Workshop on Abduction and Induction in AI and Scientific Modelling

Abduction and Induction are forms of logical reasoning with incomplete information that have many applications in AI. Abduction reasons from effects to possible causes and has been used in tasks such as planning and diagnosis. Induction learns general rules for observed data and is typically used for classification and knowledge acquisition.

This Workshop investigates the claim that Abduction and Induction play fundamental roles in the process of scientific knowledge development and that these these complementary forms of reasoning can be profitably integrated in an automated cycle of theory refinement. The contributed papers examine these issues from philosophical, cognitive and practical points of view.

Peter A. Flach
Antonis C. Kakas
Lorenzo Magnani
Oliver Ray

August 2006

Organization

Program Chairs

Peter Flach, University of Bristol, UK
Antonis Kakas, University of Cyprus, Cyprus
Lorenzo Magnani, University of Pavia, Italy
Oliver Ray, Imperial College London, UK

Program Committee

Peter Flach, University of Bristol, UK
Katsumi Inoue, National Institute of Informatics, Japan
Antonis Kakas, University of Cyprus, Cyprus
Lorenzo Magnani, University of Pavia, Italy
Stephen Muggleton, Imperial College London, UK
Oliver Ray, Imperial College London, UK
Alessandra Russo, Imperial College London, UK
Chiaki Sakama, Wakayama University, Japan

Table of Contents

Hasty Generalizers and Hybrid Abducers
External Semiotic Anchors and Multimodal Representations

Lorenzo Magnani [1]

Abstract. First of all I would like to describe inductive and abductive reasoning in the light of the agent–based framework to the aim of clarifying their fallacious character and the role of the so-called ideal systems (logical and computational). Then I will analyze some inductive and abductive types of reasoning that in the perspective of classical and informal logic are defined *fallacies*. I will describe how in an agent-based reasoning this kind of *fallacious* reasoning can in some cases be redefined and considered as a good way of reasoning. Finally, I will illustrate how what I call *manipulative abduction* can be interpreted as a form of practical reasoning a better understanding of which can furnish a description of human beings as *hybrid reasoners* in so far they are users of ideal and computational agents.

1 Beings-Like-Us as Hasty Generalizers

First of all I would like to describe inductive reasoning in the light of the so–called agent–based framework. This analysis will permit us to explain the traits of the fallacious character of induction (and abduction) and the role of the "idealized" logical systems.

It is well-known that in classical logic a good argument is a sound argument and, from a semantic point of view, it is a valid argument based on true premises. Even if this conception of good inference is usually able to model many kinds of argumentation of real human beings, its appeal to true premises is ill suited to many contexts which are often characterized by the presence of hypothetical and uncertain beliefs, by great disagreement about what is true and false, by ethical and aesthetic claims which are not easily categorized as true or false, and, finally, by variable contexts in which dramatically different assumptions may be accepted and rejected.

I share with Gabbay and Woods [8] the idea that logic can be considered a formalization of what is done by a cognitive agent. Starting from this perspective, logic is *agent-based*. In this perspective agent–based reasoning consists in describing and analyzing the reasoning occurring in problem solving situations where the agent access to cognitive resources encounters limitations such as

- bounded information
- lack of time
- limited computational capacity.

Hence, the "beings-like-us" that Woods describes in his "Epistemic bubbles" [26] discharge their cognitive agendas under press of incomplete information, lack of time, and limited computational capacity. We can consequently say that cognitive performances depend on information, time, and computational capacity. An *agent-based logic*, as a discipline that furnishes ideal descriptions of *agent-based reasoning*, returns to be thought of as a science of reasoning and considered agent-centered, task-oriented, and resource-bound. Woods says:

> So, then, a principal function of reasoning is to facilitate cognition, this means the reasoning agent is also cognitive agent. If logic is to press forward as a renewed science of reasoning, it would do well to reflect on what cognitive agency is like, on what it is like to be a knower [26, p. 732].

In dealing with these features we arrive to what has been called the "Actually Happens Rule" [26] that states that "to see what agents should do we should have to look first at what they actually do and then, if there is particular reason to do so, we would have to repair the account". This rule is a particular attractive assumption about human cognitive behavior mainly for two reasons. The first is that beings like us make a lot of errors; the other is that cognition is something that we are actually very good at.

