The two faces of abduction: bridging and separating philosophical and formal models

Alexandros Apostolidis

National and Kapodistrian University of Athens

Department of History and Philosophy of Science

Keywords: Abduction, Inference to the Best Explanation, AKM models, Formal Abduction.

1. Abstract

Abduction is a mode of inference that is pivotal in both philosophy and computer science. Contemporary consensus holds that what is commonly referred to as "philosophical abduction" encompasses two complementary methods: the heuristic and the selective. The latter is typically identified with Inference to the Best Explanation [IBE].

In recent decades, there has been an effort to formalize IBE. Notably, in the 1980s, researchers attempted to frame IBE as a specific instance of *the set covering problem*. Over the following decades, the development of the AKM framework has been developed. This framework defines the abductive problem and outlines some of the conditions that should be met for an algorithm to be considered as having successfully solved it. More recently, numerous formalizations based on the AKM framework have been proposed. These use a wide variety of logical languages in order to find either minimal or explanatory solutions.

Today, abduction is generally understood to comprise two distinct branches: philosophical abduction and formal abduction. In this presentation, their commonalities as well as their differences will be explored. This will be done using the concepts of internal and external equivalence.

Regarding internal equivalence, it will be shown that although IBE and formal abduction share certain structural features, they are not fully equivalent. There are explanatory virtues that are considered critical in IBE but have not yet been formalized in models based on the AKM framework. Similarly, it will be argued that even if IBE is treated as a black box, it is not externally equivalent to its formal counterpart, since there are several inputs for which the two methods produce divergent outputs.

Hence, although the two methods may successfully solve a subset of genuine abductive problems in similar ways, at the same time, their methodologies and the solutions they produce for other problems radically differ.

2. Philosophical Abduction

Abduction in philosophy is an inference method, introduced by Peirce. Peirce describes the general form of abduction as follows:

The surprising fact C has been observed.

But if A were true, C would be a matter of course.

Hence, there is reason to suspect that A is true.

Magnani (2001) splits abduction into two classes: creative abductions and selective abductions. The first ones refer to abductive inferences where the goal is to find an explanation. During this investigation, we may come up with new laws, new theories or new theoretical concepts in general. The second ones refer to situations where we must choose the best explanation from a given set of alternative hypotheses.

Focusing on explanatory abductions or IBEs, the last thing we should examine is the selective procedure that discriminates the best hypothesis. Kuhn (1977) introduced a list of five explanatory virtues, namely accuracy, consistency, scope, simplicity and fruitfulness. But this list is far from being complete. Hence, since 1977 many philosophers have tried to answer this question. In the meta-list of the lists of explanatory virtues several virtues like consilience, internal coherence, unifying power, completeness, importance, parsimony, and unification are mentioned.

We are far from being able to speak of IBE as a single method. IBE is not a uniquely defined procedure, in the sense that in each application, alternative explanations are not evaluated on the basis of the same virtues, each enjoying equal epistemic weight.

Thus, philosophical abduction is a method of inference consisting of a creative and a selective component, which can be equated with IBE. In its most common form, we use IBE to select the best explanation, whose goodness is evaluated based on explanatory virtues.

3. Formal Abduction

The second generation of formal models of abduction begins in 1997 with the work of Aliseda, Kuipers, Kowalski, Magnani, Meheus etc who founded the AKM frame. Nowadays, the *abductive problem* is defined as the ordered pair $<\Theta, \phi>$, where Θ stands for the set of background knowledge and ϕ is the fact to be explained. According to Aliseda (2006) the case of *abductive novelty* arises when ϕ is novel, that is neither ϕ nor its negation can be explained by Θ alone, that is both $\Theta \nvDash \phi$ and $\Theta \nvDash \neg\phi$ hold. The case of *abductive anomaly* comes up when ϕ cannot be explained by Θ alone but at the same time Θ explains the negation of ϕ, that is $\Theta \nvDash \phi$ and $\Theta \vDash \neg\phi$ hold. If we restrict the notion of explanation to the notion of tautological implication, then we may end up with five versions of abductive explanations. The most notable among them are *explanatory* and *minimal* abduction. Based on Aliseda's work, hypothesis α is explanatory if it meets:

