Abduction for (non-omniscient) agents

Fernando Soler-Toscano and Fernando R. Velázquez-Quesada
fsoler@us.es, F.R.VelazquezQuesada@uva.nl

LRBA@MALLOW2010
September, 2010

Looking for explanations (1)

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

How, in the name of good-fortune, did you know all that, Mr. Holmes?

All of us

Looking for explanations (2)

- *Your right hand is quite a size larger than your left. You have worked with it, and the muscles are more developed.*
- *. . . rather against the strict rules of your order, you use an arc-and-compass breastpin . . .*
- *The fish that you have tattooed immediately above your right wrist could only have been done in China.*
- *. . . that right cuff so very shiny for five inches, and the left one with the smooth patch near the elbow where you rest it upon the desk.*

Sherlock Holmes

Abduction

In Peirce's words:

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

Classical examples of abduction:

- Mr. Wilson uses an arc-and-compass breastpin so ***Sherlock*** suspects he is a Freemason.
- Given the symptoms A and B, the ***doctor*** suspects that the patient suffers from C.
- ***Karen*** knows that when it rains, the grass gets wet, and that the grass is wet, therefore, ***she*** suspects that it may have rained.

This work

The classical definition of an abductive problem and an abductive solution only mentions a theory and a formula.

Our goal, a ***epistemic*** and ***dynamic*** approach to abductive reasoning, that is,

- What is an abductive problem ***from an agent's information point of view***?
- What is an abductive solution ***in terms of the actions that modify the agent's information***?
- Do these notions change when we explore ***different kinds of agents***?

We will use formulas in ***dynamic epistemic logic*** style.

Abductive problem

Let Φ be a theory in some language $\mathcal{L}$ and let χ be a formula in $\mathcal{L}$.

Let $\vdash$ be a consequence relation in $\mathcal{L}$.

- The pair (Φ, χ) is a ***novel abductive problem*** if:

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

- The pair (Φ, χ) is an ***anomalous abductive problem*** if:

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

But from an agent's perspective ...

But if we read **Φ** as ***the agent's information***, then

- The agent has a ***novel χ-abductive problem*** whenever

$$\neg \mathbf{Inf}\,\chi \;\wedge\; \neg \mathbf{Inf}\,\neg\chi \tag{1}$$

- The agent has an ***anomalous χ-abductive problem*** whenever

$$\neg \mathbf{Inf}\,\chi \;\wedge\; \mathbf{Inf}\,\neg\chi \tag{2}$$

Abductive solution

Given a ***novel abductive problem*** (Φ, χ),

- the formula ψ is an ***abductive solution*** if

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

Given an ***anomalous abductive problem*** (Φ, χ),

1. perform a ***theory revision*** to get a ***novel*** problem (Φ', χ),
2. then solve (Φ', χ).

Different kinds of abductive solutions

- An abductive solution ψ is ***consistent*** if

$$\Phi, \psi \nvdash \bot$$

- An abductive solution ψ is ***explanatory*** if

$$\psi \nvdash \chi$$

- An abductive solution ψ is ***minimal*** if, for every other abductive solution ϕ,

$$\psi \vdash \phi \quad \text{implies} \quad \phi \vdash \psi$$

But from an agent's perspective ...

For a ***novel*** χ-abductive problem,

- a formula ψ is a solution if

$$\langle \mathbf{Ext}_{\psi} \rangle \, \mathbf{Inf} \, \chi$$

For an ***anomalous*** χ-abductive problem:

- a formula ψ is a solution if

$$\langle \mathbf{Rem}_{\neg\chi} \rangle \, \langle \mathbf{Ext}_{\psi} \rangle \, \mathbf{Inf} \, \chi$$

And the extra conditions

- The formula ψ is a ***consistent*** abductive solution if

$$\langle \mathbf{Ext}_{\psi} \rangle \big(\mathbf{Inf}\, \chi \land \neg \mathbf{Inf}\, \bot \big)$$

- The formula ψ is a ***explanatory*** abductive solution if

$$\neg(\psi \to \chi) \land \langle \mathbf{Ext}_{\psi} \rangle\, \mathbf{Inf}\, \chi$$

- The formula ψ is a ***minimal*** abductive solution if, for every other ϕ,

$$\langle \mathbf{Ext}_{\psi} \rangle\, \mathbf{Inf}\, \chi \land \Big(\big(\langle \mathbf{Ext}_{\phi} \rangle\, \mathbf{Inf}\, \chi \land \langle \mathbf{Ext}_{\psi} \rangle\, \mathbf{Inf}\, \phi \big) \to \langle \mathbf{Ext}_{\phi} \rangle\, \mathbf{Inf}\, \psi \Big)$$

But we have made a strong assumption

A ***theory*** is usually assumed to be closed under logical consequence.