In the following section we will discuss the case of "fallacies" as errors that people make. These errors occur in ways of reasoning and acting that from some perspectives are good and from others are bad. In dealing with this matter I will try to give an account of fallacies seen from the viewpoint of agent-based reasoning. I will try to give some examples of fallacious reasoning treating both informal fallacies (such as the inductive ones like "hasty generalization") and formal fallacies (such as abduction). I will treat induction, and abduction as fallacious ways of reasoning that in spite of their fallacious character are fruitful for the cognitive agent: a way of being rational through fallacies.

Abduction can be easily considered in the perspective of agent-based reasoning because in abductive reasoning [18] both the activity of guessing new explanatory hypotheses and the activity of selecting already existing ones, is based on incomplete information. In this case we deal with "nonmonotonic" inferences: we draw defeasible conclusions from incomplete information. From this perspective, abductive reasoning also represents a prototypical case of practical reasoning: we adopt deliberations based on incomplete information and on particular abduced hypotheses – guesses – that serve as "reasons."

1.1 Induction as a Fallacy in Organic Agents' Reasoning

As already noted, people make errors in reasoning. This means that in analyzing the beings-like-us argumentations we have to face problems regarding agent's access to cognitive resources such as infor-

[1] Department of Philosophy and Computational Philosophy Laboratory, University of Pavia, 27100 Pavia, Italy, and Department of Philosophy, Sun Yat-sen University, 510275 Guangzhou (Canton), P. R. China, email: lmagnani@unipv.it

mation, time, and computational capacity, and logical attributes such as truth-preservation. It is in this sense that I have previously said that agents discharge their cognitive agendas under press of bounded information, lack of time and limited computational capacity.

The successful use of fallacies into many kinds of reasoning can be fruitfully accounted for in the framework of agent-based reasoning. It is undeniable that in human reasoning mistakes are widespread. The peculiarity of fallacies seen in the perspective of agent–based reasoning is that mistakes that are actually committed are mistakes that do not seem to be mistakes to those who commit them. In some sense we can say that they are ways of reasoning that are felt truth preserving for the reasoner but are not considered truth preserving for the logicians!

A fallacy is a pattern of poor reasoning which appears to be a pattern of good reasoning [13]. Fallacies are forms of reasoning and argumentation typical of organic agents and in this sense we can say they are suitably shaped by evolution. Simple inductions and abductions performed more or less consciously by both humans and animals are surely two great results of this evolutionary process. Two main disciplines respectively clearly illustrate different kinds of fallacies: formal logic, which recognizes and explains "formal fallacies", and informal logic, that describes the so-called "informal fallacies". First of all, we can say that the validity of a deductive argument depends on its form, consequently, formal fallacies are arguments which have an invalid form and are not truth preserving (for example the fallacy of the "affirming the consequent" and of "denying the antecedent"). On the other hand, informal fallacies are any other invalid modes of reasoning whose failing is not strictly based on the shape of the argument (for example the "*ad hominem* argument" or the "hasty generalization").

From the point of view of classical logic a fallacy is a bad argument that looks good. From the point of view of agent-based reasoning a fallacy is not an argument that looks good but is bad, but an argument that is bad in some aspects and good in some others. Let us consider the inductive case of the so-called "hasty generalization", that can lead the cognitive agent – in spite of its fallacious character – to fruitful outcomes.

This fallacy occurs when a person (but there evidence of it also in animal cognition, for example in mouses) infers a conclusion about a group of cases based on a model that is not large enough. It has the following form:

- Sample S, which is too small, is taken from the group of persons P.
- Conclusion C is drawn about the group P based on S.

It could take also the form of:

- The person X performs the action A and has a result B.
- Therefore all the actions A will have a result B.

The fallacy is committed when not enough *A*'s are observed to warrant the conclusion. If enough *A*'s are observed then the reasoning is not fallacious, at least from the *informal* point of view. Males, driving their cars, have probably quarreled with a woman driving her car and, while quarreling, they have argued (when not shouted) "all woman are bad drivers!" That's our case of fallacious reasoning.

