper tools to do it. This 'discarding' action intends to satisfy

$$\mathrm{Inf}_{\mathrm{Ex}}\,\varphi \;\rightarrow\; [\mathrm{Dis}_{\varphi}]\,(\mathrm{Inf}_{\mathrm{Im}}\,\varphi \wedge \neg\mathrm{Inf}_{\mathrm{Ex}}\,\varphi)$$

With this modality we can now make a further distinction in the agent's explicit information, splitting it in what she knows and can actually justify (for example, if no one provides us the quadratic formula, we can still derive it by using the method of completing squares), and what she knows but cannot derive if she would have not observed it (if Holmes had not seen Mr. Wilson, he had not known the status of his cuffs). More precisely, we say that φ is *observed explicit information* if, after being discarded, becomes not derivable:

$$\mathrm{Inf}_{\mathrm{Ex}}\,\varphi \;\wedge\; \langle\mathrm{Dis}_{\varphi}\rangle\,\neg\mathrm{Inf}_{\mathrm{Der}}\,\varphi$$

On the other hand, we say that φ is *entailed explicit information* if, after being discarded, the agent can still derive it:

$$\mathrm{Inf}_{\mathrm{Ex}}\,\varphi \;\wedge\; \langle\mathrm{Dis}_{\varphi}\rangle\,\mathrm{Inf}_{\mathrm{Der}}\,\varphi$$

Abductive problems of Table 3 in which $\neg\mathrm{Inf}_{\mathrm{Der}}\,\chi$ is the case (i.e., abductive problems $(*.*.\{\mathrm{a,c}\})$) can be adapted for observed explicit information. If Ω is the

[5] In fact, non-Euclidean geometries originated when going in depth into this question.

formula that represents the abductive problem labelled as **n**, then:

$$\text{Inf}_{\text{Ex}}\,\chi \;\wedge\; \langle \text{Dis}_{\chi} \rangle\, \Omega$$

is the formula that represents the version of abductive problem **n** for observed explicit information. The solutions of these problems start with $\langle \text{Dis}_{\chi} \rangle$ and then proceed as in Table 4.

3.4 Different Notions of Information

Some of the cases we have reviewed in this section can be dropped by asking for additional requirements, now on the agent's information. Two are the main constrains we can impose. The strongest one is to assume that the information is *true*; a weaker assumption is to assume that the information is simply *consistent*.

The first assumption, *truth*, corresponds syntactically to $\text{Inf}\,\varphi \to \varphi$ in the case of omniscient agents, and to $\text{Inf}_{\text{Im}}\,\varphi \to \varphi$ in the case of non-omniscient ones. Under this assumption, the agent can face *novel* abductive problems because she does not need to have complete information about the real situation, and therefore there may be a fact χ that is not part of her information. Solutions in these cases should satisfy the further requirement of preserving the information's properties, so a solution ψ for a χ-abductive problem needs to be *true*.[6] But *anomalous* abductive problems are not possible: the agent's information is true, so if she faces a χ contradicting it, she can be sure that χ is not true, and therefore can simply discard it.

The second assumption, *consistency*, corresponds syntactically to $\neg\text{Inf}\,\bot$ in the case of omniscient agents, and $\neg\text{Inf}_{\text{Im}}\,\bot$ in the case of non-omniscient ones. With this assumption, novel abductive problems are possible, but also anomalous ones since the agent's information does not need to be true. As in the previous cases, a solution should preserve the relevant property of the agent's information, that is, the solution should not create an inconsistency when added to the agent's information. This makes all solutions consistent.

Although other requirements can be imposed, these two allow us to deal with two of the most important notions of information: *knowledge*, often assumed to be *true*, and *beliefs*, often assumed to be just *consistent*.

4 A Semantic Model

Through the discussion in Sects. 2 and 3 we have observed how considering particular kinds of agents and particular notions of information gives us different forms of abductive problems. We have identified formulas expressing the existence of diverse forms of abductive problems and also expressing that certain formula is indeed a solution. This section makes part of the discussion concrete by providing a semantic model in which some the proposed cases can be evaluated. In order to make the discussion precise, we will work with the following variation of a classic example (Aliseda 2006).

[6] Observe how this makes every solution consistent.

Mary arrives late to her apartment. She presses the light switch but the light does not turn on. Knowing that the electric line is outdated, Mary assumes that it might have failed.

Let us analyse the situation. The first step in Mary's reasoning is her irrefutable observation that the light does not turn on. After observing it, this fact becomes part of Mary's *knowledge*, but it is knowledge that is not justified (i.e., supported) by the rest of her information. In other words, before the observation, Mary did not have any piece of information that would have allowed her to predict that indeed the light would not turn on. Nevertheless, she knows that the line is outdated, and therefore after observing that the light does not work, it is reasonable for her to assume that the line has failed: if she had known that before, she would have expected the lack of light. More precisely, Mary *knows* a piece of information (the light does not turn on) and she also *knows* how she could have predicted it (if the electric line has failed, then there will be no light). Then, Mary has reasons to suspect that what she would have needed to make the prediction is actually the case (she *believes* that the electric line has failed).[7]

This shows us the needed ingredients for a semantic representation of this form of abductive reasoning. First, we need to represent not only an agent's knowledge but also her beliefs. Second, we need to represent the tools the agent has to perform inferences: these tools are precisely the ones that provide her with the possible explanations for her observation.

There are several frameworks in the literature that allow us to represent an agent's knowledge and beliefs [e.g., Kraus and Lehmann (1986), Hoek (1993), Voorbraak (1993), Baltag and Smets (2008)]. The *plausibility models* of Baltag and Smets (2008) [cf. Benthem (2007)] have the further advantage of allowing us to represent the actions that transform these notions, as observations in the case of knowledge and revision in the case of beliefs. For the last ingredient, the agent's inferential tools, we can look at the approach of Velázquez-Quesada (2010) which extends plausibility models with ideas from Fagin and Halpern (1988) and Jago (2006) in order to deal with the notions of implicit and explicit information. The resulting framework allows us to represent an agent's implicit/explicit knowledge/beliefs not only about formulas but also about rules.

