The Babelism about Induction and Abduction

Brigitte Bessant[1]

Abstract. The concepts of inductive and abductive reasoning lead to many open discussions. Some of them arise from confusions in the use of the terms 'induction' and 'abduction' that are due to the existence of numerous conflicting definitions. While both induction and abduction aim to derive hypotheses, such hypotheses are different in the general case. In this position paper, we show how it is difficult to reach a consensus on the relations between induction and abduction. Our contribution consists in improving the characterization of induction and abduction in a particular class.

1 INTRODUCTION

The characterization of inductive and abductive reasonings led and leads to numerous debates. Many works are based on these reasonings, for example in machine learning (e.g. [7, 11]), inductive logic programming (see [14] for a survey), abductive logic programming (e.g. [16]), or resolution of diagnosis problems (see [4, 17, 3] for various approaches to abduction). However many of these works just deal with one of these reasonings. We situate our paper within a very recent tendency consisting of the analysis and characterization of *relations* between induction and abduction.

Many definitions are proposed for inductive and abductive reasonings. However there is no consensus on what is induction and abduction. Actually, many people have an intuitive feeling about their nature. Definitions and modelisations are proposed using various unrelated formalisms, reflecting personal points of view about abduction and induction. How can we thus relate induction and abduction when people base their analysis on an intuitive perception of these concepts, and sometimes using different languages? We call this problem, *the babelism about induction and abduction* one.

In many works on these two models of reasoning, three kinds of knowledge can be distinguished :

- the knowledge base, namely domain **Th**(eory)
- the knowledge that triggers these reasonings, namely **O**(bservation)
- the resulting knowledge, namely **H**(ypothesis)

Commonly, we associate induction with the generalization of O. For example, from the observation of some black crows, we induce that (we generalize by) all crows are black. Abduction is associated with the inference to the best explanation of O. For example, from the observation that Jack is black, we abduce that (we explain it by) Jack is a crow, knowing that all crows are black.

[1] C.R.I.L., Université d'Artois, Rue de l'université SP16, 62307 Lens Cedex, France, e-mail : *bessant@lens.lifl.fr*.

From these intuitive definitions, it appears that, seemingly, abduction and induction both rely on more primitive notions, like generality and explanation. Unfortunately, there are no general definitions for such notions. Basically, this is one of the main causes to the babelism problem evoked above. However, our opinion is that it is not the unique one. Particularly, there can be hypotheses that are at the same time explanative and general, explanative hypotheses that take the form of facts, general ones that are non explanative, etc.

The paper is organized as follows. In the second section, we give our position on the relations between induction and abduction of AI class (see definition in section 2.2). In the third section, we propose a general discussion on the topic and we conlude.

2 A FORMAL ACCOUNT

We embrace a logical point of view on the relations between induction and abduction, restricted to what we call, the **AI explanation** problem defined as follows:

Given an observation O and a domain theory Th

find a hypothesis H such as

$$Th \wedge H \models O$$

We call such a hypothesis, an AI explanation.

2.1 Motivation

In many works in Artificial Intelligence, based on inductive and abductive reasonings, we refer, from a logical point of view, to the AI explanation problem. It appears that in the course of some informal discussions on the relations between these two models of reasoning, it becomes quickly impossible to reach a consensus when this reference is too simplified.

Generally, a first step consists in associating induction or abduction problem with the AI explanation one (i.e. as a necessary and sufficient condition).

However, it is *not a sufficient condition*. For example, given $Th \equiv shine(sun)$ and $O \equiv fly(Tweety)$, the hypothesis $H \equiv mouse(Clyde) \wedge fly(Tweety)$ is an AI explanation but is not commonly accepted as inductive or abductive.

Moreover, it can be a necessary condition for abduction, but it is not necessary for induction. For example, given $Th \equiv shine(sun)$ and $O \equiv fly(Tweety) \wedge bird(Tweety)$, the hypothesis $H \equiv \forall x(bird(x) \rightarrow fly(x))$ is an inductive one, because H is a **hypothetico-deductive** explanation of O ($bird(Tweety) \wedge H \models O$), but it is not an AI explanation ($Th \wedge H \not\models O$).