Insofar, small samples will be likely to be unrepresentative. Another simple case is the following. If we are asking one person that even met a lot of Italians what he thinks about the recently new established Italian proportional–oriented electoral system, his answer clearly would not be based on an adequate sized sample for determining what Italians in general think about the issue. This is because the answer given is based only on a reduced experience and that judgment can not be relevant in dealing with a generalization about the matter in question. This means that this fallacious argument implies that small samples are less likely to contain numbers proportional to the whole group of cases.

People often commit hasty generalizations because of bias or prejudice. For example, someone who is a sexist might conclude that all women are unfit to fly jet fighters (or to drive a car) because one woman crashed in either case. People also commonly commit hasty generalizations because of laziness or sloppiness. It is very easy to simply jump to a conclusion and much harder to gather an adequate sample and draw a justified conclusion. Thus, avoiding this fallacy requires minimizing the influence of bias and taking care to select a sample that is large and meaningful enough.

1.1.1 Casual Truth–Preserving Inferences

Moreover, we can recognize another important occurrence. I have said that people commit errors and are hasty generalizers because of prejudice, mindlessness, bias, and so on. What I am trying to underline is that the hasty generalization is not always a bad generalization for two reasons. The first is that, getting true conclusions, hasty generalization might be good if the result of the generalization we made coincides with the result of a good generalization in the philosophical – for example Millian – sense of induction (or in the sense of inductive logics). We call this case *''casual" truth preserving* feature of hasty generalization. The second reason is that, in some sense, even if we do not reach good conclusions, not exploiting the casual truth preserving feature, we can say that hasty generalization is good in some sense, obviously not in the classical logic one. We will now try to understand what it can be.

Think of a toddler that for the first time touches a stove in his kitchen [25, pp. 314–316]. His finger is now burned because the stove burns. Starting from this evidence, the hasty generalizer toddler thinks that all the stoves are hot and decides not to touch stoves anymore. This is obviously a hasty generalization:

- X of observed A are B (The stove *touched* burns).
- Therefore all A are B (*All* the stoves burn.)

Or:

- Sample S, which is too small, is taken from the group of persons P. (The toddler touches the stove and at a first touch the stove burns).
- Conclusion C is drawn about the group P based on S. (Whenever the toddler will touch the stove, it will burn).

1.1.2 Strategic Rationality

We can also say that this is a case of bad argument also from the formal point of view because it is not truth preserving, in the light of classical logic. However, in the perspective of agent-based reasoning the problem now is: can we say that this argument is good from some perspective? Indeed the hasty generalization is sometimes a "prudent" strategy. It also presents a cognitive economy: given the task of not being burnt for a second time, the hasty generalization is a kind of reasoning that is fruitful because, being a prudent strategy, it embeds the canons of *strategic rationality* in the sense of the "strife for survival". Moreover, it also involves a *cognitive success*.

First, fallacies (hasty generalization in this case) have some relevant relations with strategic rationality. However, the prudent strategy of "not touching the stove" is obviously incorrect for at least two reasons. 1) The first reason is that it is not good to generalize from only one sample available and 2) from applied natural physics, we can say that it is a state of affairs that a stove does not burn because a stove is made of iron or some other metals and metals burn only if they are overheated. So there is something "bad" in this kind of reasoning both from an informal logic point of view and from the perspective of natural physical principles of heat. But even if we recognize these wrong steps, there is an idea of some rationality embedded in this example due to the fact that the toddler prevents himself from being burned. It seems that hasty generalization (like in the case of other fallacies, too, like the fallacy of affirming the consequent can be considered resources that enter in a sort of *human survival kit* [25, p. 7]. As some unconditioned reflexes, hasty generalization is a response (in the form of a reasoning and then of an action) to something that the toddler is involved to. The cognitive result of a hasty generalization is bad but only in the sense that it does not *explain* the burnt stove. It is instead a form of good reasoning because it preserves the toddler from being burnt another time.

Second, hasty generalization also allows the toddler to produce a new *successful cognitive information.* In the perspective of the logical tradition, this piece of information is "bad" because obtained through fallacious reasoning, but in agent-based terms we notice that the same information contributes to solve the toddler's problem and, in this sense, can be endowed with "good" cognitive relevance.