4.1 Plausibility-Access Models

Definition 11 (*PA language* (Velázquez-Quesada 2010)) Given a set of atomic propositions $\mathtt{P}$, formulas φ, ψ and rules ρ of the *plausibility-access (PA)* language $\mathcal{L}$ are given by

[7] Observe how this form of abductive reasoning can be seen as a particular form of belief revision driven by the agent's inferential abilities: she has observed and therefore knows ψ, but she also knows that from φ she can derive ψ, so she will revise her beliefs in order to incorporate φ into them. The relation between abductive reasoning and belief revision has been already studied, e.g., Boutilier and Becher (1995), Aliseda (2006).

$$\varphi \quad ::= p \mid \mathrm{A}\,\varphi \mid \mathrm{R}\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\sim\rangle\,\varphi \mid \langle\leq\rangle\,\varphi$$
$$\rho \quad ::= (\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi)$$

where p is an atomic proposition in $\mathtt{P}$. For the first component of the language, formulas of the form $\mathrm{A}\,\varphi$ are read as *"the agent has access to formula* φ", and formulas of the form $\mathrm{R}\,\rho$ as *"the agent has access to rule* ρ". For the modalities, $\langle\leq\rangle\,\varphi$ is read as *"there is an at least as plausible world where* φ *holds"*, and $\langle\sim\rangle\,\varphi$ as *"there is an epistemically indistinguishable world where* φ *holds"*. Other boolean connectives as well as the universal modalities $[\sim]$, $[\leq]$ are defined as usual. For the second component, a rule ρ is a pair $(\{\psi_1, \ldots, \psi_{n_\rho}\}, \varphi)$, sometimes represented as $\{\psi_1, \ldots, \psi_{n_\rho}\} \Rightarrow \varphi$ ($\psi \Rightarrow \varphi$ in case the first component is a singleton), where $\{\psi_1, \ldots, \psi_{n_\rho}\}$ is a *finite* set of formulas, the rule's *premises* $\mathsf{pm}(\rho)$, and φ is a formula, the rule's *conclusion* $\mathsf{cn}(\rho)$. We denote by $\mathcal{L}_f$ the set of formulas of $\mathcal{L}$, and by $\mathcal{L}_r$ its set of rules.

The language $\mathcal{L}$ is the propositional one extended with four new kinds of formulas. Formulas of the form $\langle\sim\rangle\,\varphi$ and $\langle\leq\rangle\,\varphi$ will allow us to deal with the notions of knowledge and belief, respectively. Formulas of the form $\mathrm{A}\,\varphi$ and $\mathrm{R}\,\rho$ will allow us to deal with the agent's explicit beliefs/knowledge about formulas and rules.

The intuitive ideas behind the semantic model for $\mathcal{L}$ are the following. First, the agent's epistemically indistinguishable worlds are not given by a plain set but rather by a set with a *plausibility* ordering; then, while knowledge is defined in terms of all epistemically indistinguishable worlds, beliefs are defined only in terms of the most plausible (i.e., the maximal) ones (Grove 1988; Boutilier 1994; Segerberg 2001). Second, a non-omniscient agent may not have access to all formulas and rules that are true and truth-preserving at each world; then, *explicit* knowledge/beliefs are defined in terms of what the agent has actually access to (Fagin and Halpern 1988; Jago 2006). Formally, the semantic model extends the plausibility models of Baltag and Smets (2008) with two functions, indicating the formulas and the rules the agent has access to at each possible world.

Definition 12 (*PA model* (Velázquez-Quesada 2010)) Let $\mathtt{P}$ be a set of atoms. A *plausibility-access (PA) model* is a tuple $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ where

- W is a non-empty set of *possible worlds*;
- $\leq\ \subseteq (W \times W)$ is a well pre-order (a locally connected and a conversely well-founded preorder), the *plausibility relation*, representing the plausibility order of the worlds from the agent's perspective ($w \leq u$ is read as "u is at least as plausible as w");
- $V : W \to \wp(\mathtt{P})$ is an *atomic valuation function*, indicating the atomic propositions in $\mathtt{P}$ that are true at each possible world;
- $\mathsf{A} : W \to \wp(\mathcal{L}_f)$ is the *access set function*, indicating the set of formulas the agent has access to at each possible world;
- $\mathsf{R} : W \to \wp(\mathcal{L}_r)$ is the *rule set function*, indicating the set of rules the agent has access to at each possible world.

A *pointed PA model* (M, w) is a *PA* model with a distinguished world $w \in W$.

The plausibility relation $\leq$ will allow us to define the agent's beliefs as what is true in the most plausible worlds. For defining the agent's knowledge, the indistinguishability relation $\sim$ is defined as the union of $\leq$ and its converse, that is, $\sim := \leq \cup \geq$. [8]

For the semantic interpretation, the two modalities $\langle\leq\rangle$ and $\langle\sim\rangle$ are interpreted via their respective relations in the standard way; the two 'access' formulas $\mathrm{A}\,\varphi$ and $\mathrm{R}\,\rho$ simply look at the A- and R-set of the evaluation point.

Definition 13 (*Semantic interpretation*) Let (M, w) be a pointed *PA* model with $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$. Atomic propositions and boolean operators are interpreted as usual. For the remaining cases,

$$
\begin{array}{llll}
(M, w) \Vdash \mathrm{A}\,\varphi & \text{iff } \varphi \in \mathsf{A}(w) & & \\
(M, w) \Vdash \mathrm{R}\,\rho & \text{iff } \rho \in \mathsf{R}(w) & & \\
(M, w) \Vdash \langle\leq\rangle\,\varphi & \text{iff} & \text{there is a } u \in \mathrm{W} \text{ such that } w \leq u & \text{and } (M, u) \Vdash \varphi \\
(M, w) \Vdash \langle\sim\rangle\,\varphi & \text{iff} & \text{there is a } u \in W \text{ such that } w \sim u & \text{and } (M, u) \Vdash \varphi
\end{array}
$$

Defining the notions. For the notion of *implicit* knowledge, the classic *EL* approach is used: the agent knows φ implicitly iff φ is true in all the worlds she considers possible from the evaluation point. For φ to be *explicitly* known, the agent needs to have access to it in all such worlds:

The agent knows *implicitly* the *formula* φ	$K_{\mathrm{Im}}\varphi := [\sim]\,\varphi$
The agent knows *explicitly* the *formula* φ	$K_{\mathrm{Ex}}\varphi := [\sim]\,(\varphi \wedge \mathrm{A}\,\varphi)$

On the other hand, the notion of beliefs does not need to look at all the epistemically indistinguishable worlds, just to those that are the most likely to be the case: an agent believes φ *implicitly* iff φ is true in the most plausible worlds under the agent's plausibility order. Given the properties of the plausibility relation, φ is true in the most plausible worlds from the evaluation point iff there is a more plausible world from which all better ones are φ-worlds (Stalnaker 2006; Baltag and Smets 2008). For its *explicit* form, it is asked for the agent to have access to φ in these maximal worlds.

