Confirmation Theory and the Logic of Inductive Implication

Ricardo S. Silvestre[1], Tarcísio H. C. Pequeno[2]

[1] Department of Philosophy, University of Montreal,
2910 Édouard-Montpetit, Montréal, QC, H3T 1J7, Canada
`ricardo.silvestre@umontreal.ca`
[2] Department of Computer Science, Federal University of Ceará,
Bloco 910, Campus do Pici, Fortaleza-Ceará, 60455-760, Brazil
`thcp@ufc.br`

Abstract. The general purpose of this paper is to demonstrate through a well defined example how philosophy of science and Artificial Intelligence (AI) can benefit from each other by sharing some of their ideas, methods and techniques developed to tackle similar problems. The problem on which we will focus is the analysis of non-deductive inferences, which is performed in AI by the study of nonmonotonic commonsensical reasoning, and in philosophy of science by the so-called theory of inductive or evidential confirmation. After analyzing to what extent one of the most wide spread nonmonotonic formalisms – default logic – can be taken as a logic of confirmation in the philosophical sense, and carefully considering the similarities and dissimilarities of the problems faced in these contexts, an expanded version of default logic, called by us a logic of inductive implication, is introduced. It is then shown how this new framework can be used to represent different types of inductive calculi that may be of relevance to both AI and philosophy of science research. In particular, it is shown how our logic can formalize some traditional ideas concerning confirmation theory, in particular the ones proposed by Carl Hempel in his classical paper "Studies in the Logic of confirmation" of 1945 and the ones incorporated by the so-called abductive and hypothetico-deductive models. From a more theoretical point of view, we believe that these formalizations represent a relevant step in the understanding of the logic of evidential or confirmatory support

1 Introduction

In the introduction of a somewhat philosophical book of essays on Artificial Intelligence [4], the editors espouse the thesis that in the field of AI "traditional philosophical questions have received sharper formulations and surprising answers", adding that "… important problems that the philosophical tradition overlooked have been raised and solved [in AI]". They go as far as claiming that "Were they reborn into a modern university, Plato and Aristotle and Leibniz would most suitably take up appointments in the department of computer science." Even recognizing a certain amount of over enthusiasm and exaggeration in those affirmations, the fact is that there are evident similarities between some problems faced in AI practice and some classic ones dealt with within philosophical investigation. However, although there is some explicit contact between AI and philosophy in fields like philosophy of mind and philosophy of language, the effective contribution of ideas, methods and techniques from AI to philosophy is still something hard to be seen. In this paper we continue a project started in some previous works [12], [13], [14] and present what we believe to be a bridge between these two knowledge fields that, in addition to its own interest, can also serve as an example and an illustration of a whole lot of connections we hope to go over in the near future.

In order to make that bridge concrete, let us start with the not so deep observation that the study of non-deductive inferences has played a fundamental role in both artificial intelligence and philosophy of science. While in the former it has given rise to the development of *nonmonotonic logics* [8], [9], [10], in the later it has attracted philosophers in the pursuit of a so-called *logic of induction* [3], [6], [7] which was supposed to formalize the logical properties of the relation of *evidential* or *inductive confirmation*. Perhaps in virtue of the technical devices used in these two areas were *prima facie* quite different, the obvious fact that both AI researchers and philosophers were dealing with the same, or almost the same problem has remained virtually unnoticed. Of course, realizing this obvious connection is important because, at least in principle, computer scientists and philosophers can benefit from the results achieved by each other.

It is our purpose here to lay down what we believe to be an instance of (the fruitfulness of) such cooperation. From one hand, we pick one of the most wide spread nonmonotonic logics – default logic [10] – and try to find out to what extent it can be considered as a logic of induction in the philosophical sense as well as which sort of adjustments should we make to transform it into such a logic. This will be done in Section 2. In Section 3 we make use of the conclusions laid down in the previous sections to introduce a modified version of default logic which, we believe, has interesting properties from the standpoint of the theory of confirmation. In Sections 4 and 5 we try to justify this last claim by showing how our logic can be used to formalize some classical models of confirmation found in the philosophy of science literature, including the abductive and hypothetico-deducive models. Finally in Section 6 we present some concluding remarks.

2 The Logic of Induction and Default Logic

Since the time of Rudolf Carnap [3], induction has been conceived (even though not uncontroversially) as *the class of rational non truth-preserving inferences*, being the task of inductive logic to analyze such sort of inferences. The conception of non truth-preserving inference is straightforward; it means an inference whose conclusion may be false even when its premises are true. This contrasts with the second key term in the definition, “rational,” which philosophers have shown to be a quite problematic term both in its characterization and in its operationality.

Considering single defaults as inferences rules, it is clear that default logic satisfies the negative, non truth-preservingness feature of induction: conclusion β of default α:φ/β may be false even in the case where its premise α is true. But how about the hard, positive side of the concept of induction: its rationality? According to Carnap [3] and Carl Hempel [6], for example, the purpose of the logic of induction is basically one of *confirmation*, i.e., given a piece of evidence *e* and a hypothesis *h*, it should say whether (and possibility to what extent) *e* confirms or gives evidential support to *h*. In this way, the distinction between inductive inferences and fallacies will be achieved automatically: despite being non truth-preserving inferences, fallacies do not exhibit any sort of premises-conclusion confirmation relation.

Now, there is a strong parallel between default rules and the qualitative form of these confirmation sentences. Since α:φ/β allows us to infer β only provisionally, we can say it means something like "α might be taken as an evidence for the hypothesis that β is the case with the proviso that $\neg\varphi$ is not the case." Taking a well-known example, the fact that Twenty is a bird confirms or gives evidential or inductive support to the hypothesis that it flies unless we know that it does not fly. In this way, we can read α:φ/β as "α confirms or inductively supports β, with the proviso that $\neg\varphi$ is not the case," or, equivalently, "α inductively supports β unless $\neg\varphi$."