[RComp]	$\Theta, \alpha \vDash \phi$	Rule of Completeness
[RCons]	$\Theta, \{\alpha\}$ is consistent	Rule of Consistency
[RnT]	$\alpha \nvDash \phi$	Rule of non-Triviality

Moreover, it is called *minimal* if, in addition, it meets:

[RM]	No other logically weaker explanation exists	Rule of Minimality

These characterizations of the abductive problem and the abductive solutions provide the general frame, from which several AKM models emerge. Almost all these models formalize either explanatory or minimal abduction in different ways. Each of them employs a different form of logic as well as a distinct proof methodology. For example,

Aliseda (2006) formalizes both explanatory and minimal abduction in sentential logic with the use of semantic tableaux. On the other hand, Gauderis (2013) formalizes minimal abduction in a first order modal frame via adaptive logic.

These models try to formalize the selective abductions and not the creative ones. Even if it seems that they start in an explanatory vacuum and try to find one explanation, they do not introduce new concepts nor do they formulate new laws or theories. The AKM models select or form an explanation, while remaining within the already established vocabulary. Hence, they lack the creative part of the abductive syllogisms.

4. How can we compare two methods?

The question that arises is whether the two methods, namely philosophical abduction and formal abduction, are equivalent. It's easy to see that they aren't equivalent as formal models lack the creative part. The rather interesting question that emerges is whether the restriction of philosophical abduction to its selective part is equivalent to formal abduction. That is, is IBE equivalent to any of the formalizations implemented by the AKM models?

The initial idea is to compare the two methods, namely IBE and formal abduction, as if they were two programs or algorithms. But an equivalence relation between algorithms cannot be defined. Moreover, neither of the two modes of inference is purely algorithmic. The good news is that, although we cannot directly compare the two modes of inference by treating them as algorithms, we can borrow two notions from program comparison. More specifically, in what follows we will examine whether the two methods are equivalent either internally or externally.

5. Internal Equivalence

Expanding the notions of intention and extension will help us in the comparison of IBE and formal abduction. The idea behind internal equivalence is that of translatability between the two methods. If there exists a both way translation between the properties or the qualities consisting IBE on the one hand and formal abduction on the other, then the two methods are equivalent.

The most fundamental properties that IBE and formal abduction possess are the values that they use for evaluating the alternative explanations. IBE uses some (or maybe all) of the explanatory virtues mentioned above, while formal abductions use 3 or 4 among the rules of consistency, completeness, non-triviality and perhaps minimality. Recall that the rule of completeness is fulfilled if the explanation α alongside the set of background knowledge Θ, logically implies the known fact ϕ. By definition, the rule of completeness is a subspecies of the virtue of completeness. More precisely, as ϕ is already known at the time of the abductive inference, the rule of *completeness* matches the virtue of *fitting with background data* or *evidential accuracy*. The case with the rule of consistency is more complicated. If the set of background knowledge Θ contains only the theory at hand, then the rule of consistency may be identified with the virtue of internal consistency. If Θ contains the whole set of theories that we consider legitimate, then the rule of consistency may be identified with the combined virtue of internal and external consistency (or consistency per se).

The rule of non-triviality does not match any of the well-known virtues. It serves as a logical trick to avoid self explanations as it prohibits the adoption of the explanandum ϕ as the explanans α. Finally, the rule of minimality can be seen as a restricted version of the virtues of simplicity or parsimony. Minimal explanations are the logically weaker ones. In that sense, minimality may be identified with the subvirtue of beauty or logical weakness. On the other hand, some major virtues such as unification, ad hocness and some versions of simplicity, along with scope, mechanism etc, remain without a formal equivalent translation.

Hence, there are some criteria for evaluating hypotheses that they have no formal match. That is, there are some unique properties of IBE that formal abduction lacks. So, IBE and formal abduction are not internally equivalent.

6. External Equivalence

We will say that the two methods are *externally equivalent* if in every problem they end up with the same set of solutions. We do not evaluate their unique characteristics. On the contrary, we are interested only in their input and output.