The agent believes *implicitly* the *formula* φ	$B_{\mathrm{Im}}\varphi := \langle\leq\rangle\,[\leq]\,\varphi$
The agent believes *explicitly* the *formula* φ	$B_{\mathrm{Ex}}\varphi := \langle\leq\rangle\,[\leq]\,(\varphi \wedge \mathrm{A}\,\varphi)$

We have now the notions of implicit/explicit knowledge/beliefs for *formulas*. For the case of *rules*, in the implicit case, a rule ρ is translated into an implication $\mathsf{tr}(\rho)$ whose antecedent is the (finite) conjunction of the rule's premises and whose consequent is the rule's conclusion ($\bigwedge_{\psi \in \mathsf{pm}(\rho)} \psi \rightarrow \mathsf{cn}(\rho)$); in the *explicit* case, the 'access to rule' formulas are used:

[8] The indistinguishability relation should not be confused with the *equal plausibility* relation, given by the *intersection* $\leq \cap \geq$.

The agent knows *implicitly* the *rule* ρ	$K_{\mathrm{Im}}\rho := [\sim]\,\mathsf{tr}(\rho)$
The agent knows *explicitly* the *rule* ρ	$K_{\mathrm{Ex}}\rho := [\sim]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$
The agent believes *implicitly* the *rule* ρ	$B_{\mathrm{Im}}\rho := \langle\leq\rangle\,[\leq]\,\mathsf{tr}(\rho)$
The agent believes *explicitly* the *rule* ρ	$B_{\mathrm{Ex}}\rho := \langle\leq\rangle\,[\leq]\,\big(\mathsf{tr}(\rho) \wedge \mathrm{R}\,\rho\big)$

Details of this framework can be found in Velázquez-Quesada (2010). We now move on to a definition of two actions and how they can be used to represent some forms of abductive reasoning.

4.2 Operations Over Plausibility-Access Models

In a *PA* model, the agent's beliefs are given by the plausibility relation. Then, changes in beliefs can be represented by changes in this relation (Ditmarsch 2005; Benthem 2007; Baltag and Smets 2008). In particular, the act of *revising* beliefs with the aim to believe a given ϕ can be seen as a change that puts ϕ-worlds at the top of the plausibility order. There are several ways in which such a new order can be defined; each one of them can be seen as a different policy for *revising* beliefs. Here is one of the many possibilities.

Definition 14 (*Upgrade operation*) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model; let ϕ be a formula in $\mathcal{L}_f$. The *upgrade* operation yields the *PA* model $M_{\phi\Uparrow} = \langle W, \leq', V, \mathsf{A}, \mathsf{R}\rangle$, differing from M just in the plausibility order:

$$\leq' := (\leq; \phi?) \;\cup\; (\neg\phi?; \leq) \;\cup\; (\neg\phi?; \sim; \phi?)$$

The new plausibility relation, given in a *Propositional Dynamic Logic* style, states that after an explicit upgrade with ϕ, *"all ϕ-worlds become more plausible than all $\neg\phi$-worlds, and within the two zones the old ordering remains"* (Benthem 2007). More precisely, we will have $w \leq' u$ iff (1) $w \leq u$ and u is a ϕ-world, or (2) w is a $\neg\phi$-world and $w \leq u$, or (3) $w \sim u$, w is a $\neg\phi$-world and u is a ϕ-world.

PA models also have syntactic components: a set of formulas and a set of rules at each possible world. Here is an operation that modifies the first.

Definition 15 (*Uncovering operation*) Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R}\rangle$ be a *PA* model; let Φ be a set of formulas. Define $T_\Phi(M, w)$ as

$$T_\Phi(M, w) := \{\phi \mid \phi \in \Phi \text{ and } (M, w) \Vdash \phi\} \;\cup\; \{\neg\phi \mid \phi \in \Phi \text{ and } (M, w) \nVdash \phi\}$$

that is, $T_\Phi(M, w)$ is a set of formulas that contains, for each $\phi \in \Phi$, ϕ itself if ϕ holds at (M, w), and $\neg\phi$ if ϕ does not hold at (M, w).

The *uncovering* operation produces the *PA* model $M_{+\Phi} = \langle W, \leq, V, \mathsf{A}', \mathsf{R}\rangle$, differing from M just in the access set function, given now for each $w \in W$ by:

$$\mathsf{A}'(w) := \mathsf{A}(w) \;\cup\; T_\Phi(M, w)$$

This operation makes explicit the formulas in Φ. More precisely, for every $w \in W$, $\mathsf{A}(w)$ will be extended with the formulas in Φ that w satisfies and with the negations of the formulas in Φ that it does not satisfy.

For the language, we introduce the modalities $\langle \phi \Uparrow \rangle \varphi$ and $\langle +\Phi \rangle \varphi$, with their semantic interpretation given below.

Definition 16 Let $M = \langle W, \leq, V, \mathsf{A}, \mathsf{R} \rangle$ be a *PA* model. Then,

$$(M, w) \Vdash \langle \phi \Uparrow \rangle \varphi \quad \text{iff} \quad (M_{\phi\Uparrow}, w) \Vdash \varphi$$
$$(M, w) \Vdash \langle +\Phi \rangle \varphi \quad \text{iff} \quad (M_{+\Phi}, w) \Vdash \varphi$$

Based on these two operations, we will now provide an action that represents an abductive reasoning step.

4.3 An Action Representing Abductive Reasoning

The key observation of the discussed Mary's example is that the described form of abductive reasoning can be seen as a form of belief revision guided by the agent's inferential abilities: if the agent knows explicitly a rule and its conclusion, then it is reasonable for her to believe *explicitly* all the rule's premises. Following this idea, here is our proposal for an action representing this form of abductive reasoning.

Definition 17 (*Abductive reasoning with rule σ*) In order to represent abductive reasoning with a given rule σ, we introduce a new modality that allows us to build formulas of the form $\langle \mathrm{Abd}_\sigma \rangle \varphi$, read as *"the agent can perform an abductive step with rule σ after which φ is the case"*. The semantic interpretation of formulas of this form is defined in the following way:

$$\big(M, w\big) \Vdash \langle \mathrm{Abd}_\sigma \rangle \varphi \quad \text{iff} \quad \big(M, w\big) \Vdash K_{\mathrm{Ex}}\sigma \wedge K_{\mathrm{Ex}}\mathsf{cn}(\sigma) \text{ and } \big((M_{+\mathsf{pm}(\sigma)})_{P_\sigma \Uparrow}, w\big) \Vdash \varphi$$

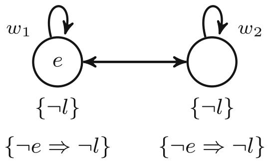
where the formula $P_\sigma := \bigwedge_{\psi \in \mathsf{pm}(\sigma)} \psi$ is the conjunction of all premises of σ (if $\mathsf{pm}(\sigma)$ is empty, then $P_\sigma := \top$).