From this perspective, it is clear why the positive side of inductive inferences is not taken into account by default logic. While a logic of induction is supposed to function as a black box having as input several pair of sentences *e-h* and as output a smaller number of sentences of the form "*e* inductively supports *h*," default logic deals only with the output of such box, functioning as a tool for representing such confirmation sentences and detaching their hypotheses from the evidences. As a consequence of that, it is aloof from the problem of positively characterizing inductive inferences: such a task is left to the knowledge engineer, who can use default logic to formalize rational non truth-preserving inferences but also any sort of nonsense. In order to really deal with both input and output, we would need to have a sort of meta-default logic able to automatically "generate" defaults from defaults or from something else than defaults as well as to reason about them. In other words, we would have to have a calculus of defaults.

Now, how about if we had such calculus of defaults? What sort of new resources would we need to have in addition to the ones provided by default logic? And, which is more important, what kind of advantages this logic would give to us? Before answering these questions, let us take a look at another feature of the philosophical project of building a logic of induction which happens to be of fundamental importance for our comparative study: the notion of probability.

If we have that evidence *e* confirms or supports hypothesis *h*, it is natural to wonder what we can conclude about *h* when *e* is true. Despite the diversity of approaches, all theorists agree on one basic point: given that *e* confirms *h* and that *e* is true, whatever we conclude about *h*, it should reflect the uncertainty inherent to inductive inferences. Almost invariably some *probability* notion has been chosen to do this job: even though from "e confirms h" and "e is true" we cannot conclude that *h* is true, we can conclude that it is probable. This notion of probability can be seen as an epistemic label we attach to inductive conclusions in order to make explicit their defeasible character. Carnap calls it pragmatical probability; we will use the less problematic term "*plausibility*."

It is important to note that characterizing inductive inferences in terms of pragmatical probability or plausibility implies changing our previous understanding of confirmation statements. Considering that the truth of evidence *e* warrants us to inductively conclude the plausibility of hypothesis *h* instead of its truth, what *e* confirms or evidentially supports is not the truth of *h*, but its plausibility. Therefore, rather than saying that *e* confirms or inductively supports *h*, we should say that *e* confirms or inductively supports the *plausibility* of *h*. And given that "*h* is plausible" will be possibly inferred, the whole thing might be read as "*e* inductively implies the plausibility of *h*." We call such statements *inductive implications*.

Turning back to AI, it is pretty clear that default logic embodies the mentioned detachment mechanism. After all, its whole purpose is exactly to find out which consequents of defaults can be concluded. So therefore reading $\alpha{:}\varphi/\beta$ as "α inductively implies β unless $\neg\varphi$" as well seems quite suitable. But here we arrive at a breakdown in our reading of defaults in terms of confirmation theory. Since original default logic provides no means to distinguish nonmonotonically obtained conclusions from deductively obtained ones, there is no way to represent "β is plausible" and consequently no way to represent the statement "α inductively implies 'β is plausible' unless $\neg\varphi$."

One might be wondering whether there is some harm in that. Well, it does harm from a philosophical point of view, by not representing the epistemological state of matters properly and consequently promoting the confusion between well-established conclusions and defeasible ones. From the point of view of formal reasoning, this lack of terminological precision does also have some problematic consequences, which will be centered around the question of whether or not nonmonotonic conclusions should be treated in further reasoning upon them in the same way as monotonic ones. More specifically, it prevents us from seriously attacking the so-called problem of *anomalous extensions*, which is due to the arising of contradictions among extensions. By considering seriously the distinction between ontological and epistemological contradictions [9], which are the ones we are faced with when dealing with inductive inferences, we can make use of an existing distinction between deductively and inductively obtained formulae to treat these later paraconsistently, tolerating so-called plausible contradictions without allowing a trivialization of the theory [2], [5], [9].

3 A Logic of Inductive Implication

According to what we have discussed so far, any attempt to transform default logic into a logic of induction should encompass two things: (1) it must allow us to inferentially obtain defaults or inductive implications, or, which would be even better, to represent through a calculus of default or inductive implication specific ways according to which defaults are inferred, and (2) mark the consequent of defaults with some plausibility symbol so as to be able to express its special epistemic status and tolerate the contradictions that may arise from their use.

For the second task, we shall refer to some recent theoretical results pointing to the connections between paraconsistent logic and modal logic [1], [2] and use this latter as a sort of logic of plausibility. Taking $\Diamond\alpha$ as meaning "α is plausible," we shall force the consequents of all defaults to be of the form $\Diamond\alpha$. As result of that, the contradictions that might arise from the use of defaults will necessarily be of the form $\{\Diamond\alpha, \Diamond\neg\alpha\}$, being easily manageable by modal logic.

From a semantic point of view, the possible worlds of this plausibility interpretation of modal logic will be taken as *plausible worlds*. Now the interesting point about this kind of possible world is that since the consequent of our defaults will be $\Diamond$-marked, there will be a close parallel between plausible worlds and the extensions that would be generated by the corresponding $\Diamond$-less defaults. Given a particular ordinary default theory T, the contradictions that may eventually be inferred from it will be accommodated in different self-consistent extensions. If α belongs to all these extensions we say that it is a skeptical consequence of T; if it belongs to at least one we say it is a credulous consequence of T. Now, by marking the consequent of defaults with $\Diamond$ and using a modal logic as our underlying monotonic logic, contradictory conclusions will be accommodated in the same extension and the set of old extensions will correspond to a semantic interpretation to this new modal extension, each extension being an individual plausible world. In this way, $\Diamond$ will correspond to the credulous consequence relation, being thus called credulous plausibility, and $\Box$ ($\Box\alpha =_{\text{def}} \neg\Diamond\neg\alpha$) to the skeptical one, representing then a sort of skeptical plausibility notion [2].

