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 $\chi$ is observed.
2. But if $\varphi$ were true, $\chi$ would be a matter of course.
3. Hence, there is reason to suspect that $\varphi$ 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, an 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 $\Phi$ be a theory in some language $\mathcal{L}$ and let $\chi$ be a formula in $\mathcal{L}$. Let $\vdash$ be a consequence relation in $\mathcal{L}$.

- The pair $(\Phi, \chi)$ is a **novel abductive problem** if:
  $$\Phi \not\vdash \chi \quad \text{and} \quad \Phi \not\vdash \neg \chi$$

- The pair $(\Phi, \chi)$ is an **anomalous abductive problem** if:
  $$\Phi \not\vdash \chi \quad \text{and} \quad \Phi \vdash \neg \chi$$
But from an agent’s perspective . . .

But if we read $\Phi$ as *the agent’s information*, then

- The agent has a *novel $\chi$-abductive problem* whenever
  \[ \neg \text{Inf } \chi \land \neg \text{Inf } \neg \chi \]  
  (1)

- The agent has an *anomalous $\chi$-abductive problem* whenever
  \[ \neg \text{Inf } \chi \land \text{Inf } \neg \chi \]  
  (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 \((\Phi, \chi)\),

1. perform a theory revision to get a novel problem \((\Phi', \chi)\),
2. then solve \((\Phi', \chi)\).
Different kinds of abductive solutions

- An abductive solution $\psi$ is **consistent** if
  \[ \Phi, \psi \not\vdash \bot \]

- An abductive solution $\psi$ is **explanatory** if
  \[ \psi \not\vdash \chi \]

- An abductive solution $\psi$ is **minimal** if, for every other abductive solution $\phi$,
  \[ \psi \vdash \phi \quad \text{implies} \quad \phi \vdash \psi \]
But from an agent’s perspective . . .

For a *novel* $\chi$-abductive problem,
- a formula $\psi$ is a solution if

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

For an *anomalous* $\chi$-abductive problem:
- a formula $\psi$ is a solution if

$$\langle \text{Rem}_{\neg \chi} \rangle \langle \text{Ext}_\psi \rangle \text{Inf} \chi$$
And the extra conditions

- The formula $\psi$ is a *consistent* abductive solution if

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

- The formula $\psi$ is a *explanatory* abductive solution if

$$\neg (\psi \rightarrow \chi) \land \langle \text{Ext}_\psi \rangle \text{Inf } \chi$$

- The formula $\psi$ is a *minimal* abductive solution if, for every other $\phi$,

$$\langle \text{Ext}_\psi \rangle \text{Inf } \chi \land \left( (\langle \text{Ext}_\phi \rangle \text{Inf } \chi \land \langle \text{Ext}_\psi \rangle \text{Inf } \phi) \rightarrow \langle \text{Ext}_\phi \rangle \text{Inf } \psi \right)$$
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 *omniscient* agent.

What if she is not?
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 $\chi$-abductive problem appears when $\chi$ is not part of the agent’s *explicit* information.
THE NEW ABDUCTIVE PROBLEMS

- **Novel** $\chi$-abductive problems:

  $\neg\text{Inf}_E \chi \land \neg\text{Inf}_E \neg\chi$ \land \begin{cases} 
  \neg\text{Inf}_I \chi \land \neg\text{Inf}_I \neg\chi & (1,1) \\ 
  \text{Inf}_I \chi \land \neg\text{Inf}_I \neg\chi & (1,2) \\ 
  \neg\text{Inf}_I \chi \land \text{Inf}_I \neg\chi & (1,3) \\ 
  \text{Inf}_I \chi \land \text{Inf}_I \neg\chi & (1,4) 
\end{cases}

- **Anomalous** $\chi$-abductive problems:

  $\neg\text{Inf}_E \chi \land \text{Inf}_E \neg\chi$ \land \begin{cases} 
  \neg\text{Inf}_I \chi \land \neg\text{Inf}_I \neg\chi & (2,1) \\ 
  \text{Inf}_I \chi \land \neg\text{Inf}_I \neg\chi & (2,2) \\ 
  \neg\text{Inf}_I \chi \land \text{Inf}_I \neg\chi & (2,3) \\ 
  \text{Inf}_I \chi \land \text{Inf}_I \neg\chi & (2,4) 
\end{cases}
There is a natural relation between implicit and explicit information:

$$\text{Inf}_E \phi \rightarrow \text{Inf}_I \phi$$

- **Novel** $\chi$-abductive problems:
  $$\neg \text{Inf}_E \chi \wedge \neg \text{Inf}_E \neg \chi \wedge$$
  $$\begin{cases}
  \neg \text{Inf}_I \chi \wedge \neg \text{Inf}_I \neg \chi & (1,1) \\
  \text{Inf}_I \chi \wedge \neg \text{Inf}_I \neg \chi & (1,2) \\
  \neg \text{Inf}_I \chi \wedge \text{Inf}_I \neg \chi & (1,3) \\
  \text{Inf}_I \chi \wedge \text{Inf}_I \neg \chi & (1,4)
  \end{cases}$$

- **Anomalous** $\chi$-abductive problems:
  $$\neg \text{Inf}_E \chi \wedge \text{Inf}_E \neg \chi \wedge$$
  $$\begin{cases}
  \neg \text{Inf}_I \chi \wedge \neg \text{Inf}_I \neg \chi & (2,1) \\
  \text{Inf}_I \chi \wedge \neg \text{Inf}_I \neg \chi & (2,2) \\
  \neg \text{Inf}_I \chi \wedge \text{Inf}_I \neg \chi & (2,3) \\
  \text{Inf}_I \chi \wedge \text{Inf}_I \neg \chi & (2,4)
  \end{cases}$$
Now for the solutions . . . (1)

(1.1) The truly novel case:

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \neg \text{Inf}_{\text{Im}} \neg \chi \]

A solution is a formula \( \psi \) such that

\[ \langle \text{Ext}_{\psi} \rangle \text{Inf}_{\text{Ex}} \chi \]

A solution can also be a formula \( \psi \) and a reasoning \( \alpha \) such that

\[ \langle \text{Ext}_{\psi} \rangle \left( \text{Inf}_{\text{Im}} \chi \land \langle \alpha \rangle \text{Inf}_{\text{Ex}} \chi \right) \]
Now for the solutions . . . (2)

(2.3) The truly anomaly case:

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg \chi \]

A solution takes two steps: a revision to remove \( \neg \chi \)

\[ \langle \text{Rem}_{\neg \chi} \rangle \left( \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \neg \text{Inf}_{\text{Im}} \neg \chi \right) \]

and then to solve now (1.1).
In a diagram
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 $\text{Inf}_{\text{Im}} \varphi$, and what the agent can actually derive, the *subjective* implicit information $\text{Inf}_{\text{Der}} \varphi$.
- $\text{Inf}_{\text{Der}} \varphi$ holds when the agent can perform a sequence of reasoning steps $\langle \alpha \rangle$ that make $\varphi$ explicit information:
  \[
  \text{Inf}_{\text{Der}} \varphi \rightarrow \langle \alpha \rangle \text{Inf}_{\text{Ex}} \varphi
  \]
- We assume:
  \[
  \text{Inf}_{\text{Ex}} \varphi \rightarrow \text{Inf}_{\text{Der}} \varphi \rightarrow \text{Inf}_{\text{Im}} \varphi
  \]
Some of the eleven abductive problems:

\[-\text{Inf}_{\text{Ex}} \chi \land \neg\text{Inf}_{\text{Ex}} \neg\chi \land \neg\text{Inf}_{\text{Im}} \chi \land \neg\text{Inf}_{\text{Im}} \neg\chi \land \neg\text{Inf}_{\text{Der}} \chi \land \neg\text{Inf}_{\text{Der}} \neg\chi\]  

