

# Abduction for (non-omniscient) agents

Fernando Soler-Toscano and Fernando R. Velázquez-Quesada

## 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, making special emphasis on non-ideal agents. We begin by exploring what an abductive problem is in terms of an agent's information, and what an abductive solution is in terms of the actions that modify it. Then we explore the different kinds of abductive problems and abductive solutions that arise when we consider agents whose information is not closed under logical consequence, and agents whose reasoning abilities are not complete.

## 1 Abductive reasoning

*Beyond the obvious facts that he has at some time done manual labour, that he takes snuff, that he is a Freemason, that he has been in China, and that he has done a considerable amount of writing lately, I can [get] nothing else.*

*Sherlock Holmes  
The Red-Headed League*

Among the non-monotonic reasoning processes, abduction [1] is one of the most important. Usually described as the process of *looking for an explanation*, it has been recognized as one of the most commonly used in our daily activities. 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 he (Mr. Wilson) has done a considerable amount of writing lately. Given the symptoms  $A$  and  $B$ , a doctor suspects that the patient suffers from  $C$ . 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.

But though traditional examples of abductive reasoning are given in terms of an agent's information and its changes, classical 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. After recalling the classical definitions of an abductive problem and an abductive solution (the rest of the current section), we explore what an abductive problem is in terms of the agent's information, and what an abductive solution is in terms of the actions that modify it (Section 2). Then we focus on non-ideal agents, analyzing not only the cases that arise when the agent's information is not closed under logical consequence (Section 3) but also those that arise when the agent's reasoning abilities are not complete (Section 4). We finish with a summary, proposing lines for further work (Section 5).

In this paper we will use the term *information* in the most general sense, with the notions of *knowledge* or *belief* being particular instances that impose further restrictions, like truth or consistency. Moreover, though we will use formulas in *Epistemic Logic* (EL; [9]) and *Dynamic Epistemic Logic* style (DEL; [5]), we will not commit ourselves to any particular semantic model. The main goal of this work is to explore the possibilities and concepts that emerge from a dynamic epistemic analysis of abductive reasoning.

## 1.1 The classical approach to abduction

Traditionally, it is said that there is an abductive problem when there is a formula  $\chi$  that is not predicted by the current theory  $\Phi$ . Recently, it has been observed that, even if the theory does not entail  $\chi$ , it might entail its negation. Following [1], we can identify two basic abductive problems.

**Definition 1.1** (Abductive problem). Let  $\Phi$  and  $\chi$  be a theory and a formula, respectively, in some language  $\mathcal{L}$ . Let  $\vdash$  be a consequence relation on  $\mathcal{L}$ .

The pair  $(\Phi, \chi)$  is a *novel abductive problem* when neither  $\chi$  nor  $\neg\chi$  are consequences of  $\Phi$ , i.e., when

$$\Phi \not\vdash \chi \quad \text{and} \quad \Phi \not\vdash \neg\chi$$

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

$$\Phi \not\vdash \chi \quad \text{and} \quad \Phi \vdash \neg\chi$$

Traditionally, a solution for an abductive problem  $(\Phi, \chi)$  is a formula  $\psi$  that, together with  $\Phi$ , entails  $\chi$ . This solves the problem because now the theory is strong enough to *explain*  $\chi$ . The *anomalous* case requires an extra initial step, since adding directly such  $\psi$  will make the theory to entail both  $\chi$  and  $\neg\chi$ . The agent should perform first a *theory revision* that stop  $\neg\chi$  from being a consequence of  $\Phi$ . Here are the formal definitions.

**Definition 1.2** (Abductive solution).

- Given a *novel* abductive problem  $(\Phi, \chi)$ , the formula  $\psi$  is an *abductive solution* if

$$\Phi, \psi \vdash \chi$$

- Given an *anomalous* abductive problem  $(\Phi, \chi)$ , the formula  $\psi$  is an *abductive solution* if it is possible to perform a theory revision to get a *novel* problem  $(\Phi', \chi)$  for which  $\psi$  is a solution.

In some cases, Definition 1.2 is too weak since it allows trivial solutions, like  $\chi$  itself. Again, following [1], it is possible to make a further classification.