For the first task, we shall consider an expansion of a specific modal calculus (interpreted as a calculus of plausibility) in such a way that inductive implications or defaults are added to the logical language and treated as atomic formulae by its axiomatic machinery. In this way we will be able to make defaults and ordinary formulae to interact with the help of standard logical connectives as well as have defaults appearing as the prerequisite, justification or consequent of another default. This logic will be used as the monotonic basis of our meta-default logic. In our representation of inductive implications, we shall change Reiter's notation in such a way as to better reflect our new interpretation. We also drop the explicit reference to the normal part of the default [2], automatically preventing the so-called abnormal defaults.

Definition 1. Let $\Im$ be a language. The *inductive language* $\Im_{\succ}$ built over $\Im$ is defined as follows: (i) If $\alpha\in\Im$ is such that it contains no one of $\Im$'s logical symbols, then $\alpha\in\Im_{\succ}$; (ii) If $\oplus$ is a monadic logical symbol of $\Im$ along with one of its non-logical complements, if there is any, and $\alpha\in\Im_{\succ}$, then $(\oplus\alpha)\in\Im_{\succ}$; (iii) If $\oplus$ is a dyadic logical symbol of $\Im$ and $\alpha,\beta\in\Im_{\succeq}$, then $(\alpha\oplus\beta)\in\Im_{\succ}$; (iv) If $\alpha,\beta,\varphi\in\Im_{\succ}$, then $(\alpha\succ\beta\not\precsim\varphi)\in\Im_{\succ}$; (v) Nothing else belongs to $\Im_{\succ}$.

Item (iv) defines our version of Reiter's defaults, which we shall call from now on simply inductive implications. $\alpha\succ\beta\not\precsim\varphi$ means "α inductively implies β unless φ" and is equivalent to default $\alpha{:}\beta\wedge\neg\varphi/\beta$. Right away we see that given propositional language L, for example, $L_{\succ}$ will contain formulae like $\alpha\succ(\beta\succ\varphi\not\precsim\lambda)\not\precsim\phi$, $(\beta\succ\varphi\not\precsim\lambda)\succ\alpha\not\precsim\phi$, $\alpha\wedge(\varphi\succ\lambda\not\precsim\beta)$ and $(\alpha\succ\beta)\rightarrow((\beta\succ\varphi)\rightarrow(\alpha\succ\varphi))$. That is to say, besides allowing us to represent specific ways according to which defaults are inferred, by putting $\succ$ and $\not\precsim$ on the same level as the other logical connectives, $\Im_{\succ}$ fully explores the representational potential of defaults. We call α the antecedent of the inductive implication, β its consequent and φ its exception. $\alpha\succ\beta$ is an abbreviation of $\alpha\succ\beta\not\precsim\bot$ and $\beta\not\precsim\varphi$ is an abbreviation of $\top\succ\beta\not\precsim\varphi$. We call any formula that is not an inductive implication an ordinary formula.

We represent a modal calculus M by a triple $<\Im,\Theta,\Lambda>$, where $\Im$ is its language, Θ its set of modal operators and Λ its set of axiom schemas and inference rule schemas. If we want to know the real set of axioms of M, for example, we have just to take the set of all formulae of $\Im$ satisfying one of the axiom schemas of Λ. We also say that $\Im$ is based on Θ. Classical logic could be represented in this notation, for example, by taking $\Im$ as propositional or first-order logic, $\Theta = \varnothing$ and Λ as one of its sets of axioms.

Definition 2. Let M = $<\Im,\Theta,\Lambda>$ be a modal calculus. The *pseudo-inductive modal logic* M' based on M is the modal calculus $<\Im_{\succ},\Theta,\Lambda>$, where $\Im_{\succ}$ is the inductive language built over $\Im$.

The set of axioms of M' is simply the set of all formulae of $\Im_{\succ}$ which satisfies at least one of Λ's axiom schemas. The term "pseudo-inductive" indicates that the calculus in question is deductive rather than inductive but nevertheless contains and reasons (deductively) about inductive implications.

Definition 3. A *pseudo-inductive logic of plausibility* P is a pair <M', Θ'> where M' = $<\Im_{\succ},\Theta,\Lambda>$ is a pseudo-inductive modal logic and $\Theta'\subseteq\Theta$ is a set of modal operators.

The difference between P and M' is that in P we have chosen a subset of Θ to be our plausibility modal operators. Because of that we call it a *logic of plausibility*. The definitions below show how Θ' shall play the role of a plausibility modality.

Definition 4. Let $\Im$ be a modal language based on a set of modal operators Θ and $\theta\in\Theta$ a modal operator. The notion of θ-*formula* is defined as follows: (i) If $\alpha\in\Im_{\succ}$ is of the form $\theta\varphi$, then α is a θ-inductive formula; (ii) If $\alpha\in\Im_{\succ}$ is a θ-inductive formula, then $\alpha\wedge\beta$, $\alpha\vee\beta$, $\alpha\rightarrow\beta$ and $\forall x\alpha$ are also θ-inductive formulae; (iii) If $\beta\in\Im_{\succ}$ is a θ-inductive formula, then $\alpha\succ\beta\not\precsim\varphi$ is a θ-inductive formula; (iv) Nothing else is a θ-inductive formula.

Definition 5. Let P = <M', Θ'> be a pseudo-inductive logic of plausibility with M' = $<\Im_{\succ},\Theta,\Lambda>$. The P-*inductive language* $\Im_P$ is defined as follows: (i) If $\alpha\in\Im_{\succ}$ is an ordinary closed formulae, then $\alpha\in\Im_P$; (ii) If $\alpha\in\Im_{\succ}$ is a closed θ-inductive formula such that $\theta\in\Theta'$, then $\alpha\in\Im_P$; (iii) Nothing else belongs to $\Im_P$. We call any set $A\subseteq\Im_P$ a P-*theory*.

Bellow we lay down the definitions which make clear how P-theories shall be used in order to generate what we have called P-extensions, which in its turn will be used to define the concepts of P-inductive consequence relation and inductive basis. The full explanation of these notions and their use will be given in the following sections.