So our agent's information is closed under logical consequence, i.e., we have an ***ominscient*** agent.

What if she is not?

Implicit and explicit information

Then we should make a difference between

- what the agent actually has, her ***explicit information***, and
- what follows logically from it, her ***implicit information***.

A χ-abductive problem appears when χ is not part of the agent's ***explicit*** information.

The new abductive problems

- *Novel* χ-abductive problems:

$$\neg \mathbf{Inf}_{\mathbf{Ex}}\, \chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \;\wedge\; \begin{cases} \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \neg \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (1,1) \\ \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \neg \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (1,2) \\ \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (1,3) \\ \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (1,4) \end{cases}$$

- *Anomalous* χ-abductive problems:

$$\neg \mathbf{Inf}_{\mathbf{Ex}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \;\wedge\; \begin{cases} \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \neg \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (2,1) \\ \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \neg \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (2,2) \\ \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (2,3) \\ \mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi & (2,4) \end{cases}$$

But not all are possible

There is an natural relation between implicit and explicit information:

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

- *Novel* χ-abductive problems:

$$\neg\mathbf{Inf}_{\mathbf{Ex}}\,\chi \;\wedge\; \neg\mathbf{Inf}_{\mathbf{Ex}}\,\neg\chi \;\wedge\; \begin{cases} \neg\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \neg\mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (1,1) \\ \mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \neg\mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (1,2) \\ \neg\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (1,3) \\ \mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (1,4) \end{cases}$$

- *Anomalous* χ-abductive problems:

$$\neg\mathbf{Inf}_{\mathbf{Ex}}\,\chi \;\wedge\; \mathbf{Inf}_{\mathbf{Ex}}\,\neg\chi \;\wedge\; \begin{cases} \text{\sout{$\neg\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \neg\mathbf{Inf}_{\mathbf{Im}}\,\neg\chi$}} & \text{\sout{(2,1)}} \\ \text{\sout{$\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \neg\mathbf{Inf}_{\mathbf{Im}}\,\neg\chi$}} & \text{\sout{(2,2)}} \\ \neg\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (2,3) \\ \mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \mathbf{Inf}_{\mathbf{Im}}\,\neg\chi & (2,4) \end{cases}$$

Now for the solutions . . . (1)

(1.1) The ***truly novel*** case:

$$\neg \mathbf{Inf}_{\mathbf{Ex}}\, \chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathbf{Im}}\, \chi} \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi$$

A solution is a ***formula*** ψ such that

$$\langle \mathbf{Ext}_{\psi} \rangle\, \mathbf{Inf}_{\mathbf{Ex}}\, \chi$$

A solution can also be a ***formula*** ψ and a ***reasoning*** α such that

$$\langle \mathbf{Ext}_{\psi} \rangle \left(\mathbf{Inf}_{\mathbf{Im}}\, \chi \wedge \langle \alpha \rangle\, \mathbf{Inf}_{\mathbf{Ex}}\, \chi \right)$$

Now for the solutions . . . (2)

(2.3) The ***truly anomaly*** case:

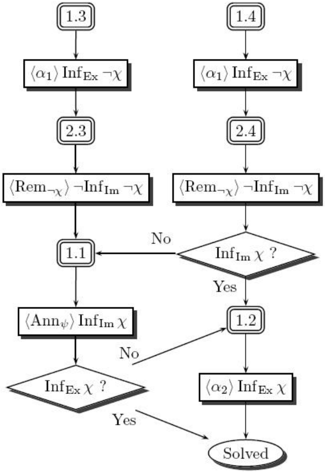
$$\neg \mathbf{Inf}_{\mathrm{Ex}}\, \chi \;\wedge\; \underline{\mathbf{Inf}_{\mathrm{Ex}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathrm{Im}}\, \neg\chi$$

A solution takes two steps: a ***revision*** to remove $\neg\chi$

$$\langle \mathbf{Rem}_{\neg\chi} \rangle \Big(\neg \mathbf{Inf}_{\mathrm{Ex}}\, \chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Ex}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\, \chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Im}}\, \neg\chi} \Big)$$

and then to solve now ***(1.1)***.