Hence, if we wish to establish the non-equivalence of IBE and formal abduction, it suffices to find an abductive problem (or a single class of abductive problems) where they end up with different outcomes. That is, a problem where the conclusion of the IBE is way different than the conclusion of each AKM model. We argue that the various AKM models face serious difficulties in cases where the problem of *causal determination* arises, that is, in cases in which two or more distinct events c1, c2, ... jointly cause an effect e. For example, when iron is exposed to moisture or oxygen, oxidation occurs. Usually, an exposed piece of iron rusts as both moisture and atmospheric air are present simultaneously.

IBE evaluates potential explanations according to their conformity with a set of desirable explanatory virtues and licenses the acceptance of the best among them. We will examine three of these virtues in detail, namely *elegance*, *parsimony* and *completeness*, as they will help us build the case for the external differentiation between IBE and formal abduction.

We will conclude that simple explanations satisfy the virtues of elegance, parsimony, and completeness. Complex explanations satisfy parsimony and completeness. Superfluous explanations satisfy only completeness. Non-explanations satisfy none of these virtues. In the previous example, if we adopt the conjunction of the three virtues mentioned above, we arrive at the following order:

{moisture} > {moisture, oxygen} > {moisture, nitrogen} > {nitrogen}

In a nutshell, IBE, by employing appropriate explanatory virtues, can distinguish among and rank possible explanations in such a way that simple explanations are considered better than complex ones, complex ones better than superfluous, and the latter better than non-explanations. Thus, in cases of causal overdetermination, where two or more causes/explanations of a phenomenon are present, simple explanations will be preferred over complex ones, without, however, excluding the latter or treating them as equally good as explanations that include hypotheses not involved in the causal chain.

On the other hand, the formal models either confront some among these four species of explanations as equally good or they merge a couple of classes of different explanations. More specifically, the set of virtues that explanatory abduction formalizes does not contain elegance or parsimony. In the previous example, the picture we obtain is the following:

$$\{\text{moisture}\} \approx \{\text{moisture, oxygen}\} \approx \{\text{moisture, nitrogen}\} > \{\text{nitrogen}\}$$

As explanatory abduction allows the generation of complex explanations, it can cope with cases of simultaneous causation.

On the contrary, the set of virtues that the minimal abduction formalizes does contain simplicity, as a simple explanation is, by definition, logically weak. Hence, only simple explanations will appear in the set of the solutions. At this point, the picture that emerges is the following:

$$\{\text{moisture}\} > \{\text{moisture, oxygen}\} \approx \{\text{moisture, nitrogen}\} \approx \{\text{nitrogen}\}$$

Thus, models of minimal abduction fail to address cases of causal overdetermination, while models of explanatory abduction can handle such cases, albeit at the cost of generating a large number of solutions with superfluous content. Hence, in any abductive problem where causal overdetermination arises, IBE yields a different evaluation of the alternative hypotheses than both explanatory and minimal models.

Even a combination of the two methods, first applying explanatory abduction to screen off non-explanations and then applying minimal abduction to the remaining explanations, would not solve the problem. In both methods, complex and superfluous explanations are treated in the same way: either as equally good or as equally bad, in explanatory and minimal abduction respectively. Thus, in AKM models, there is no way to discriminate between a complex and a superfluous explanation. Therefore, IBE and formal abduction are not externally equivalent.

Hence, we are dealing with two distinct methods, each with its own advantages and disadvantages. Although they share a common field of application and some explanatory values, we are far from characterizing them as equivalent. Likewise, we are far from assuming that IBE is the descriptive version of AKM models or that the latter is the formal version of the former.

References

Aliseda-Llera, A. (2006). Abductive reasoning: Logical investigations into discovery and explanation. Dordrecht, the Netherlands: Springer.

Gauderis, T. (2012). Modelling abduction in science by means of a modal adaptive logic. Foundations of Science, 18(4), 611-624.

Kuhn, T. (1977). Objectivity, Value Judgement, and Theory Choice. In The Essential Tension, Chicago: University of Chicago Press, 320-333.

Magnani, L. (2001). Abduction, reason and science: Processes of discovery and explanation. Springer Science & Business Media.