Definition 6. Let P = <M', Θ'> be a pseudo-inductive logic of plausibility with M' = $<\Im_{\succsim},\Theta,\Lambda>$, $A \subseteq \Im_{\succsim}$ a P-theory and $S \subseteq \Im_{\succsim}$ a set of closed formulae. $\Gamma(S) \subseteq \Im_{\succsim}$ is the smallest set satisfying the following conditions: (i) $A \subseteq \Gamma(S)$; (ii) If $\Gamma(S) \vdash_{M'} \alpha$ then $\alpha \in \Gamma(S)$; (iii) If $\alpha \succsim \beta \not\succsim \varphi \in A$, $\alpha \in \Gamma(S)$ and $\neg\beta \notin S$ and $\varphi \notin S$, then $\beta \in \Gamma(S)$. A set of formulae E is a P-*extension* of A iff $\Gamma(E) = E$, that is, E is a fixed point of operator Γ.

Definition 7. Let P = <M', Θ'> be a pseudo-inductive logic of plausibility with M' = $<\Im_{\succsim},\Theta,\Lambda>$, $A \subseteq \Im_{\succsim}$ a P-theory and $\alpha \in \Im_{\succsim}$ a formulae. α is a P-*inductive consequence* of A (in symbols: $A \vdash_P \alpha$) iff, for all P-inductive extensions E of A, $\alpha \in E$.

Definition 8. An *inductive basis* L is a pair $<P,\vdash_P>$ where P is a pseudo-inductive logic of plausibility and $\vdash_P$ is its relation of inductive consequence.

About the technicalities of our formulation, we first observe that in order to represent default statements such as "typically birds fly" we do not need to consider schemas of defaults like Reiter does. Rather, we just use $\forall$ along with an open formula, in this case an inductive implication, and obtain formulae of the sort $\forall x(\text{bird}(x) \succsim \text{flies}(x))$. Second, as already mentioned, we make the test of consistency of the consequent inside the very definition of extension, which automatically prevent so-called abnormal defaults. Third, by restricting ourselves to P-theories, we make sure that every ordinary formula inferred nonmonotonically will be marked with the plausibility modalities of Θ'. Besides being in accordance with the philosophy of confirmation we have sketched above, this also allows us to properly deal with contradictions. Since item (ii) uses the inference relation of M' ($\vdash_{M'}$) instead of classical logic's, taking any normal modal logic as M and $\Theta' = \{\Diamond\}$ enables us to reason about plausible contradictions ($\Diamond\alpha$ and $\Diamond\neg\alpha$) without trivializing the theory. Finally, as already observed, M' is capable of reasoning about inductive implications themselves. For instance, we will be able to set "laws" (either monotonically or nonmonotonically) about inductive implications such as what we shall call the transitivity property of defaults: $(\alpha \succsim \beta) \rightarrow ((\beta \succsim \varphi) \rightarrow (\alpha \succsim \varphi))$.[1]

4 Hempel's Calculus of Confirmation

We now define the inductive basis we shall use in the rest of the paper. It will be built upon modal logic S5 with $\Diamond$ as its primitive modal operator meaning "it is plausible that."

---

[1] For more on the technical properties of our logic and its relation with Reiter's default logic see [14] and [15].

Definition 9. Let S5' be the pseudo-inductive modal logic based on modal calculus S5. The pseudo-inductive logic of plausibility $P_{\diamond}$ is the pair $<S5', \{\diamond\}>$ and the inductive basis $L_{\diamond}$ is the pair $<P_{\diamond}, \vdash_{P\diamond}>$.

Since an inductive basis does not set any property of inductive implications, it cannot perform the task of generating defaults we have agreed an inductive logic should perform. The only thing it does concerning formulae of the form $\alpha \succ \beta \not\precsim \varphi$ is to detach the consequent from the antecedent. It is like a calculus of material implication provided with MP but with no axioms about $\rightarrow$. However, akin to such an implication-axiom-less calculus, an inductive basis provides the basic tools with which we can build so-called inductive axioms and obtain something worthy of being called a logic of induction or a calculus of inductive implication.

Definition 10. Let $P = <M', \Theta'>$ be a pseudo-inductive logic of plausibility with $M' = <\Im_{\succ}, \Theta, \Lambda>$, $A \subseteq \Im_{\succ}$ a P-theory, $T \subseteq \Im_{\succ}$ a P-theory called the set of inductive axioms and $\alpha \in \Im_{\succ}$ a formulae. α is a T-P-*inductive consequence* of A (in symbols: $A \vdash_{T\text{-}P} \alpha$) iff $T \cup A \vdash_{P} \alpha$.

Definition 11. A *logic of induction* or *calculus of inductive implication* C is a triple $<P, T, \vdash_{T\text{-}P}>$ where $P = <M', \Theta'>$ is a pseudo-inductive logic of plausibility with $M' = <\Im_{\succ}, \Theta, \Lambda>$, $T \subseteq \Im_{\succ}$ is a P-theory representing the set of inductive axioms and $\vdash_{T\text{-}P}$ is the T-P relation of inductive consequence. We also refer to $\vdash_{T\text{-}P}$ as $\vdash_{C}$.

Now we may ask: Which sort of inductive implications are worth being taken as inductive axioms? Considering that inductive implication sentences $\alpha \succ \diamond\beta \not\precsim \varphi$ are our representation of confirmation statements of the form "α inductively *confirms* β unless φ," it seems reasonable to try to answer this question by looking at some general conditions philosophers have proposed to characterize the minimal properties that every definition of confirmation is supposed to satisfy (which are more or less like the properties represented in an implication calculus which are supposed to set the basic properties of a specific kind of implication sentence.) The set of conditions we will examine here the one proposed by Carl Hempel's in his classical paper of 1945 "Studies in the Logic of Confirmation" added by a few more conditions proposed by other philosophers.