Let us spell out this definition. The agent can perform an abductive step with rule σ after which φ is the case, $(M, w) \Vdash \langle \mathrm{Abd}_\sigma \rangle \varphi$, if and only if she knows explicitly the rule and its conclusion, $(M, w) \Vdash K_{\mathrm{Ex}}\sigma \wedge K_{\mathrm{Ex}}\mathsf{cn}(\sigma)$, and, after making explicit the premises of σ and then putting on top of the plausibility order those worlds that satisfy all of them, φ is the case, $(M_{+\mathsf{pm}(\sigma)}{}_{P_\sigma \Uparrow}, w) \Vdash \varphi$. In other words, an abductive step with a rule σ is an action that, provided that the rule and its premises are explicitly known, makes explicit σ's premises and then lifts those worlds satisfying all of them. In the particular case in which σ's premises are propositional formulas, after this sequence of operations, the agent will believe σ's premises.[9]

[9] This does not hold in the general case because the operations do change the model (the access set function and the plausibility relation), therefore affecting the truth-value of formulas that can see such change (formulas including A, $\langle \sim \rangle$ or $\langle \leq \rangle$); cf. Holliday and Icard (2010).

4.4 An Example

In the *PA* model below, each possible world shows the atomic propositions that are true at it, and its A- and R-sets appear below it. In the model, Mary knows explicitly both that the light does not turn on ($\neg l$) and that if the electric line fails ($\neg e$), then there will be no light (this rule is abbreviated as $\neg e \Rightarrow \neg l$). Nevertheless, Mary does not believe, neither explicitly nor implicitly, that the electric line fails. The formulas on the right of the diagram express all this; they are true in every world in the model, so no evaluation point is specified.



- $K_{\text{Ex}}\, \neg l \;\wedge\; K_{\text{Ex}}(\neg e \Rightarrow \neg l)$
- $\neg B_{\text{Ex}}\, \neg e$
- $\neg B_{\text{Im}}\, \neg e$

If we understand information as knowledge, this is a case of an abductive problem with explicit knowledge of the observation $\neg l$, as explained in Subsect. 3.3. The modality $\langle \text{Dis}_\chi \rangle$ represents an action that removes χ from the explicit (but not from the implicit) knowledge, so we can define it in the *PA* framework as an operation that removes χ from the A-sets of the worlds in the model [cf. the *dropping* operation of Benthem and Velázquez-Quesada (2010)]. Then, by applying $\langle \text{Dis}_{\neg l} \rangle$ to the *PA* model above, we get a similar model in which all A-sets are empty, therefore representing a situation in which, though Mary does not have explicit knowledge of $\neg l$, she has implicit knowledge about it. Now, if we understand inference as the application of explicitly known rules with explicitly known premises (Grossi and Velázquez-Quesada 2009), then Mary cannot make $\neg l$ explicit only by inference (that is, only by applying the rule $\neg e \Rightarrow \neg l$) because, even though she knows explicitly the rule, she does not know explicitly the single premise. Hence, $\neg l$ is objectively ($\text{Inf}_{\text{Im}}\, \neg l$) but not subjectively ($\neg\text{Inf}_{\text{Der}}\, \neg l$) implicit knowledge (Subsect. 3.2). In summary, the abductive problem she faces is given by

$$\langle \text{Dis}_{\neg l} \rangle \left(\neg\text{Inf}_{\text{Ex}}\, \neg l \;\wedge\; \neg\text{Inf}_{\text{Ex}}\, l \;\wedge\; \neg\text{Inf}_{\text{Der}}\, \neg l \;\wedge\; \neg\text{Inf}_{\text{Der}}\, l \;\wedge\; \text{Inf}_{\text{Im}}\, \neg l \;\wedge\; \neg\text{Inf}_{\text{Im}}\, l \right)$$

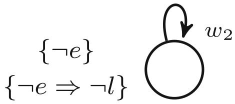
This corresponds to case (1.2.a) on Table 3 with the modification introduced in Subsect. 3.3 (χ becomes $\neg l$ and $\neg\chi$ becomes l).[10] According to Table 4 plus the modification described in Subsect. 3.3, in order for $\neg e$ to be a solution, it should satisfy

$$\langle \text{Dis}_{\neg l} \rangle \, \langle \text{Add}_{\neg e} \rangle \, \langle \neg e \Rightarrow \neg l \rangle \, \text{Inf}_{\text{Ex}}\, \neg l$$

While the effect of $\langle \text{Dis}_{\neg l} \rangle$ has been already described, $\langle \text{Add}_{\neg e} \rangle$ can be understood as a form of public announcement (Plaza 1989; Gerbrandy 1999) of $\neg e$ that removes

[10] Strictly speaking, $\neg\chi$ should become $\neg\neg l$, but using l makes the example clearer. Most importantly, as we explain further on, it is easy to provide Mary with the reasoning ability to get l from $\neg\neg l$.

the worlds where $\neg e$ is false (that is, w_1) and adds $\neg e$ to the A-sets of the remaining worlds in which the formula is true [cf. the *explicit seeing* operation of Benthem and Velázquez-Quesada (2010)]. Then, by applying $\langle \text{Dis}_{\neg l} \rangle$ and then $\langle \text{Add}_{\neg e} \rangle$ to the initial *PA*-model, we get:



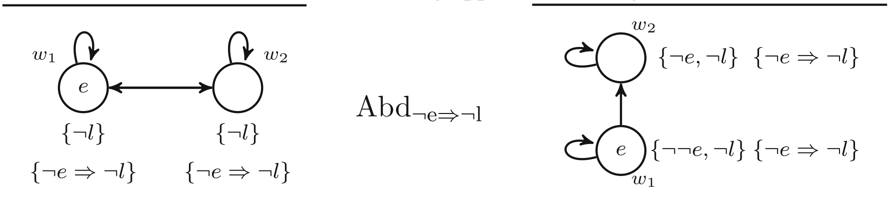
Now Mary's information state is given by

$$\neg \text{Inf}_{\text{Ex}}\, \neg l \ \wedge\ \neg \text{Inf}_{\text{Ex}}\, l \ \wedge\ \text{Inf}_{\text{Der}}\, \neg l \ \wedge\ \neg \text{Inf}_{\text{Der}}\, l \ \wedge\ \text{Inf}_{\text{Im}}\, \neg l \ \wedge\ \neg \text{Inf}_{\text{Im}}\, l$$

which corresponds to the abductive problem (1.2.b) of Table 3, as indicated in Table 4. Mary has now subjectively implicit information of $\neg l$; hence she only needs to apply her reasoning abilities to make $\text{Inf}_{\text{Ex}}\, \neg l$ true. Thus, $\neg e$ is indeed an abductive solution for the problem.