In a diagram


1.3
⟨α1⟩ InfEx ¬χ
2.3
⟨Rem¬χ⟩ ¬InfIm ¬χ
1.1
⟨Annψ⟩ InfIm χ
InfEx χ ?
No
Yes
1.4
⟨α1⟩ InfEx ¬χ
2.4
⟨Rem¬χ⟩ ¬InfIm ¬χ
InfIm χ ?
No
Yes
1.2
⟨α2⟩ InfEx χ
Solved

Objective vs subjective information

- 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 have complete reasoning abilities
- She may not be able to derive all logical consequences of her explicit information
- We can distinguish two kinds of implicit information: what follows logically from the agent's explicit information, the ***objective*** implicit information $\mathbf{Inf}_{\mathbf{Im}}\,\varphi$, and what the agent can actually derive, the ***subjective*** implicit information $\mathbf{Inf}_{\mathbf{Der}}\,\varphi$
- $\mathbf{Inf}_{\mathbf{Der}}\,\varphi$ holds when the agent can perform a sequence of reasoning steps $\langle\alpha\rangle$ that make φ explicit information:

$$\mathbf{Inf}_{\mathbf{Der}}\,\varphi \rightarrow \langle\alpha\rangle\,\mathbf{Inf}_{\mathbf{Ex}}\,\varphi$$

- We assume:

$$\mathbf{Inf}_{\mathbf{Ex}}\,\varphi \rightarrow \mathbf{Inf}_{\mathbf{Der}}\,\varphi \rightarrow \mathbf{Inf}_{\mathbf{Im}}\,\varphi$$

Problems with derivable information

Some of the eleven abductive problems:

$$\neg \mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge \neg \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \qquad (1.1.a)$$

$$\neg \mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \neg \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \neg \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \qquad (1.3.a)$$

$$\neg \mathrm{Inf}_{\mathrm{Ex}}\,\chi \wedge \mathrm{Inf}_{\mathrm{Ex}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Im}}\,\chi \wedge \mathrm{Inf}_{\mathrm{Im}}\,\neg\chi \wedge \neg \mathrm{Inf}_{\mathrm{Der}}\,\chi \wedge \mathrm{Inf}_{\mathrm{Der}}\,\neg\chi \qquad (2.3.c)$$

- (1.1.a) Extended truly novel case
- (1.3.a) Subjective novelty with objective anomaly
- (2.3.c) Extended anomaly

Abductive solutions . . . (1)

(1.1.a) The ***extended truly novel*** case:

$$\neg \mathbf{Inf}_{\mathrm{Ex}}\,\chi \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Ex}}\,\neg\chi \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\,\chi \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\,\neg\chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Der}}\,\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Der}}\,\neg\chi$$

A solution is a ***formula*** ψ such that

$$\langle \mathbf{Ext}_{\psi} \rangle\, \mathbf{Inf}_{\mathrm{Ex}}\,\chi$$

A solution can also be a ***formula*** ψ and a ***reasoning*** α such that

$$\langle \mathbf{Ext}_{\psi} \rangle \big(\mathbf{Inf}_{\mathrm{Der}}\,\chi \wedge \langle \alpha \rangle\, \mathbf{Inf}_{\mathrm{Ex}}\,\chi \big)$$

Abductive solutions . . . (2)

(2.3.c) The ***extended anomaly*** case:

$$\neg \mathbf{Inf}_{\mathrm{Ex}}\, \chi \;\wedge\; \underline{\mathbf{Inf}_{\mathrm{Ex}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathrm{Im}}\, \neg\chi \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Der}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathrm{Der}}\, \neg\chi$$

A solution requires first a revision $\langle \mathbf{Rem}_{\neg\chi} \rangle$ such that

$$\langle \mathbf{Rem}_{\neg\chi} \rangle \left(\neg \mathbf{Inf}_{\mathrm{Ex}}\, \chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Ex}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Im}}\, \chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Im}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathrm{Der}}\, \chi \;\wedge\; \underline{\neg \mathbf{Inf}_{\mathrm{Der}}\, \neg\chi} \right)$$

and takes us to case ***(1.1.a)***.