In [8], Carl Hempel proposed a set of necessary conditions that any model of confirmation is supposed to satisfy: (I) *Entailment condition*: if statement α entails (i.e., logically implies) statement β, then β should be confirmed by α; (II) *Consequence condition*: if statement α confirms statement β and β logically implies statement φ, then α should also confirm φ; (III) *Equivalence condition*: if statement α confirms statement β and β is logically equivalent to φ, then α should also confirm φ; (IV) *Weak Consistency condition*: if statement α confirms statement β and α is not self-contradictory, then α and β should be logically compatible; (IV') *Strong Consistency condition*: if statement α confirms statements β and φ and α is not self-contradictory, then β and φ should be logically compatible.

This list might be lengthened by some few additional conditions: (V) *Inverse Equivalence condition*: if statement α confirms statement β and α is equivalent to statement φ, then φ should confirm β; (VI) *Transitivity condition*: if statement α

confirms statement β and β confirms statement φ, then α should confirm φ. Together, the set of conditions I-VI entails the following derivate conditions: (VII) *Inverse Consequence condition*: if statement α entails statement β and β confirms statement φ, then α confirms φ; (VIII) *Inclusion Condition*: every statement confirms itself.[2]

Since we shall interpret $\alpha\succsim\diamond\beta\not\precsim\varphi$ as "α confirms $\diamond\beta$ unless φ," condition IV' is automatically satisfied by $L_\diamond$. What follows below are the axioms of what we shall name Hempel calculus of confirmation.

Definition 12. Let P be a pseudo-inductive logic of plausibility. The *Hempel confirmation axioms* T_H in $\Im_P$ is the set composed by all formulae of $\Im_P$ satisfying the following schemas of formula:

I:	$(\alpha\rightarrow\beta)\succ(\alpha\succsim\diamond\beta)\not\precsim((\alpha\leftrightarrow\bot)\vee(\top\leftrightarrow\beta))$	*Entailment*
II:	$(\beta\rightarrow\varphi)\succ((\alpha\succsim\diamond\beta\not\precsim\phi)\rightarrow(\alpha\succsim\diamond\varphi\not\precsim\phi))\not\precsim(\top\leftrightarrow\varphi)$	C*onsequence*
III:	$(\alpha\succsim\diamond\beta\not\precsim\phi)\rightarrow((\beta\leftrightarrow\varphi)\rightarrow(\alpha\succsim\diamond\varphi\not\precsim\phi))$	*Equivalence*
IV:	$(\alpha\succsim\diamond\beta\not\precsim\varphi)\rightarrow((\alpha\wedge\beta\rightarrow\bot)\rightarrow\bot)$	*Weak consistency*
V:	$(\alpha\succsim\diamond\beta\not\precsim\phi)\rightarrow((\alpha\leftrightarrow\varphi)\rightarrow((\varphi\succsim\diamond\beta\not\precsim\phi))$	*Inverse Equivalence*
VI:	$(\alpha\succsim\diamond\beta\not\precsim\phi)\rightarrow((\beta\succsim\diamond\varphi\not\precsim\phi')\rightarrow(\alpha\succsim\diamond\varphi\not\precsim\phi\vee\phi'))$	*Transitivity*
VII:	$(\alpha\rightarrow\beta)\succ((\beta\succsim\diamond\varphi\not\precsim\phi)\rightarrow(\alpha\succsim\diamond\varphi\not\precsim\phi))\not\precsim(\alpha\leftrightarrow\bot)$	*Inverse Consequence*
VIII:	$\alpha\succsim\diamond\alpha$	*Inclusion*

Definition 13. Let T be Hempel confirmation axioms in $\Im_{P\diamond}$. The *Hempel calculus of confirmation* C_H is the triple $<P_\diamond,T,\vdash_{T\text{-}P\diamond}>$.

The reason for representing I, II and VII through an inductive implication instead of a material implication formula is due to inability of → to capture the relevance aspect required by a confirmation relation. If we represent I by $(\alpha\rightarrow\beta)\rightarrow(\alpha\succsim\beta)$, for example, we will have that for any sentence α and β, $\alpha\succsim\diamond\top$ and $\bot\succsim\diamond\beta$, i.e., the plausibility of a tautological formula is confirmed by α and a contradictory one confirms $\diamond\beta$, for any two formulae α and β. And because we have represented I and II as an inductive implications instead as material implication formulae, we cannot infer VII and VIII from them. Consequently we have to add them to our set of inductive axioms.

5 The Abduction and Hypothetico-Deductive Models

As it can be easily seen, the axioms of the Hempel calculus of confirmation function exactly like a calculus of confirmation, setting the general properties according to which confirmation statements are supposed to be obtained from prior, already existing confirmation statements. They however do not say anything about how these priori confirmation statements are supposed to be generated, which is the task of what we have been calling *model of confirmation*. In this section we will examine

---

[2] Formulations of V and VI have appeared, respectively, in [11] and [7]. VII is obtained from I and VI and VIII is an special case of I.

how to obtain through our framework a problem-free representation of the *hypothetico-deductive* (H-D) model and its brother-model the *abductive* model.

What we call the abductive model of confirmation is the confirmatory form of the so-called abductive reasoning: α confirms β if $\beta \rightarrow \alpha$ (if α is true we then get that β is plausible, ending then the abductive reasoning chain.) It is present in Hempel’s 1945 article in the form of the (IX) *Converse Consequence condition*: if statement α confirms statement β and statement φ logically implies β, then α confirms φ, which along with VIII implies the (X) *Converse Entailment condition*: if statement α logically implies statement β, then β confirms α. Now we may think that all we have to do obtain a Hempelian abductive model is to add these conditions to Hempel calculus of confirmation. Not quite so. As Hempel showed, these two conditions are incompatible with his previous conditions.