(1.1.a)

\[-\text{Inf}_{\text{Ex}} \chi \land \neg\text{Inf}_{\text{Ex}} \neg\chi \land \neg\text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg\chi \land \neg\text{Inf}_{\text{Der}} \chi \land \neg\text{Inf}_{\text{Der}} \neg\chi\]  

(1.3.a)

\[-\text{Inf}_{\text{Ex}} \chi \land \text{Inf}_{\text{Ex}} \neg\chi \land \neg\text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg\chi \land \neg\text{Inf}_{\text{Der}} \chi \land \text{Inf}_{\text{Der}} \neg\chi\]  

(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:

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \neg \text{Inf}_{\text{Im}} \neg \chi \land \neg \text{Inf}_{\text{Der}} \chi \land \neg \text{Inf}_{\text{Der}} \neg \chi \]

A solution is a *formula* \( \psi \) such that

\[ \langle \text{Ext}_\psi \rangle \text{Inf}_{\text{Ex}} \chi \]

A solution can also be a *formula* \( \psi \) and a *reasoning* \( \alpha \) such that

\[ \langle \text{Ext}_\psi \rangle \left( \text{Inf}_{\text{Der}} \chi \land \langle \alpha \rangle \text{Inf}_{\text{Ex}} \chi \right) \]
Abductive solutions . . . (2)

(2.3.c) The extended anomaly case:

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg \chi \land \neg \text{Inf}_{\text{Der}} \chi \land \text{Inf}_{\text{Der}} \neg \chi \]

A solution requires first a revision \( \langle \text{Rem}_{\neg \chi} \rangle \) such that

\[
\langle \text{Rem}_{\neg \chi} \rangle \left( \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \neg \text{Inf}_{\text{Im}} \neg \chi \land \neg \text{Inf}_{\text{Der}} \chi \land \neg \text{Inf}_{\text{Der}} \neg \chi \right)
\]

and takes us to case (1.1.a).
Abductive solutions ...(3)

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

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Der}} \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg \chi \land \neg \text{Inf}_{\text{Ex}} \chi \land \neg \text{Inf}_{\text{Der}} \chi \]

- 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 \( \psi \) or rules \( \alpha \) and a reasoning \( \alpha \) such that
  \[ \langle \text{Ext}_{\psi/\alpha} \rangle (\text{Inf}_{\text{Der}} \neg \chi \land \langle \alpha \rangle \text{Inf}_{\text{Ex}} \neg \chi) \]
- She arrives to

\[ \neg \text{Inf}_{\text{Ex}} \chi \land \text{Inf}_{\text{Ex}} \neg \chi \land \neg \text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg \chi \land \neg \text{Inf}_{\text{Der}} \chi \land \text{Inf}_{\text{Der}} \neg \chi \]

that is (2.3.c)
The observation $\chi$ 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 $\langle \text{Dis}_\varphi \rangle$ of *discarding* formula $\varphi$. It verifies:

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

We can distinguish:

- **Observed explicit information**, where
  $$\text{Inf}_{\text{Ex}} \varphi \land \langle \text{Dis}_\varphi \rangle \neg \text{Inf}_{\text{Der}} \varphi$$

- **Entailed explicit information**, where
  $$\text{Inf}_{\text{Ex}} \varphi \land \langle \text{Dis}_\varphi \rangle \text{Inf}_{\text{Der}} \varphi$$

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.
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, $\text{Inf } \varphi \rightarrow \varphi$. This is the case of *knowledge*.
  - *Consistency* in the agent’s information, $\neg\text{Inf } \bot$. This corresponds to *belief*.
- With consistent information there are no abductive problems where $\text{Inf}_{\text{Im}} \chi \land \text{Inf}_{\text{Im}} \neg\chi$
- 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*
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?