Abductive solutions . . . (3)

(1.3.a) The ***subjective novelty with objective anomaly*** case:

$$\neg \mathbf{Inf}_{\mathbf{Ex}}\, \chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \;\wedge\; \underline{\mathbf{Inf}_{\mathbf{Im}}\, \neg\chi} \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Der}}\, \chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Der}}\, \neg\chi$$

- There is an objective but not subjective anomaly for the agent
- If she does not detect if, she will try to solve ***(1.3.a)*** as ***(1.1.a)***
- But if she gets to extend her information with some ***formula*** ψ or ***rules*** α and a ***reasoning*** α such that

$$\langle \mathbf{Ext}_{\psi/\alpha} \rangle \left(\mathbf{Inf}_{\mathbf{Der}}\, \neg\chi \wedge \langle \alpha \rangle\, \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \right)$$

- She arrives to

$$\neg \mathbf{Inf}_{\mathbf{Ex}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathbf{Ex}}\, \neg\chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Im}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathbf{Im}}\, \neg\chi \;\wedge\; \neg \mathbf{Inf}_{\mathbf{Der}}\, \chi \;\wedge\; \mathbf{Inf}_{\mathbf{Der}}\, \neg\chi$$

that is ***(2.3.c)***

Observed vs entailed explicit information

- The observation χ to explain in an abductive problem is supposed to be surprising.
- But sometimes we face the task of ***explaining explicit information***, for example an observation we have just made.
- In some cases we know ***that*** something is true but we do not know ***why*** is it true.
- We use an operation $\langle \mathbf{Dis}_\varphi \rangle$ of ***discarding*** formula φ. It verifies:

$$\langle \mathbf{Dis}_\varphi \rangle \neg \mathbf{Inf}_{\mathbf{Ex}}\, \varphi$$

- We can distinguish:
  - ***Observed explicit information***, where

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

  - ***Entailed explicit information***, where

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

- Previous abductive problems can be adapted for observed explicit information.

Refining abduction (1): abduction as becoming aware

- Working with non-omniscient agents shows the importance of ***reasoning***.
- Some abductive problems are solved when we become aware of some information that was previously ***implicit***, maybe *on the tip of the tongue*.
- Example:
  - In an Smullyan island of knights (always tell the truth) and knaves (always lie) we find John and Bill. Then John says: *we are both knaves*
  - How to explain this declaration that seems paradoxical?
  - We are implicitly informed about the solution, we only need a reasoning to make it explicit.

Refining abduction (2): abduction as adding rules

- Working with agents with non-complete reasoning abilities highlights the ***importance of rules***.
- Some abductive problems are solved when ***it is proposed a new rule*** that makes *a matter of course* something previously unexplained.
- This is the case of many ***scientific discoveries***.
- Example:
  - The Aristotelian theory of motion establishes that objects move only as long as a force is applied to them.
  - In other case objects move naturally to the center of the Earth following a straight line.
  - It does not explain the parabolic movement of projectiles.
  - Modern mechanics (Galileo & Newton) introduce laws (***new rules***) that allow to explain this paradoxical movement.

Refining abduction (3): abduction as proving the obvious

- The distinction between observed and entailed explicit information allows us to model an interesting kind of abductive problem.
- We may ask if something that is obvious for us can be explained with the rest of our information.
- Example:
  - The ***fifth postulate*** is explicit information in Euclidean geometry.
  - Can it be proved from the first four postulates?
  - This is equivalent to ask whether the fifth postulate is ***observed*** (non-derivable) or ***entailed*** (derivable) explicit information in Euclidean geometry.
  - Non-Euclidean geometries originated when going in depth into this abductive problem.

Knowledge and belief cases

- We can reduce the amount of abductive problems by assuming:
  - ***Truth*** in the agent's information, that is, $\mathbf{Inf}\,\varphi \rightarrow \varphi$. This is the case of ***knowledge***.
  - ***Consistence*** in the agent's information, $\neg\mathbf{Inf}\,\bot$. This corresponds to ***belief***.
- With consistent information there are no abductive problems where $\mathbf{Inf}_{\mathbf{Im}}\,\chi \wedge \mathbf{Inf}_{\mathbf{Im}}\,\neg\chi$
- With true information and observation, there are no possible anomalous abductive problems
- It is interesting to mix notions of knowledge and belief:
  - The background ***theory*** is ***known***
  - The abductive ***solution*** is ***believed***

So far

- Subjective perspective to an abductive problem.
- Dynamic perspective to abductive solution.
- Non-omniscient agents.
- Implicit information: objective vs subjective
- Explicit information: derivable vs no derivable
- Knowledge/belief cases.

Still to do

- Concrete semantic model.
- How to find abductive solutions?
- How to select the “best” of them?
- What the agent will do with this “best” solution?