Consider for example the Stark effect (Se) which is known to confirm quantum mechanics (Qm). Let also Bl be the hypothesis that black cats bring bad luck. Trivially, $\mathrm{Qm} \wedge \mathrm{Bl} \rightarrow \mathrm{Qm}$. But since Se confirms Qm, we have by IX that Se confirms $\mathrm{Qm} \wedge \mathrm{Bl}$. Now, given that $\mathrm{Qm} \wedge \mathrm{Bl} \rightarrow \mathrm{Bl}$, by II we have that the absurd conclusion that Se supports the hypothesis that black cats bring bad luck. In fact, Se will confirm not only Bl, but any statement expressible in the language at hand. Things get still worse when we consider X, which leads to the conclusion that any pair of statement *e-h* whatsoever is such that *e* confirms *h*. The same unwanted conclusion could be derived if we consider disjunctive statements rather than conjunctive ones. Since $\alpha \rightarrow \alpha \vee \beta$, by I we have that α confirms $\alpha \vee \beta$. But since $\beta \rightarrow \alpha \vee \beta$, by IX we have that α confirms β. Similarly, by X we have that $\alpha \vee \beta$ confirms β. Since by I α confirms $\alpha \vee \beta$, by the transitivity condition we have that α confirms β.

Because of these problems, Hempel rejected X along with the definition of confirmation which brought it into the discussion: the *prediction-criterion of confirmation*. This prediction-criterion of confirmation is nothing less than a formulation of the so-called *hypothetico-deductive model* of confirmation, whose importance for the contemporary theory of science is such that some philosophers went so far as claiming it to be the official “scientist’s philosophy of science.” Hempel’s formulation of the H-D model goes as follows: α confirms β if (i) $\vdash \alpha \leftrightarrow \alpha' \wedge \alpha''$; (ii) $\{\alpha' \wedge \beta\} \vdash \alpha''$ and (iii) $\{\alpha'\} \nvdash \alpha''$. In words: if α is composed by two statements α' and α'' and α'' can be logically deducted from α' in conjunction with β but not from α' alone, then β is confirmed by α. Since this definition of confirmation satisfies conditions IX and X, it is easy to conclude that it will be plagued by the same problems we have shown in connection with the abductive model.

About the cause of these problems, in contrast to what many philosophers have held, it is not on the H-D and abductive models themselves, but in the tools we are using to represent them. In a nutshell, they all come from the irrelevance feature of classical entailment[3]. When, for instance, we say that if α entails β then β confirms α, we expect that all parts of α are necessary for the derivation of β and therefore deductively connected with it. Now, if α and β are such connected and we conjoin φ to α, trivially β is logically entailed by $\varphi \wedge \alpha$; but φ plays no role at all in the deri-

---

[3] Some few philosophers have already pointed this out. See for instance [16].

vation of β from $\alpha \wedge \varphi$. Therefore we are not in any way ready to say that $\alpha \wedge \varphi$ confirms β, even though α alone does. The same thing happens when we take the disjunction of β and φ. All the incompatibility between IX and X and Hempel's former conditions as well as the problems we have identified with axioms I and II come from this irrelevance feature of classical entailment.

What follows now is an attempt to formulate the abductive and H-D models inside our framework which is not plagued by the mentioned problems. We first define a few abbreviations: (i) $\alpha \Leftrightarrow \beta =_{\mathrm{def}} (\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha)$; (ii) $\alpha \rhd \beta =_{\mathrm{def}} \beta \succ \diamond\alpha$; (iii) $\alpha \unrhd \beta =_{\mathrm{def}} (\alpha \rightarrow \beta) \wedge (\alpha \rhd \beta)$; (iv) $\alpha \ntriangleright \beta =_{\mathrm{def}} (\alpha \rhd \beta) \succ \diamond\bot$. $\alpha \Leftrightarrow \beta$ is a simple abbreviation meaning that α and β are "implicationally connected" to each other. $\alpha \rhd \beta$ is an alternative way of writing $\alpha \succ \diamond\beta$ which will be of some help in our task of representing the abductive method of confirmation. $\alpha \rhd \beta$ can be read as "α is confirmed by β." $\alpha \unrhd \beta$ is intent to represent a situation where α relevantly implies β. It depends directly on what we have called abduction model of confirmation: if α (relevantly) implies β, then β confirms α. That is to say, supposing that we have such a model, if β confirms α and $\alpha \rightarrow \beta$, then α relevantly implies β. Finally, $\alpha \ntriangleright \beta$ represents a situation where α and β are such that, due to the lack of a relevant entailment connection between α and β, α cannot be confirmed by β through the abduction model. We show below the basic axioms which will make use of these abbreviations.

Definition 14. Let P be a pseudo-inductive logic of plausibility. The *Abduction axioms* T_{Ab} in $\mathfrak{I}_{P}$ is the set composed by all formulae of $\mathfrak{I}_{P}$ satisfying the following schemas of formula:

X: $(\alpha \rightarrow \beta) \succ (\alpha \rhd \beta) \not\lesssim (\alpha \ntriangleright \beta)$

Ab1: $((\alpha \leftrightarrow \alpha' \wedge \alpha'') \wedge (\alpha' \unrhd \beta)) \succ (\alpha \ntriangleright \beta) \not\lesssim ((\alpha'' \unrhd \beta) \vee (\alpha' \Leftrightarrow \alpha''))$

Ab2: $((\beta \leftrightarrow \beta' \vee \beta'') \wedge (\alpha \unrhd \beta')) \succ (\alpha \ntriangleright \beta) \not\lesssim ((\alpha \unrhd \beta'') \vee (\beta'' \Leftrightarrow \beta'))$

The purpose of the above axioms is basically to define what we have been calling abductive confirmation. X, which is a more sophisticated formulation of the converse entailment condition, sets the basic abductive criterion according to which formula α confirms formula β: if $\alpha \rightarrow \beta$ then α is confirmed by β. However, as we have seen, material implication does not embody the relevant aspects required by an abductive model of confirmation: sometimes even though $\alpha \rightarrow \beta$, due to α's not being relevantly connected with β, α is not confirmed by α. It is the goal of the exception part of X, $\alpha \ntriangleright \beta$, to block these non-relevance cases and therefore prevent $\alpha \rhd \beta$ from being concluded from $\alpha \rightarrow \beta$. These non-relevance cases are formally defined by axioms Ab1 and Ab2, which basically take into account the conjunction and disjunctive problems of abduction which we have discussed above.