**Definition 1.3** (Classification of abductive solutions). Let  $(\Phi, \chi)$  be an abductive problem. An abductive solution  $\psi$  is

<i>consistent</i> if	$\Phi, \psi \not\vdash \perp$
<i>explanatory</i> if	$\psi \not\vdash \chi$
<i>minimal</i> if, for every other abductive solution $\varphi$ ,	$\psi \vdash \varphi$ implies $\varphi \vdash \psi$

The consistency requirement discards those  $\psi$  inconsistent with  $\Phi$  and the explanatory requirement discards  $\chi$  itself. Minimality works as the Occam's razor, asking for the solution  $\psi$  to be logically equivalent to any other solution it implies.

## 2 From an agent's perspective

Most of the examples of abductive reasoning involve an agent and its information. It is Holmes who observes that Mr. Wilson's right cuff is very shiny; it is a doctor who observes the symptoms  $A$  and  $B$ ; it is Karen who observes that the grass is wet. So when does an agent has an abductive problem  $(\Phi, \chi)$ ? By interpreting  $\Phi$  as the agent's information, we get the following definitions. We use formulas in *EL* style, where  $\text{Inf } \varphi$  is read as " $\varphi$  is part of the agent's information".

**Definition 2.1** (Subjective abductive problem). Let  $\chi$  be a formula.

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

$$\neg\text{Inf } \chi \wedge \neg\text{Inf } \neg\chi \tag{1}$$

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

$$\neg\text{Inf } \chi \wedge \text{Inf } \neg\chi \tag{2}$$

So an agent has a  $\chi$ -abductive problem when  $\chi$  is not part of her information. What about an abductive solution? Definition 1.2 states that  $\psi$  is a solution to a novel problem if, when added to the theory  $\Phi$ , we get a theory that entails  $\chi$ . But a theory is actually closed under logical consequence, so  $\psi$  is a solution if, when added to the theory, makes  $\chi$  part of the theory too. The *anomalous* case needs another step, since a revision is required first.

We have identified  $\Phi$  with the agent's information. Then, a solution for the subjective *novel* case is a formula  $\psi$  that, when added to the agent's information, makes the agent informed about  $\chi$ . This highlights the fact the requisites of a solution involve an *action*; an action that changes the agent's information by adding  $\psi$  to it. In the subjective *anomalous* case, the action was already clear, since the theory should be modified. But now we can see that the requisites for this case involves *two* actions: removing a piece of information and then incorporating a new one.

We will express changes in the agent's information by using formulas in *DEL* style. In particular, formulas of the form  $\langle \text{Add}_\phi \rangle \varphi$  will be read as " $\phi$  can be added to the agent's information and, after that,  $\varphi$  is the case", and formulas of the form  $\langle \text{Rem}_\phi \rangle \varphi$  will be read as " $\phi$  can be removed from the agent's information and, after that,  $\varphi$  is the case".

**Definition 2.2** (Subjective abductive solution). Suppose an agent has a *novel*  $\chi$ -abductive problem, that is,  $\neg \text{Inf } \chi \wedge \neg \text{Inf } \neg \chi$  holds. A formula  $\psi$  is an *abductive solution* to this problem if, when added to the agent's information, the agent becomes informed about  $\chi$ . In a formula,

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

Now suppose the agent has an *anomalous*  $\chi$ -abductive problem, that is,  $\neg \text{Inf } \chi \wedge \text{Inf } \neg \chi$  holds. A formula  $\psi$  is an *abductive solution* to this problem if the agent can revise her information to remove  $\neg \chi$  from it and, after it, the incorporation of  $\psi$  makes  $\chi$  part of her information. In a formula,

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

What about the further classification for abductive solutions? We can also provide formulas that characterize them.

**Definition 2.3** (Classification of subjective abductive solutions). Suppose an agent has a  $\chi$ -abductive problem. A formula  $\psi$  is a(n)

- *consistent* abductive solution if it is a solution and can be added to the agent's information without making the latter inconsistent:

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

- *explanatory* abductive solution if it is a solution and it does not imply  $\chi$ , that is, it only *complements* the agent's information to produce  $\chi$ :

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

- *minimal* abductive solution if it is a solution and, for any  $\varphi$ , if  $\varphi$  is a solution that becomes part of the agent's information after  $\psi$  is added, then  $\psi$  also becomes part of the agent's information after  $\varphi$  is added.