Following a classic terminology in Philosophy of Science, by proving that $\neg e$ satisfies all the requisites to be a good explanation we have only dealt with the *context of justification* of abductive reasoning. There are three important points still to be discussed. First, where should the agent look for the possible explanations of the surprising observation (what is commonly called the *context of discovery* of abductive reasoning)? Our proposal is that candidates for solutions of a given χ-abductive problem can be found in the premises of those rules that would have allowed the agent to derive χ. Second, though in our example there is only one candidate for a solution, in general there will be more than one explanation for a given abductive problem, and in such cases the agent has to choose one (some) of them. Our proposal for dealing with this stage of the abductive process will be presented in Sect. 5. Third, once that the agent has chosen her explanation(s), she should somehow incorporate it(them) into her information. For this we will use the abductive operation of Definition 17 in which the explanation is not incorporated as knowledge (though plausible, the chosen explanation does not need to be the case) but rather as *belief*.

Back in our story, Mary has observed that the light does not turn on, therefore reaching an information state described in the left model of the diagram below. In order to explain the lack of light, Mary applies abductive reasoning with the rule $\neg e \Rightarrow \neg l$ (we have already justified why $\neg e$ is a good candidate for an explanation). The model that results from this reasoning appears on the right.



There are two changes in the new model with respect to the original one. First, the unique premise of the chosen rule has been made explicit by adding it to the A-set of w_2 and by adding its negation to the A-set of w_1. Second, there is a new plausibility ordering: now the unique $\neg e$-world, w_2, has become more plausible than the unique e-world, w_1. In this resulting model Mary still knows explicitly that there is no light

($K_{\text{Ex}}\, \neg l$) and still knows explicitly the rule that links that with the failure of the electric line ($K_{\text{Ex}}(\neg e \Rightarrow \neg l)$). But, as a result of abductive reasoning, Mary now believes (both implicitly and explicitly) that the electric line has failed ($B_{\text{Ex}}\, \neg e \;\wedge\; B_{\text{Im}}\, \neg e$).

Note how even if the original model had w_2 above w_1 as the sentence *"Knowing that the electric line is outdated ..."* in the original statement (page 14) may suggest, the operation does makes a difference since it makes explicit the premises of the chosen rule.

5 Selecting the Best Explanation

So far our proposal does not deal with the process of choosing the best explanation. In general, there will be more than one solution for a given abductive problem: Mary observed the lack of light with a failure in the electric line, but it could be also the case that the switch or even the bulb was broken. The selection of *the best explanation* is a fundamental problem in abductive reasoning [see, e.g. Lipton (1991) and Hintikka (1998)], and in fact many authors consider it the heart of abductive reasoning. But beyond logical requisites to avoid triviality [e.g., Definition 3], the definition of suitable criteria is still an open problem. In this section we adapt the ideas presented in Nepomuceno-Fernández et al. (2013) to our non-omniscient agents case.

Some typical selection criteria are restrictions on the logical form of the solutions, so that only *abducible* formulas are selected as possible explanations. For example, if we only allow conjunctions of literals as abductive solutions, a criterion for *minimal* explanation can be established on the length of these conjuncts. But finer criteria to select between two equally valid solutions require *contextual aspects* (Aliseda 2006). A typical option is the use of *preferential models* (e.g., Makinson (2003)), but such approaches are often criticized because they introduce an external resource.

Our subjective approach to abductive reasoning gives us a different perspective. The electric line in Mary's house is outdated; knowing this, Mary considers its failure more likely than other possible explanation, and hence she explains the lack of light by assuming that the electric line has failed. But Mary's friend Gaby does not know that the line is old, and therefore she explains the same fact by assuming what she considers more likely to happen: the bulb has failed. The two explanations are equally "logical" since a failure on the electric line or a broken bulb would be enough to explain why the light does not turn on; what makes Mary to choose the first and Gaby choose the second is that they have different knowledge and different beliefs. This suggests that instead of looking for criteria to select *the* best explanation, we should rather look for criteria to select *the agent*'s best explanation.

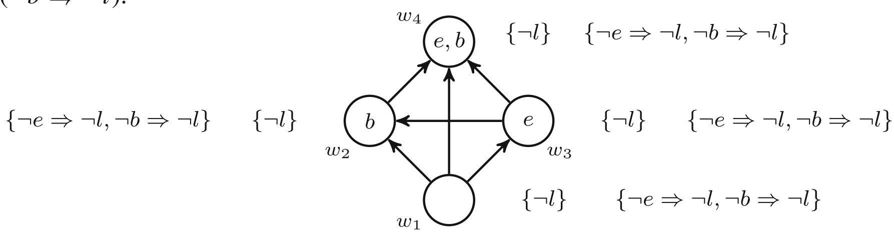
More precisely, and as the example shows, the explanation an agent will choose for a given observation depends not only on the logical properties of the candidates, but also on what the agent herself knows and considers more likely to be the case. This criteria is not "logical" in the classical sense, since it is not based exclusively on what can be derived from the explanation or even the number of needed inference steps. Nevertheless, it is logical in a broader sense since it depends on the agent's information: her knowledge and her beliefs.

Here is how this idea reflects on the previous section's proposal. We have suggested that the agent's reasonable candidates for explanation of an observed χ are the premises of those rules that would allow her to derive χ. In general, the agent might have several ways in which she could have derived the observed χ (she might have several rules with χ as conclusion). But the agent already has knowledge and beliefs, so she already considers, implicitly or explicitly, some of these explanations more likely than the others. In our example, Mary explains the light problem by assuming that the line is failing not because she does not consider the possibility that the bulb might be broken, but because after the light does not turn on, the failure on the line is, from her perspective, more likely to be the case. Consider this extended version of Mary's adventure.

> Mary arrives late to her apartment. She presses the light switch but the light does not turn on. Knowing that the electric line is outdated, Mary assumes that it might have failed, but then she takes a quick look at the alarm clock on her night table and realizes that it is working. This makes her to change her mind, and now she assumes that the bulb is broken.