At the second step, we could add more information about H, like: H is general (e.g. H is a universal quantified formula), for induction, and H is a fact (e.g. H is a ground formula),

for abduction. But this does not solve the problems evoked above.
Such considerations correspond to implicit assumptions that we link to the epistemological point of view of Hempel (in [6]) described as:
given a phenomenon to be explained, namely explanandum, a **deductive-nomological** explanation of this phenomenon is the knowledge called explanans such as

$$explanans \models explanandum$$

Explanans is characterized by Hempel as a conjunction of a set of covering laws represented by L and particular conditions represented by C

$$explanans \equiv L \wedge C$$

In the following section, we propose to refine the model of AI explanation by a logical characterization of some minimal properties for inductive and abductive hypotheses.

2.2 Formal definitions

Logical setting
We consider a first-order language $\mathcal{L}$ constituted by an infinite countable set of classical symbols (variables, connectors $\vee, \wedge, \neg, \rightarrow$, universal and existential quantifiers $\forall$, $\exists$) and by an infinite countable set of non-logical symbols (constants, functions, relations). Classical rules of formation of terms and formulas, are considered to be known. Let us simply note that capital roman letters $A, B, C \ldots$ denote well-formed formulas of $\mathcal{L}$. $M(A)$ is the set of models of A (interpretations where A is true). $\overline{A}$ is the deductive closure of A (i.e. $\overline{A} = \{B$ formulas of $\mathcal{L}$ such that $A \models B\}$).

Définition 1 *Induction and abduction are of* **AI class** *iff they are based*
- on the logical model of AI explanation
- on the epistemological model of deductive-nomological explanation
Induction of AI class is defined as the inference of covering laws and abduction of AI class as the inference of particular conditions.

We first give the logical characterization of knowledge common to induction and abduction.

- Th is a **domain theory** (or knowledge base) iff
  1. Th is a closed formula of $\mathcal{L}$
  2. $M(Th) \neq \emptyset$
- O is an **observation** iff
  1. O is a closed formula of $\mathcal{L}$
  2. $M(O) \neq \emptyset$
  3. If Th is a domain theory, then $Th \not\models O$ and $M(Th \wedge O) \neq \emptyset$
- H is a **hypothesis** iff
  1. H is a closed formula of $\mathcal{L}$
  2. Given Th and O, $Th \wedge O \not\models H$ and $M(Th \wedge O \wedge H) \neq \emptyset$

We consider that the domain theory Th and the observation O are consistent ($M(Th) \neq \emptyset$ and $M(O) \neq \emptyset$) and that, O and H, have the particularity to be consistent with the, a priori knowledge ($M(Th \wedge O) \neq \emptyset$ and $M(Th \wedge O \wedge H) \neq \emptyset$). Indeed, the reason is that there exists at least, one common model to Th, O and H that is the intentional interpretation (or real world).
The validity of a hypothesis H is not known ($Th \wedge O \not\models H$). At last, the observation O has a particularity linked to its part in the trigger of induction and abduction that is that, a priori, we do not have a proof of its validity ($Th \not\models O$).

- L is a **law** iff it is semantically[2] equivalent to a universally quantified formula.
- C is a **particular condition** iff it is a ground formula

Définition 2 (Induction of AI class) *Given an observation O and a domain theory Th, the minimal properties for a hypothesis H induced from O and Th are:*
(I1) *H is a hypothesis*
(I2) *H is a law*
(I3)
1 - Either $H \models O$
2 - Or H is such that there exists a particular condition C such that $H \not\models O$, $C \not\models O$, $C \wedge H \models O$, and

$$\textit{either (a) } C \in \overline{Th} \quad \textit{or} \quad \textit{(b) } C \in \overline{O}$$

Définition 3 (Abduction of AI class) *Given an observation O and a domain theory Th, the minimal properties for a hypothesis H abduced from O and Th are:*
(A1) *H is a hypothesis*
(A2) *H is a particular condition*
(A3) *H is such that there exists a law $L \in \overline{Th}$ such that $H \not\models O$, $L \not\models O$, $L \wedge H \models O$*

(I1)(I2)(I3) (resp. (A1)(A2)(A3)) determine the set of possible inductive (resp. abductive) hypotheses. We do not address here, the preference criteria that permits to select interesting hypotheses.

2.3 Some examples

Given a domain theory $Th \equiv shine(sun)$, let an observation $O \equiv fly(Tweety) \wedge bird(Tweety)$.