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

### 3 A non-omniscient agent

In the classical definition of an abductive problem, the set of formulas  $\Phi$  is understood as a *theory*, usually assumed to be closed under logical consequence, as we mentioned before. If this is the case, then we have actually revised an omniscient case.

But our agent does not need to be ideal. And if the agent's information is not closed under logical consequence, then we should make a difference between the information she actually has, her *explicit* information, and what follows logically from it, her *implicit information* [12; 11; 15].

#### 3.1 Abductive problems

A non-omniscient agent has an abductive problem whenever  $\chi$  is not part of her *explicit* information. As a consequence, the modality  $\text{Inf}$  in Definition 2.1 becomes  $\text{Inf}_{\text{Ex}}$ . But now each of our two abductive problems splits into four, according to the agent's *implicit* information ( $\text{Inf}_{\text{Im}}$ ) about  $\chi$  and  $\neg\chi$ . These eight cases include inconsistent situations in which the agent is implicitly informed about both  $\chi$  and  $\neg\chi$ . They can be discarded under particular interpretations of the agent's information, like *knowledge* or *consistent beliefs*, but we have chosen to keep them here for the sake of generality.

Still, not all these cases are possible. We have said that *implicit* information is what follows logically from the *explicit* one, so explicit information itself should be implicit information, that is,

$$\text{Inf}_{\text{Ex}} \varphi \rightarrow \text{Inf}_{\text{Im}} \varphi$$

By assuming this formula, we can drop the cases in which some formula is in the agent's explicit information, but not in her implicit one.

**Definition 3.1** (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.

$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge$	{	$\neg\text{Inf}_{\text{Im}} \chi \wedge \neg\text{Inf}_{\text{Im}} \neg\chi$	<b>(1.1)</b>
		$\text{Inf}_{\text{Im}} \chi \wedge \neg\text{Inf}_{\text{Im}} \neg\chi$	<b>(1.2)</b>
		$\neg\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(1.3)</b>
		$\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(1.4)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \text{Inf}_{\text{Ex}} \neg\chi \wedge$	{	$\neg\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(2.3)</b>
		$\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(2.4)</b>

Table 1: Abductive problems for non-omniscient agents.

Let us review each one of the *novel* cases. In case **(1.1)**, the *truly novel* one, the agent lacks explicit and implicit information about both  $\chi$  and  $\neg\chi$ ; the formula  $\chi$  is a real novelty for her. But in case **(1.2)**, the *not implicit novelty* one, though the agent does not have explicit information about neither  $\chi$  nor  $\neg\chi$ , she has implicit information about  $\chi$ . In other words,  $\chi$  is a novelty for the agent's explicit information, but not for her implicit one since  $\chi$  follows logically from what she explicitly has. In case **(1.3)**, the *implicitly anomaly* one, the agent lacks explicit information about both  $\chi$  and  $\neg\chi$ , but implicitly she is informed about  $\neg\chi$ . Finally we have the *implicitly inconsistent* case, **(1.4)**, in which the agent lacks explicit information about both  $\chi$  and  $\neg\chi$ , but has an implicit inconsistency.

Now for the *anomalous* cases. Case **(2.3)** is the *truly anomalous* one: the agent has both explicit and implicit information about  $\neg\chi$ , and lacks both explicit and implicit information about  $\chi$ . In the remaining one **(2.4)**, called *anomaly with implicit inconsistency*, though the agent has explicit information about  $\neg\chi$  but not about  $\chi$ , the latter follows from her explicit information.

**Omniscience as a particular case** An agent is omniscient when she has explicitly all her implicit information. With this extra requirement, expressed by the formula  $\text{Inf}_{\text{Im}} \varphi \rightarrow \text{Inf}_{\text{Ex}} \varphi$ , cases **(1.2)**, **(1.3)**, **(1.4)** and **(2.4)** can be discarded since explicit and implicit information do not coincide. This leaves us only with cases **(1.1)** and **(2.3)**; exactly the two cases of Definition 2.1.

### 3.2 Abductive solutions

We have defined non-omniscient  $\chi$ -abductive problems as situations in the agent is not *explicitly* informed about  $\chi$ . Accordingly, for defining a solution, we will look for an action (or a sequence of them) that makes the agent *explicitly* informed about  $\chi$ , without having neither implicit nor explicit information about  $\neg\chi$ . We will focus on cases **(1.1)**, **(1.2)**, **(1.3)** and **(2.3)**, leaving the inconsistent ones, **(1.4)** and **(2.4)**, for future work.