After observing that the light does not turn on, Mary has two hypotheses: *"the electric line fails"* and *"the bulb is broken"*. She chooses the first as her explanation because at her apartment it is more common than the second. But then she observes that the alarm clock works; since this contradicts her first hypothesis, she discards the first and selects the second.

In order to show how we can capture Mary's reasoning in the described framework, consider the *PA* model shown below, representing Mary's knowledge and beliefs immediately after she observes (and therefore knows explicitly) that the light is not working. In the model (with reflexive arrows omitted), she has also explicit knowledge of the following rules: *"if the electric line fails* ($\neg e$) *then the light does not work* ($\neg l$)" ($\neg e \Rightarrow \neg l$) and *"if the bulb is broken* ($\neg b$) *then the light does not work* ($\neg l$)" ($\neg b \Rightarrow \neg l$).



The most plausible world (w_4) is the one in which nothing fails, as it is usual at Mary's apartment. Nevertheless, from Mary's perspective, a problem with the bulb is less plausible than a failure in the electric line (w_3 is below w_2), and the least plausible situation is the one in which both fail (w_1 is at the bottom). This is Mary's personal plausibility ordering, according to her previous experience, and it will be the key for selecting *her* best explanation. Observe also how though Mary believes that nothing fails (w_1 is the most plausible world), this belief is just *implicit* and not *explicit*: the information about the failure of the electric line or the failure of the bulb is just implicit because the A-sets do not contain formulas about them.

After the observation, Mary does not have an explanation for $\neg l$: she does not know/believe explicitly anything that entails $\neg l$ with the rules she knows (in particular, she knows/believes explicitly neither $\neg e$ nor $\neg b$). Thus, there is an abductive problem, and in order to solve it Mary can perform an abductive step with any of the two rules she knows explicitly: $\neg e \Rightarrow \neg l$ or $\neg b \Rightarrow \neg l$. Our proposal here is that Mary does not need to choose one of these rules; what she can do is to consider *all of them*, and then allow her previous plausibility order to decide which will be the chosen explanation, that is, which explanation she will believe explicitly. This solution is simpler but nevertheless more natural since the selection of the best explanation relies on Mary's previous beliefs. In the following section we provide the definitions to make precise this idea.

5.1 A General Action Representing Abductive Reasoning

Definition 18 (*Abductive reasoning*) In order to represent general abductive reasoning, we introduce a new modality that allows us to build formulas of the form $\langle \mathrm{Abd}_\chi \rangle\, \varphi$, read as *"the agent can perform an abductive step for formula* χ *after which* φ *is the case"*. In order to provide the semantic interpretation of these formulas, we first make the following definitions.

Denote by R_χ the set of the rules explicitly known by the agent at the pointed model (M, w) that have χ as conclusion, that is,

$$R_\chi := \left\{ \sigma \in \mathcal{L}_r \mid (M, w) \Vdash K_{\mathrm{Ex}}\sigma \text{ and } \mathsf{cn}(\sigma) = \chi \right\}$$

Then denote by P_{R_χ} the disjunction of the conjunction of all premises of each rule in $R\chi$, that is,

$$P_{R_\chi} := \bigvee_{\sigma \in R_\chi} \bigwedge_{\psi \in \mathsf{pm}(\sigma)} \psi$$

Finally, denote by $\mathsf{pm}(R_\chi)$ the set of all premises of rules in $R\chi$:

$$\mathsf{pm}(R_\chi) := \bigcup_{\sigma \in R_\chi} \mathsf{pm}(\sigma)$$

If $R_\chi \neq \emptyset$, the semantic interpretation of formulas of the form $\langle \mathrm{Abd}_\chi \rangle\, \varphi$ is defined in the following way:

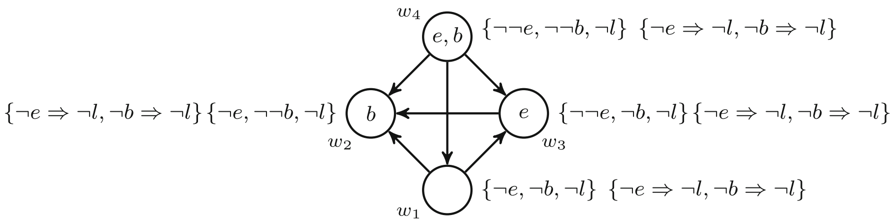
$$\big(M, w\big) \Vdash \langle \mathrm{Abd}_\chi \rangle\, \varphi \;\text{ iff }\; \big(M, w\big) \Vdash K_{\mathrm{Ex}}\chi \quad \text{and} \quad \big((M_{+\mathsf{pm}(R_\chi)})_{P_{R_\chi}\Uparrow}, w\big) \Vdash \varphi$$

Let us spell out the semantic interpretation. If the agent knows explicitly at least one rule with χ as conclusion (the $R_\chi \neq \emptyset$ requirement), then she can perform an abductive step for formula χ after which φ is the case, $(M, w) \Vdash \langle \mathrm{Abd}_\chi \rangle\, \varphi$, if and only if she knows explicitly χ, $(M, w) \Vdash K_{\mathrm{Ex}}\chi$, and after making explicit the all premises of all rules for χ she knows explicitly and then lifting those worlds that satisfy all the premises of at least one such rule, φ is the case, $(M_{+\mathsf{pm}(R_\chi)}{}_{P_{R_\chi}\Uparrow}, w) \Vdash \varphi$.

Observe the effect of this abductive step (i.e., of the sequence *uncovering* then *upgrade*), again for the case in which the premises of all such rules are propositional formulas. It will make explicit the premises of all the relevant rules and then it will lift those worlds that satisfy the conjunction of all the premises of at least of one of such rules (that is, P_{R_χ}). But recall the effect of the upgrade operation with P_{R_χ}: *"all P_{R_χ}-worlds become more plausible than all $\neg P_{R_\chi}$-worlds, and within the two zones the old ordering remains"*. Then, in the upper zone, the worlds that will be at the top will be those that satisfy P_{R_χ} and were already more plausible than the rest. Hence, the agent will believe the premises that were already more plausible for her.