Ab1 says that if α is equivalent to the conjunction of α' and α'', and α' relevantly implies β, then to conjoin α' and α'' and write $\alpha' \wedge \alpha'' \rightarrow \beta$ will be a trivialization with no relevance content. Therefore $\alpha \ntriangleright \beta$. Of course there are exceptions to this. The first one is α'' relevantly implying β, in which case $\alpha' \wedge \alpha''$ should be confirmed by β (which will be obtained by using $\alpha' \wedge \alpha'' \rightarrow \beta$ along with X.) Also, if $\alpha' \rightarrow \alpha''$ or $\alpha'' \rightarrow \alpha'$ then $\alpha \ntriangleright \beta$ should not be the case, for if $\alpha' \rightarrow \alpha''$ then α' will be equivalent to $\alpha' \wedge \alpha''$, and if $\alpha'' \rightarrow \alpha'$, by transitivity $\alpha'' \rightarrow \beta$ and therefore $\alpha'' \rhd \beta$. Hence, $\alpha' \wedge \alpha'' \rhd \beta$. One could think that this second part of $\alpha' \Leftrightarrow \alpha''$ was not needed at all, for, since $\alpha'' \rhd \beta$ (which is obtained by using X along with $\alpha'' \rightarrow \beta$,) the situation was already contemplated by the inductive implication Ab2. However, taking Ab1 without $\alpha'' \rightarrow \alpha'$ in its exception part, and $\alpha'' \rightarrow \alpha'$ and $\alpha' \unrhd \beta$ as valid formulae (which implies $\alpha'' \leftrightarrow \alpha' \wedge \alpha''$) entails a conflict between X and Ab1: by using Ab1 first and concluding $\alpha'' \ntriangleright \beta$ (which could be done because we have not used yet X. to conclude $\alpha'' \rhd \beta$ and be able to block Ab1) we will not be able to use X and conclude $\alpha'' \rhd \beta$. Therefore two extensions would arise. In order to prevent that, we have to consider $\alpha'' \rightarrow \alpha'$ in the very exception part of Ab1.

For Ab2 the reasoning is almost the same. If β is equivalent to the disjunction of β' and β'', and α relevantly implies β', then to write $\alpha \rightarrow \beta' \vee \beta''$ means to go against our relevance principle, for β'' plays no role at all in the derivation of $\beta' \vee \beta''$ from α. Therefore $\alpha \ntriangleright \beta$. About the exceptions, we have first that if α relevantly implies β'' then α should be confirmed by $\beta' \vee \beta''$. Also, if $\beta'' \rightarrow \beta'$ or $\beta' \rightarrow \beta''$ then $\alpha \ntriangleright \beta$ should not be the case, for if $\beta'' \rightarrow \beta'$ then β' will be equivalent to $\beta' \vee \beta''$, and if $\beta' \rightarrow \beta''$ then by transitivity $\alpha \rightarrow \beta''$ and therefore $\alpha \rhd \beta''$. Hence $\alpha \rhd \beta' \vee \beta''$. About the objection that it is not necessary to consider $\beta' \rightarrow \beta''$ as an exception, taking Ab2 without $\beta' \rightarrow \beta''$ in its exception, and $\beta' \rightarrow \beta''$ and $\alpha \unrhd \beta'$ as valid formulae (which implies $\beta'' \leftrightarrow \beta' \vee \beta''$) entails a conflict between X and Ab2: by using Ab2 first and concluding $\alpha \ntriangleright \beta''$ (which could be done because we have not used yet X. to conclude $\alpha \rhd \beta''$ and be able to block Ab2) we will not be able to use X and conclude $\alpha \rhd \beta''$. Therefore two extensions would arise.

It is easy to see that this formulation of the abduction model of confirmation is not plagued by the relevance problems we have been talking about here. Below we have what we can call the abduction logic of induction.

Definition 15. Let T be the abduction axioms in $\Im_{P\diamond}$. The *abduction model of confirmation* C_{Ab} is the triple $<P\diamond, T, \vdash_{T\text{-}P\diamond}>$.

With these abduction axioms at hand we can also define a relevance-problem-free H-D model.

Definition 16. Let P be a pseudo-inductive logic of plausibility. The *H-D axioms* T_{Ab} in $\Im_P$ is the set composed by all formulae of $\Im_P$ satisfying the following schema of formula:

H-D: $(\beta \leftrightarrow \beta' \wedge \beta'') \wedge (\alpha \wedge \beta' \rightarrow \beta'') \succcurlyeq (\beta \succcurlyeq \diamond\alpha) \not\lesssim (\alpha \wedge \beta' \ntriangleright \beta'') \vee (\top \rightarrow \alpha)$

We are using the formulation proposed by Hempel which we have shown at the beginning of this section. There will be two kinds of exceptions to this rule. The first obviously are situations where $\alpha \wedge \beta'$ does not relevantly imply β''. The second are cases where α is a tautology. The reason for this second sort of exception is that since $\beta' \leftrightarrow T \wedge \beta'$, we do not want to take $T \wedge \beta' \rightarrow \beta''$ as an irrelevant implication. Therefore $T \wedge \beta' \ntriangleright \beta''$ will not be the case. But we are also not ready to say that $\beta' \wedge \beta''$ confirm T. The only alternative then is to consider this case as a separated exception. Concerning Hempel's three conditions, we note that the third one $(\{e'\} \nvdash$

e”, which would be represented in our notation by introducing $\beta' \rightarrow \beta''$ in the exception part of H-D) is already contemplated by $\alpha \wedge \beta' \not\succ \beta''$.

Definition 17. Let T_{Ab} and $T_{H\text{-}D}$ be the abduction axioms in $\Im_{P\diamond}$ and the H-D axioms in $\Im_{P\diamond}$, respectively. The *H-D model of confirmation* $C_{H\text{-}D}$ is the triple $<P_\diamond, T_{Ab} \cup T_{H\text{-}D}, \vdash_{T\text{-}P\diamond}>$.