Consider the *truly novel* case **(1.1)**: the agent lacks explicit and implicit information about both  $\chi$  and  $\neg\chi$ . Then, just like in the omniscient case, a solution is a formula  $\psi$  that when added to the agent's *explicit* information makes the agent explicitly informed about  $\chi$ .

Now consider the *not implicit novelty* case (1.2): though the agent does not have  $\chi$  explicitly, she has it *implicitly*. A solution for case (1.1), adding some  $\psi$ , would also work here, but the agent does not really need this external interaction, since a non-omniscient agent has another possibility: she can make the implicit  $\chi$  explicit by performing the adequate *reasoning steps*. And this gives us new possibilities not only this case. For example, in (1.1), the agent does not need a  $\psi$  that makes  $\chi$  explicit after added: a  $\psi$  that makes  $\chi$  *implicit* is also a solution, since now she can make  $\chi$  explicit by only reasoning. In fact, there are several strategies for solving each one of the abductive problems, but for simplicity we will focus on the most representative one for each one of them.

In case (1.3), reasoning will only make the anomaly explicit. But then the agent will be in the *truly anomaly* case (2.3), which can be solved by revising the agent's information to remove  $\neg\chi$  from the explicit and implicit part, and then adding a  $\psi$  that makes  $\chi$  part of her explicit information.

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

**Definition 3.2** (Non-omniscient abductive solutions). Solutions for consistent non-omniscient abductive problems are provided in Table 2.

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

Table 2: Solutions for consistent non-omniscient abductive problems.

Note how actions take us from some abductive problem to another. In case (1.3), the proper reasoning will take the agent to case (2.3) from which, by applying the proper revision, the agent will reach case (1.1), where a new piece of information is needed. The flowchart of Figure 1 shows this.

**Classification of abductive solutions** The extra requisites of Definition 2.3 can be adapted in this non-omniscient case. For the consistency and the explanatory requirements there are no important changes: we just require for the agent's *implicit* (and therefore her explicit information too) to be consistent at the end of the sequence of actions ( $\neg\text{Inf}_{\text{Im}}\perp$ ), and for the formula  $\psi$  to not imply  $\chi$  ( $\neg(\psi \rightarrow \chi)$ ) in the cases in which it is needed

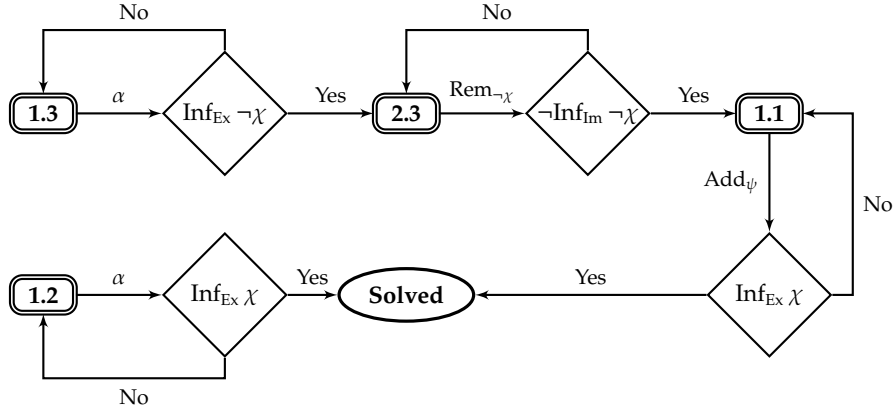


Figure 1: Flowchart of abductive solutions for non-omniscient agents.

((1.1), (1.3) and (2.3)). The minimality requirement now gives us more options. We can define it over the action  $\langle \text{Add}_\psi \rangle$ , looking for the weakest formula  $\psi$ , but it can also be defined over the action  $\langle \text{Rem}_{\neg\chi} \rangle$ , 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.

## 4 A non-dynamically-omniscient agent

Even though the agents of the previous section are non-omniscient, there is still an idealization about them. We have defined the agent's implicit information as what follows logically from her explicit information, but a more 'real' agent does not need to be *dynamically omniscient* in the sense that she does not need to have complete reasoning abilities. In other words, she may not be able to derive all logical consequences of her explicit information. This difference is important, because then a solution for a  $\chi$ -abductive problem does not need to be as strong as a formula that, when added, also informs explicitly the agent about  $\chi$ ; it can also be some formula that, when added, allow the agent to *derive*  $\chi$ .