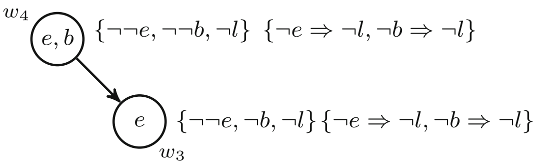
5.2 The Example Revisited

We left Mary trying to explain why the light does not work. She has two rules with $\neg l$ as conclusion, $\neg e \Rightarrow \neg l$ and $\neg b \Rightarrow \neg l$. Hence, after $\langle \text{Abd}_{\neg l} \rangle$, the formulas $\neg e$, e, $\neg b$, b will become explicit and, more importantly, there will be a new plausibility ordering among the worlds: those satisfying $\neg e \vee \neg b$ (w_1, w_2 and w_3) become more plausible than those satisfying its negation (w_4). Still, within the upper zone, the old ordering will be kept: w_2 will be still the most plausible among these three. This gives us the following model:



Observe how $\langle \text{Abd}_{\neg l} \rangle\, B_{\text{Ex}} \neg e$ holds: now Mary believes that an electric problem is responsible for the lack of light.

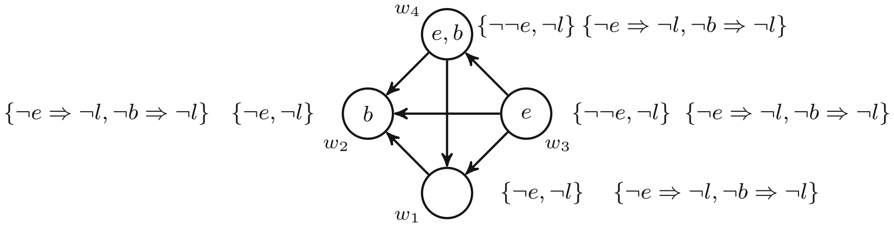
Mary's best explanation relies on her previous beliefs. This can lead to incorrect explanations, as in our case: when she observes the alarm clock working, she realizes that the electric line has not failed. This observation of $\neg\neg e$ can be represented by $\langle \text{Add}_{\neg\neg e} \rangle$ [11], thus producing the following model:



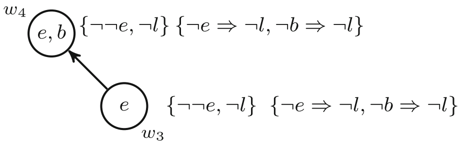
[11] Such action is more naturally represented as the announcement of e, but in our model Mary has explicitly $\neg\neg e$. There are two options here. The most elegant one is to introduce an action representing Mary's inference from $\neg\neg e$ to e [see, again, Velázquez-Quesada (2010)], and then work with the announcement in its most natural form (e). Due to space reasons, here we have chosen to work with this special form of observing that the electric line works.

The failed hypothesis emphasizes the fact that abduction is a non-monotone form of reasoning: the explanation chosen as the best can be refuted by further information, as Mary knows now. But note how in the model that results from observing $\neg\neg e$, the abductive problem $\neg l$ is still explained, only this time the explanation is $\neg b$. Before observing $\neg\neg e$, Mary explicitly believed $\neg e$ as the explanation of $\neg l$, but after the observation she believes explicitly $\neg b$. She does not need a further abductive step, as the first one reordered her beliefs to make more plausible the worlds where *some* explanation for $\neg l$ is true. Then, just as the original explanation was chosen according to her previous beliefs, once these beliefs change, a new explanation appears.

This does not happen if we apply the action $\langle \text{Abd}_\sigma \rangle$ of the previous section. If Mary chooses $\neg e \Rightarrow \neg l$ (according to her original preference for $\neg e$ over $\neg b$), the model that results from applying $\langle \text{Abd}_{\neg e \Rightarrow \neg l} \rangle$ is:



Then, a further observation of $\neg\neg e$ produces the following model in which $\neg l$ is unexplained, that is, Mary does not have any explicit information that would allow her to derive $\neg l$.



This example shows two characteristics of $\langle \text{Abd}_\chi \rangle$. First, the 'best explanation' is selected according to the previous beliefs of the agent, thus given us *the agent's* best explanation. Second, as these beliefs change, the best explanation changes too, without needing further abductive steps. In fact, if $[\psi!]$ represents the observation of a propositional formula ψ (modelled as a public announcement), and the rules that the agent knows for χ contain only propositional formulas, the following holds:

$$\Vdash \langle \text{Abd}_\chi \rangle [\psi!] \varphi \;\leftrightarrow\; [\psi!] \langle \text{Abd}_\chi \rangle \varphi$$

The formula is valid, first because $\langle \text{Abd}_\chi \rangle$ does not affect the truth-value of propositional formulas, and thus the worlds $[\psi!]$ removes when applied after $\langle \text{Abd}_\chi \rangle$ are exactly those it removes when applied before $\langle \text{Abd}_\chi \rangle$; second, because the plausibility order among ψ-worlds is not affected by $[\psi!]$, so $\langle \text{Abd}_\chi \rangle$ has the same effect regardless whether it takes place before or after $[\psi!]$; third, because the effect of these two actions on A-sets is the same, regardless of the order in which they are applied; fourth, because none action changes the rules the agent explicitly knows.[12]

[12] If ψ is not propositional, the effect of observing it is not interchangeable with an abductive step; see Holliday and Icard (2010).

The possibility of interchanging abductive steps and observations is a very interesting property. As the example of Mary shows, new observations after the abductive step may change the hypothesis that is believed as the best explanation. The just shown property guarantees that after new observations there is no need of further abductive steps: the agent's information (and so her beliefs about the best explanation) is the same if she performs the abductive operations before or after observing new information.

6 Summary and Future Work

We have presented definitions of abductive problem and abductive solution in terms of an agent's information and the way it changes, showing how different examples of abductive reasoning correspond to different kinds of agents (e.g., those whose information is not closed under logical consequence, or those whose reasoning abilities are not complete). Then we have shown how a syntactic extension of the possible worlds model allows us to represent certain forms of abductive reasoning.

Our work is just an initial exploration of abductive reasoning for non-omniscient agents, and there are many aspects yet to be studied. The most important one is the semantic representation of abductive reasoning, and though we have shown that certain forms can be represented, a deep study of the diverse types that can be described is still pending. In order to achieve this, we first should provide a proper semantic definition for the actions we have sketched through the work: $\text{Add}_{\varphi/\alpha}$, Rem_{φ}, α and Dis_{φ}. Natural candidates can be found in the current *DEL* literature (Plaza 1989; Gerbrandy 1999; Benthem 2007; Baltag and Smets 2008; Velázquez-Quesada 2010). Once a full setting has been formally defined, the next step is a proper comparison between it and other approaches to abductive reasoning, in order to identify strengths and weaknesses and, more importantly, in order to obtain a better picture of the abductive process.