6 Conclusion

In this work we have tried to establish some connections between the field of non-monotonic logic in AI and the philosophical field of inductive logic. We tried to materialize our conclusions by proposing a logical system inspired in Reiter’s default logic that fulfills the purpose of a logic of induction. We think our work represents a valuable contribution to both AI and philosophy in that it shows how AI tools can be useful in treating traditional philosophical problems such as the one of formalizing a theory of confirmation, and on the other hand how ideas traditionally cultivated inside philosophical fields can throw some light upon AI problems, namely the problem of anomalous extensions and the problem of representation of defeasible reasoning.

About how much our logic departs from traditional default logics, the following points can be mentioned: (1) Defaults are represented through $\alpha \succ \beta \not\precsim \varphi$, which is structurally equivalent to Reiter’s semi-normal default $\alpha{:}\beta \wedge \neg\varphi/\beta$ and is read as “α inductively implies β unless φ.” (2) Implicit in this notation is the taking into account of the consistency of the consequent in the definition of extension and the exclusion of abnormal defaults. (3) Rather than using classical logic as our underlying monotonic logic, we suggest that a traditional modal logic be used instead and all defaults have their consequents marked with $\diamond$, meaning “is plausible that.” This has the advantage that now we are able to keep contradictions in the same extension and reason paraconsistently about them as well as to properly represent other cases of anomalous extensions. (4) Finally, defaults or, in our terminology, inductive implications are part of the logical language and treated by the axiomatic machinery more or less like atomic formulae. This has important representational consequences. First, we are able to represent universal defaults with the help of the logical symbol $\forall$ and, which is more important, represent, both monotonically and non-monotonically, laws about defaults.

About this last point, we should mention that our work opens the door to something which has never been done before: the construction of a calculus of defaults or inductive implications akin to the calculus of material implication contained in classical logic. In the same way that MP detaches the consequence from the antecedent of material implication sentences and $\rightarrow$-axioms set the additional properties of $\rightarrow$, we can say that the definition of extension acts exactly like MP, being our responsibility to lay down the inductive axioms which will effectively set the properties we want inductive implications to possess. We have illustrated that by formalizing through what we called a calculus of confirmation the properties which some philosophers have thought any definition of confirmation should have. In addition to that, we have also provided a sound formalization of one of the most important and problematic models of confirmation found in philosophy of science literature: the hypothetico-deductive model. We believe that these two applications – the Hempel model of confirmation and the H-D model of confirmation – illustrate well the representational power of our logic as well as its relevance for both AI and confirmation theory. From a more general point of view, they also serve as evidence for the thesis that the problems dealt with in the fields of nonmonotonic logic and confirmation theory are more similar than what has been commonly thought and can be dealt with inside the same representational framework.

References

1. Bèziau, J. Y.: S5 is a Paraconsistent Logic and so is First-Order Classical Logic. Logical Studies 9 (2002), http://www.logic.ru/LogStud/09/No9-01.html.
2. Buchsbaum, A. Pequeno, T., Pequeno, M.: The Logical Expression of Reasoning. To appear in: Béziau, J., Krause, D. (eds.): New Threats in Foundations of Science. Papers Dedicated to the Eightieth Birthday of Patrick Suppes. Kluver, Dordrecht (2005).
3. Carnap, R.: Logical Foundations of Probability. U. of Chicago Press, Chicago (1950)
4. Ford, M. , Glymour, C., Hayes, P. (eds.): Android Epistemology. The MIT Press (1995).
5. Gabbay, D., Hunter A.: Making Inconsistency Respectable. In: Jorrand, P., Kelemen, J. (eds.): Proc. of Fundamental of AI Research, Springer-Verlag (1991) 19-32.
6. Hempel, Carl G.: Studies in the Logic of Confirmation. Mind 54 (1945) 1-26, 97-121.
7. Hesse, M.: Theories and the Transitivity of Confirmation. Philosophy of Science 37 (1970) 50-63.
8. McCarthy, J.: Applications of Circumscription to Formalizing Commonsense Knowledge. Artificial Intelligence 26 (1986) 89-116.
9. Pequeno, T., Buchsbaum, A.: The Logic of Epistemic Inconsistency. In: Allen, J., Fikes, R., Sandewall, E. (eds.): Principles of Knowledge Representation and Reasoning: Proc. of Second International Conference. Morgan Kaufmann, San Mateo (1991) 453-460.
10. Reiter, R.: A Logic for Default Reasoning. Artificial Intelligence 13 (1980) 81–132.
11. Scheffler, I.: The Anatomy of Inquiry. Alfred A. Knopf, New York (1963).
12. Silvestre, R., Pequeno, T: A Logical Treatment of Scientific Anomalies or AI Meets Philosophy of Science. In: Arabnia, H, Joshua, R., Mun, Y. (eds.): Proceedings of the 2003 Int. Conference on Artificial Intelligence, CSRA Press (2003) 669-675.
13. Silvestre, R., Pequeno, T: Is Plausible Reasoning a Sensible Alternative for Inductive-Statistical Reasoning? In: Bazzan, L.C, Labidi, S. (eds.): Advances in Artificial Intelligence – SBIA 2004 (LNAI 3171), Springer-Verlag, Berlin-Heidelberg (2004) 124-133.
14. Silvestre, R., Pequeno, T: A Logic of Inductive Implication or AI Meets Philosophy of Science II. In: Kégl, B., Lapalme, G. (eds.): Advances in Artificial Intelligence – AI 2005 (LNAI 3501), Springer-Verlag, Berlin-Heidelberg (2005) 232-243.
15. Silvestre, R.: Induction and Plausibility: A Formal Approach from the Standpoint of Artificial Intelligence. PhD dissertation, Department of Philosophy, University of Montreal, Montreal, Canada (2005).
16. Waters, C.: Relevance Logic Brings Hope to Hypothetico-Deductivism. Philosophy of Science 54 (1987) 453-464.