$F_1 \equiv mouse(Clyde) \wedge fly(Tweety)$ is neither an inductive hypothesis because of (I2), nor an abductive one because of (A3): Th does not contain a law.
$F_2 \equiv \forall x(fly(x) \wedge bird(x))$ is an inductive hypothesis: (I1) (I2) (I3-1) are verified.
$F_3 \equiv \forall x(fly(x) \rightarrow bird(x))$ (resp. $F_4 \equiv \forall x(bird(x) \rightarrow fly(x))$) is an inductive hypothesis: (I1) (I2) (I3-2-b) are verified with the particular condition C (such that $C \in \overline{O}$) equivalent to $fly(Tweety)$ (resp. $bird(Tweety)$).
$F_5 \equiv \forall x(crow(x) \rightarrow (fly(x) \wedge bird(x))$ is not an inductive hypothesis: (I3) is not verified. However, if we add $crow(Tweety)$ to Th, H_5 becomes inductive: (I3-2-a) is verified.

[2] A is semantically equivalent to B iff $M(A) = M(B)$.

Th does not contain law, so the set of possible abductive hypotheses is empty because of (A3).
If we add $F_6 \equiv \forall x(gold(x) \rightarrow shine(x))$ to the theory, the only ground formulas that we can consider such that $Th \wedge F \models O$, are $F \equiv O \wedge G$ with any formula G. However, it is not an abductive hypothesis ($F \models O$).
If we add F_5 to Th then $crow(Tweety)$ is an abductive hypothesis: (A1)(A2)(A3) are verified.

From an epistemological point of view, we adopt the Hempel's one (with the notion of law and particular condition) rather than Pierce's one [15]. The reason is that it seems that in an important class of AI works, the intuitive definitions are closer to the deductive-nomological explanation. From a logical point of view, (I2) and (A2) are syntactical considerations that could be too restrictive. However, we think that this captures a non negligible part of AI works. We are working on a semantical characterization of the notions of law and particular condition, taken in a more general sense. Our main purpose was to show that from a logical point of view, a too simplified reference to the AI explanation problem can lead to debates directly linked to the babelism problem, when we attempt to characterize some relations between induction and abduction. In our definition, these reasonings are forms of AI explanation except for induction (I1)(I2)(I3-2-b). Abduction is domain-dependent. Indeed, it is the inference of particular conditions that, by epistemological definition, depends on a law. If this law were not in the domain theory, we should infer it as an hypothesis too, and it would not be an abductive reasoning but an explanative one. Induction and abduction of AI class are two different reasonings because of (I2) and (A2), but related by the AI explanatory reasoning.

3 DISCUSSION AND CONCLUSION

Is induction a form of abduction? Is abduction a form of induction? or are they different and how are they related? Following Michalski's definitions [12] abduction is a form of induction (with the inductive inference is defined as "hypothezing premise P, given BK and C such as $P \wedge BK \models C$"). However, if we consider Pierce's formulation [15] of abduction: "The surprising fact, C, is observed, but if A were true, C would be a matter of course, hence there is reason to suspect that A is true", induction as defined by R.S. Michalski is a form of abduction. Another approach to abduction is to consider the domain theory Th as a set of beliefs (not as certain knowledge) and then, to define it as the inference of H such as: from Th revised by H, we can deduce O. Following Michalski's definition, this form of abduction is neither induction nor abduction. The difficulty to relate induction and abduction, in a general way, is that various definitions are proposed for characterizing some basic notions, common to induction and abduction.
For example, let us consider that of explanation. We think that sometimes confusion arises because "explanation = explanation + explanation". Indeed, the knowledge H explains the knowledge O because of a particular relation between the two. Sometimes we consider that the explanation of O is H, sometimes it is this particular relation and sometimes it is H and the relation. We distinguish three levels of explanation: the knowledge H that leads to O (e1), the relation that links H and O (e2), and the whole (e3 = e1 + e2). When we relate induction and abduction, it is important that the implicit level of explanation chosen in the description of induction be the same than the implicit level chosen for the abduction one. According to the level (e1), induction and abduction are the same reasoning: an explanatory non truth-preserving one (that is simply, a hypothetical reasoning), while according to the level (e2) or (e3), they are different.
One problem is when we consider the informal definition, commonly accepted, of abduction, that is to say inference to the best explanation; induction is a form of abduction, whose the criteria chosen for the best explanation is "the more general one". The implicit level of explanation is (e3). However, sometimes induction is not a form of abduction if we consider the level (e1) (for example, when we distinguish a particular form for the hypothesis inferred), or if we consider the level (e2) (for example, when we distinguish the process of construction of the hypothesis).
We think that a general consideration on the relations between induction and abduction should lay on, at least, two notions: the explanation and the confirmation. The explanatory power of a hypothesis H translates the relevance of H in relation to the observation O, and the confirmatory power of O translates the relevance of O in relation to H. This duality is found for example, in P. Flach's work [5] consisting of a meta-level characterization of induction by rationality postulates. The duality may also be an argument, to separate induction from abduction. For example, the relevance of O in relation to H depends on other observations in induction, that are not of the same nature than in abduction. The confirmation problem is studied for induction (see [13]), although it is rarely expressed for abduction.
In short, we emphasized the babelism problem and we discuss on some relevant parameters to the distinction between induction and abduction. We proposed an improvement of the characterization of induction and abduction in AI class. We are actually working on a more general approach, in a semantical setting.