### 4.1 Abductive problems

Now we can make a further refinement. We can distinguish between what follows logically from the agent's explicit information, the *objective* implicit information  $\text{Inf}_{\text{Im}}$ , and what the agent can actually derive, the *subjective* implicit information  $\text{Inf}_{\text{Der}}$ . In other words,  $\text{Inf}_{\text{Der}} \varphi$  holds when the agent can perform a sequence of reasoning steps that make  $\varphi$  explicit information. In particular, an empty sequence of reasoning steps makes explicit the information that is already explicit, so we assume

$$\text{Inf}_{\text{Ex}} \varphi \rightarrow \text{Inf}_{\text{Der}} \varphi$$



Though not *complete*, we can assume that the agent's reasoning abilities are *sound*. This makes subjective implicit information part of objective implicit one, giving us

$$\text{Inf}_{\text{Der}} \varphi \rightarrow \text{Inf}_{\text{Im}} \varphi$$

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  $\text{Inf}_{\text{Der}} \chi$  and  $\text{Inf}_{\text{Der}} \neg\chi$  hold or not. Our two assumptions allow us to discard some of the cases, leaving us with the following.

**Definition 4.1** (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.

$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \neg\text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \right\} \wedge \neg\text{Inf}_{\text{Im}} \chi \wedge \neg\text{Inf}_{\text{Im}} \neg\chi$	<b>(1.1.a)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \begin{array}{l} \neg\text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \\ \text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \end{array} \right\} \wedge \text{Inf}_{\text{Im}} \chi \wedge \neg\text{Inf}_{\text{Im}} \neg\chi$	<b>(1.2.a)</b> <b>(1.2.b)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \begin{array}{l} \neg\text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \\ \neg\text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi \end{array} \right\} \wedge \neg\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(1.3.a)</b> <b>(1.3.c)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \begin{array}{l} \neg\text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \\ \text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi \\ \neg\text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi \\ \text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi \end{array} \right\} \wedge \text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(1.4.a)</b> <b>(1.4.b)</b> <b>(1.4.c)</b> <b>(1.4.d)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \neg\text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi \right\} \wedge \neg\text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(2.3.c)</b>
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \text{Inf}_{\text{Ex}} \neg\chi \wedge \left\{ \text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi \right\} \wedge \text{Inf}_{\text{Im}} \chi \wedge \text{Inf}_{\text{Im}} \neg\chi$	<b>(2.4.d)</b>

Table 3: Abductive problems with subjective/objective implicit information.

## 4.2 Abductive solutions

Recall that different abductive problems in Table 1 can have the same solution. For example, though abductive problem **(1.2)**, the *non-implicit novelty* case, can be solved by means of reasoning steps (see Table 2), we mentioned that it can also be solved like case **(1.1)**. But the further refinement that we have just done really makes them different. In case **(1.2.b)**,  $\chi$  is subjective implicit information, so the agent can derive it and solve the problem by only reasoning. Nevertheless, this is not possible in **(1.2.a)** since  $\chi$  is objective but not subjective implicit information. The agent cannot derive  $\chi$ ; she needs a formula that, when added to her explicit information, makes  $\chi$  explicit (like in the truly novel case); or, more interesting, she can *extend her reasoning abilities* with a formula/rule that allows her to derive  $\chi$ .

The same happens with other cases; consider those derived from (1.3). In (1.3.c) the anomaly will be detected so the agent can start with a revision of her information. But in (1.3.a) the anomaly cannot be derived, so we have an objective but not subjective anomaly. The agent cannot detect and, moreover, cannot derive the anomaly, so a better approach for a solution is to solve (1.3.a) as a novel abductive problem. In fact, we can say that (1.3.a) is an *objective anomaly* but a *subjective novelty*.

Just like actions of reasoning, revision and addition can take us from one abductive problem of Table 1 to another, they also allows us to move between the abductive problem of Table 3. Again, we will focus on the consistent cases, discarding (1.4.\*) and (2.4.d).