Acknowledgments We thank Johan van Benthem and Ángel Nepomuceno-Fernández for their valuable comments and suggestions on early versions of this paper. We acknowledge support from the project FFI2011-15945-E (Ministerio de Economía y Competitividad, Spain) and the Excellence Research Project of the Junta de Andalucía P10-HUM-5844. We would also like to express our gratitude to the anonymous referees.

References

Aliseda, A. (2006). *Abductive reasoning. Logical investigations into discovery and explanation, synthese library series* (Vol. 330). Berlin: Springer.

Baltag, A., & Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, & M. Wooldridge (Eds.), *Logic and the foundations of game and decision theory (LOFT7), texts in logic and games, vol. 3*, (pp. 13–60). Amsterdam University Press.

Boissier, O., Seghrouchni, A. E. F., Hassas, S., Maudet, N. (eds). (2010). *Proceedings of the multi-agent logics, languages, and organisations federated workshops (MALLOW 2010), vol. 627*, CEUR Workshop Proceedings, Lyon, France.

Boutilier, C. (1994). Toward a logic for qualitative decision theory. In J. Doyle, E. Sandewall, & P. Torasso (Eds.), *KR 94* (pp. 75–86). Bonn, Germany: Morgan Kaufmann.

Boutilier, C., & Becher, V. (1995). Abduction as belief revision. *Artificial Intelligence*, *77*(1), 43–94. doi:10.1016/0004-3702(94)00025-V.

Fagin, R., & Halpern, J. Y. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, *34*(1), 39–76. doi:10.1016/0004-3702(87)90003-8.
Gärdenfors, P. (Ed.). (1992). *Belief revision. no. 29 in Cambridge tracts in theoretical computer science*. Cambridge: Cambridge Press.
Gerbrandy, J. (1999). Dynamic epistemic logic. In L. S. Moss, J. Ginzburg, & M. de Rijke (Eds.), *Logic, language and computation* (Vol. 2, pp. 67–84). Stanford, CA: CSLI Publications.
Grossi, D., Velázquez-Quesada, F. R. (2009). Twelve Angry Men: A study on the fine-grain of announcements. In: X. He, J. F. Horty, E. Pacuit (eds.) *LORI, Springer, lecture notes in computer science, vol. 5834*, pp. 147–160. doi:10.1007/978-3-642-04893-7_12.
Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, *17*(2), 157–170. doi:10.1007/BF00247909.
Hartshorne, C., Weiss, P. (eds). (1934). *Collected papers of Charles S. Peirce, vol. V: Pragmatism and pramaticism*. Harvard Universit Press, Cambridge.
Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.
Hintikka, J. (1998). What is abduction? The fundamental problem of contemporary epistemology. *Transactions of the Charles S Peirce Society*, *34*(3), 503–533.
Holliday, W. H., & Icard, T. F. (2010). Moorean phenomena in epistemic logic. In L. Beklemishev, V. Goranko, & V. Shehtman (Eds.), *Advances in modal logic* (pp. 178–199) College Publications.
Jago, M. (2006). Rule-based and resource-bounded: A new look at epistemic logic. In T. Ågotnes & N. Alechina (Eds.), *Proceedings of the workshop on logics for resource-bounded agents, organized as part of the 18th European summer school on logic, language and information (ESSLLI)* (pp. 63–77). Spain: Malaga.
Kraus, S., & Lehmann, D. J. (1986). Knowledge, belief and time. In: L. Kott (ed.) *ICALP, Springer, lecture notes in computer science, vol. 226*, pp. 186–195.
Levesque, H. J. (1984). A logic of implicit and explicit belief. In: *Proceedings of AAAI-84*, Austin, TX, pp. 198–202.
Lipton, P. (1991). *Inference to the best explanation*. New York: Routledge.
Makinson, D. (2003). Bridges between classical and nonmonotonic logic. *Logic Journal of the IGPL*, *11*(1), 69–96.
Nepomuceno-Fernández, A., Soler-Toscano, F., & Velázquez-Quesada, F. R. (2013). An epistemic and dynamic approach to abductive reasoning: Selecting the best explanation. *Logic Journal of IGPL*, *21*(6), 943–961. doi:10.1093/jigpal/jzt013.
Peirce, C. S. (1911). A letter to J. H. Kehler, nem 3:203–204, 1911. Available at http://www.helsinki.fi/science/commens/terms/retroduction.html.
Plaza, J. A. (1989). Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, & Z. W. Ras (Eds.), *Proceedings of the 4th international symposium on methodologies for intelligent systems*, Oak Ridge National Laboratory, ORNL/DSRD-24, Tennessee, USA, pp. 201–216.
Segerberg, K. (2001). The basic dynamic doxastic logic of AGM. In: Williams and Rott (2001), pp. 57–84.
Soler-Toscano, F., & Velázquez-Quesada, F. R. (2010). Abduction for (non-omniscient) agents. In: Boissier et al. (2010).
Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, *128*(1), 169–199. doi:10.1007/s11098-005-4062-y.
van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, *17*(2), 129–155.
van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge: Cambridge University Press.
van Benthem, J., & Velázquez-Quesada, F. R. (2010). The dynamics of awareness. *Synthese (Knowledge, Rationality and Action)*, *177*(1), 5–27. doi:10.1007/s11229-010-9764-9.
van der Hoek, W. (1993). Sytems for knowledge and belief. *Journal of Logic and Computation*, *3*(2), 173–195.
van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthese*, *147*(2), 229–275. doi:10.1007/s11229-005-1349-7.
van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic, synthese library series, vol. 337* Berlin: Springer.
Vardi, M. Y. (1986). On epistemic logic and logical omniscience. In J. Y. Halpern (Ed.), *TARK* (pp. 293–305). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Velázquez-Quesada, F. R. (2010). Dynamic epistemic logic for implicit and explicit beliefs. In: Boissier et al. (2010).

Velázquez-Quesada, F. R., Soler-Toscano, F., & Nepomuceno-Fernández, A. (2013). An epistemic and dynamic approach to abductive reasoning: Abductive problem and abductive solution. *Journal of Applied Logic*, *11*(4), 505–522. doi:10.1016/j.jal.2013.07.002.

Voorbraak, F. (1993). As far as I know: Epistemic logic and uncertainty. PhD thesis, Utretch University, Utretch, The Netherlands.

Williams, M. A., & Rott, H. (Eds.). (2001). *Frontiers in belief revision. No. 22 in applied logic series*. Dordrecht: Kluwer.