ACKNOWLEDGEMENTS

This work has been supported by the Ganymède II project of the Contrat de Plan Etat/Nord-Pas-de-Calais and by the MESR (Ministère de l'Enseignement Supérieur et de la Recherche). We would like to thank E. Grégoire and P. Marquis for helpful discussions on the topic.

REFERENCES

[1] P. Besnard, E. Grégoire, 1992. *About Deductive Generalization*, in D.Pearce, G.Wagner (Eds), Logics in AI, LNCS 633, pp212-231
[2] L. Console, D. Theseider Dupre, P. Torasso, 1991. *On the Relationship between Abduction and Deduction* , in Journal of Logic Computational, 1-5, pp661-690
[3] P.T. Cox, T. Pietrzykowski, 1987. *General diagnosis by abductive inference*, in Proc. IEEE Symposium on Logic Programming, San Francisco, pp183-189
[4] J. de Kleer, C.R. Williams, 1987. *Diagnosing multiple faults*, Artificial Intelligence, vol32, pp97-130
[5] P.A. Flach, 1996. *Rationality Postulates for Induction*, in proc. TARK'96, pp267-281
[6] C.G. Hempel, 1972. *Elements d'épistémologie*, Armand Colin

eds. Translation of the book *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs, N.J., 1966

[7] Y.Kodratoff, J.G.Ganascia, 1986. *Improving the Generalization Step in Learning*, Machine Learning II, R.S.Michalski & al. (eds), Morgan Kaufmann, Paolo Alto, pp215-244

[8] P. Lipton, 1991. *Inference to the best explanation*, London and New-York

[9] P. Marquis, 1991. *Contribution à l'étude des méthodes de construction d'hypothèses en intelligence artificielle*, Doctorat thesis, university of Nancy I

[10] M.C. Mayer, F. Pirri, 1996. *Abduction is not deduction-in-reverse*, dans journal d'IGPL, vol. 4, no 1, pp1-14

[11] R.S. Michalski, 1983. *A theory and methodology of inductive learning* in Machine Learning: an AI approach, T.M. Mitchell et al., San Mateo, CA

[12] R. Michalski, 1993. *Inferential Theory of Learning as a Conceptual Basis for Multistrategy Learning*, in Machine Learning 11, pp11-151

[13] H. Mortimer, 1988. *The logic of induction* , Ellis Horwood Limited

[14] S. Muggleton, L. De Raedt, 1994. *Inductive Logic Programming : theory and methods*, in Jounal of logic programming, 19-20, pp 629-679

[15] C.S. Pierce, 1931. *Collected papers of Charles Sanders Pierce* vol2, C.Harsthorne & P.Weiss eds, Harvard University Press, Cambridge

[16] D. Poole, R. Goebel, R. Aleluinas, 1987. *Theorist: a logical reasoning system for default and diagnosis* in N.Cercone, G.McCalla eds., The knowledge Frontier, pp331-352, Springer-Verlag

[17] D. Poole, 1989. *Normality and faults* in logic-based diagnosis in proc. IJCAI'89, pp1304-1310