**Definition 4.2** (Extended abductive solutions). Solutions for consistent extended abductive problems are provided in Table 4. It 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. There are six operations/conditions which are described below, from left to right.

- Action  $\langle \text{Add}_{\psi} \rangle$  consists in adding  $\psi$  to the agent's explicit information. The aim is to make the agent explicitly informed about  $\chi$ .
- Action  $\langle \text{Add}_{\psi/\alpha} \rangle$  extends the agent's explicit information by adding a formula  $\psi$  or some inference resource  $\alpha$  (e.g., a rule) with the aim to provide the agent with enough information so she can derive  $\chi$ .
- Action  $\langle \alpha \rangle$  consists on the application of reasoning steps. The goal here is to make  $\chi$  explicit.
- Action  $\langle \text{Add}_{\psi/\alpha} \rangle$  is just as before. The goal is that after the action, the agent should be able to derive  $\neg\chi$ .
- Action  $\langle \alpha \rangle$  is just as before, this time with the aim to make  $\neg\chi$  explicit.
- Action  $\langle \text{Rem}_{\neg\chi} \rangle$  removes  $\neg\chi$  from the agent's explicit information, but the goal here is to remove it from her implicit information as well.

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

Table 4: Solutions for consistent extended abductive problems.

Table 4 establishes a natural path to solve a consistent extended abductive problem in. The longest path corresponds to case **(1.3.a)** in which the agent does not have explicit information about neither  $\chi$  nor  $\neg\chi$  and, though  $\neg\chi$  follows logically from her explicit information, she cannot derive it. A sequence of actions to solve this problem is, first, to provide the agent with enough information so she can derive  $\neg\chi$ , turning this case into **(1.3.c)**. Then, after reasoning to derive  $\neg\chi$  she will have an explicit anomaly, case **(2.3.c)**. From here she needs to revise her information to remove  $\neg\chi$  from it and, once she has done this and reached case **(1.1.a)**, she needs to extend her information with some  $\psi$  that will make her be explicitly informed about  $\chi$ .

### 4.3 Collapsing the cases

We have taken the perspective of an outsider. 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 explicit information and what she can derive from it, that is, the subjective implicit one. In other words, unaccessible inconsistencies should not matter!

We can simplify the solution of abductive problem by observing 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 not relevant. For example, abductive problems **(1.{1,2,3,4}.a)** are in fact the same from the agent's point of view, and she will try to solve them in the same way. By reducing Table 3 according to the agent's subjective information we get:

$\neg\text{Inf}_{\text{Ex}} \chi \wedge \neg\text{Inf}_{\text{Ex}} \neg\chi \wedge$	$\left\{ \begin{array}{ll} \neg\text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi & \mathbf{(1.\{1,2,3,4\}.a)} \\ \text{Inf}_{\text{Der}} \chi \wedge \neg\text{Inf}_{\text{Der}} \neg\chi & \mathbf{(1.\{2, 4\}.b)} \\ \neg\text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi & \mathbf{(1.\{3, 4\}.c)} \\ \text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi & \mathbf{(1.4.d)} \end{array} \right.$
$\neg\text{Inf}_{\text{Ex}} \chi \wedge \text{Inf}_{\text{Ex}} \neg\chi \wedge$	$\left\{ \begin{array}{ll} \neg\text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi & \mathbf{(2.3.c)} \\ \text{Inf}_{\text{Der}} \chi \wedge \text{Inf}_{\text{Der}} \neg\chi & \mathbf{(2.4.d)} \end{array} \right.$

Note how these classes correspond to abductive problems in Table 1 in which  $\text{Inf}_{\text{Der}}$  appears in place of  $\text{Inf}_{\text{Im}}$ . Then the abductive solutions in Table 3 can be considered the *objective solutions* for the eleven objectively different abductive problems. But if we consider only the information accessible to the agent then the *subjective paths* she follows to solve abductive problems are similar to the Figure 1.

## 5 Summary and future work

We have presented definitions of *novel* and *anomalous* abductive problems from a subjective perspective. We have focused not only on omniscient agents, but also on those whose information is not closed under logical consequence and those whose reasoning abilities are not complete. Moreover, we have also provided definitions for what an abductive solution is for each one of these problems, identifying actions that allow the agent to move from one problem to another and the conditions that such actions should verify in order to provide an abductive solution.

Our work is just an initial exploration on abductive reasoning for (non-omniscient) agents, and there are many aspects yet to be studied. To begin with, we can still make a further refinement in our notions of information, this time relative to the explicit one. Among our explicit information, there are things that are supported by the rest of our information. Consider, for example, the quadratic formula for solving quadratic equations: even if we do not know or forget the formula, we can derive it if we know the process of completing squares. But we also have information that is not supported by the rest; things that, if not observed, we would not have. Consider our initial example of Mr. Wilson having a very shiny right cuff: Holmes did not know him before, so there is no way he could have derived this information about his cuff without observing him. And in fact, the intuitive idea of abductive reasoning is closer to situations of this kind in which we try to find support (justification) for facts that, being observed, have become part of our explicit information.

But we can also look for a more concrete form of what we already have. First, we have talked about information, but we can study more specific notions, like *knowledge* and *beliefs*, by asking for more specific requirements, like truth or consistency. Moreover, the real definite step will be given by providing a semantic model in which we can represent not only the introduced notions of information (explicit, subjective implicit, objective implicit) in their knowledge and belief versions, but also provide a concrete definition for the discussed actions that modify them: adding external information, reasoning in order to extend the explicit one or removing part of it. Our current efforts are oriented to *dynamic epistemic* approaches, following not only ideas like public announcements [14; 6] and revision [3; 2], but also the finer grained notions of dropping [4] and inference [7].

And finally we can look at, possibly, the most important question in abductive reasoning. We know now what an abductive solution is but, how can we find them? In other words, given a particular semantic model representing an agent's information, can we provide a procedure that returns 'the set of abductive solutions'? And though an abductive solution can be a formula that, when assumed, provides the agent with explicit information about the observed  $\chi$ , the most interesting solutions are those that allow

the agent to *derive*  $\chi$ , linking tightly the search for solutions with the agent's reasoning tools. This emphasizes the need of a semantic model that allows us to represent an agent's inference dynamics.

Even more: an important topic in abductive reasoning is the selection of the best explanation (e.g., [13] and [10]). Beyond logical requisites to avoid triviality, the definition of a suitable criteria to model the agent's preference for solutions is still an open problem. By using Kripke frames, we can provide some criteria based in a *plausibility* measure among accessible worlds [3; 2]. And at last (but not least), once the agent has selected her 'best' explanation, what shall she do with it?

**Acknowledgements** We thank Johan van Benthem and Ángel Nepomuceno-Fernández for their valuable comments and suggestions.

## References

- [1] A. Aliseda. *Abductive Reasoning. Logical Investigations into Discovery and Explanation*, volume 330 of *Synthese Library Series*. Springer, 2006.
- [2] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. AUP, 2008.
- [3] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.
- [4] J. van Benthem and F. R. Velázquez-Quesada. Inference, promotion, and the dynamics of awareness. Technical Report PP-2009-43, ILLC, Universiteit van Amsterdam, 2009.
- [5] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007.
- [6] J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, ILLC, Universiteit van Amsterdam, 1999. ILLC Dissertation Series DS-1999-01.
- [7] D. Grossi and F. R. Velázquez-Quesada. *Twelve Angry Men: A study on the fine-grain of announcements*. In X. He, J. F. Horty, and E. Pacuit, editors, *LORI*, volume 5834 of *LNCS*, pages 147–160. Springer, 2009.
- [8] J. Y. Halpern, editor. *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge, Monterey, CA, March 1986*, San Francisco, CA, USA, 1986. Morgan Kaufmann Publishers Inc.

- [9] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, N.Y., 1962.
- [10] J. Hintikka. What is abduction? The fundamental problem of contemporary epistemology. *Transactions of the Charles S. Peirce Society*, 34(3):503–533, 1998.
- [11] G. Lakemeyer. Steps towards a first-order logic of explicit and implicit belief. In Halpern [8], pages 325–340.
- [12] H. J. Levesque. A logic of implicit and explicit belief. In *Proc. of AAAI-84*, pages 198–202, Austin, TX, 1984.
- [13] P. Lipton. *Inference to the Best Explanation*. Routledge, New York, 1991.
- [14] J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, Tennessee, USA, 1989. Oak Ridge National Laboratory, ORNL/DSRD-24.
- [15] M. Y. Vardi. On epistemic logic and logical omniscience. In Halpern [8], pages 293–305.