I have contended above that fallacies are forms of reasoning and argumentation typical of organic agents and in this sense we have concluded they are part of a "survival kit" suitably shaped by evolution. I have also added that induction and abduction performed more or less consciously by both humans and animals are surely two great results of this evolutionary process. We know that in the last centuries humans have also characterized induction and abduction in various "ideal" philosophical and logical ways, so going beyond the spontaneous use of those kinds of thinking I have just illustrated. Already Mill provided "Methods" for Induction and Peirce integrated abduction and induction through the famous syllogistic framework where the two non–deductive inferences can be clearly distinguished: it has to be noted that Mill also said that what he called "institutions" rather than individuals are the real embodiment of "inductive logics". Following this Millian perspective Gabbay and Woods also add that it is typical of human individuals to function as *practical agents* and that it is typical of "institutions" to function as *theoretical agents* [8, p. 14]; moreover, agents tend toward enhancement of cognitive assets when this enables the achievement of cognitive goals previously unaffordable or anattainable. The ideal agents (logical and computational) I will describe in the following sections are theoretical agents, that *mimic* "institutions", in Millian sense, more than individuals' reasoning performances.

To clarify the process that underlies the formation of ideal inductive and abductive agents I have to briefly introduce in the following subsection the distinction between internal and external representations.

2 External and Internal Representations

2.1 Logic Programs as Agents: External Observations and Internal Knowledge Assimilation

As I will illustrate in the following subsection it is in the area of distributed cognition that the importance of the interplay between internal and external representations has recently acquired importance (cf. for example Clark [4] and Hutchins [14]). This perspective is particularly coherent with the agent–based framework I have introduced above, as we will see. It is interesting to note that a clear attention to the agent–based nature of cognition and to its interplay between internal and external aspects can be found in the area of logic programming. Indeed, logic programs can be seen in an agent-centered, computationally-oriented and purely syntactic perspective. Already in 1994 Kowalski [15] in "Logic without model theory" introduced a knowledge assimilation framework for rational abductive agents, to deal with incomplete information and limited computational capacity.

"Knowledge assimilation" is the assimilation of new information into a knowledge base, "as an alternative understanding of the way in which a knowledge base formulated in logic relates to externally generated input sentences that describe experience". The new pragmatic approach is based on a proof-theoretic assimilation of observational sentences into a knowledge base of sentences formulated in a language such as CL.[2] Kowalski proposes a pragmatic alternative view that contrasts with the model-theoretic approach to logic. In model theory notions such as *interpretation* and *semantic structures* dominate and are informed by the philosophical assumption that experience is caused by an independent existing "reality composed of individuals, functions and relations, separate from the syntax of language"

On the contrary logic programs can be seen as agents endowed with deductive databases considered as *theory presentations* from which logical consequences are derived, both in order to *internally* solve problems with the help of *theoretical sentences* and in order to assimilate new information from the *external* world of observations (*observational sentences*). The part of the knowledge base, which includes observational sentences and the theoretical sentences that are used to derive conclusions that can be compared with observations sentences, is called *world model*, considered a completely syntactic concept: "World models are tested by comparing the conclusions that can be derived from them with other sentences that record inputs, which are observational sentences extracted – *assimilated* – from experience". The agent might generate outputs – that are generated by some plan formation process in the context of the agents's "resident goals" – which affect its environment and which of course can affect its own and other agents' future inputs. Kowalski concludes "The agent will record the output, predict its expected effect on the environment using the 'world model' and compare its expectations against its later observations".

The epistemological consequence of this approach is fundamental: in model theory truth is a static correspondence between sentences and a given state of the world. In Kowalski's computational and "pragmatic" theory, the important is not the correspondence between language and experience, but the appropriate assimilation of an inevitable and continuous flowing input stream of "external" observational sentences into an ever changing "internal" knowledge base (of

[2] CL, computational logic, refers to the computational approach to logic that has proved to be fruitful for creating non–trivial applications in computing, artificial intelligence, and law.

course the fact that computational resources available are bounded suggests to the agent to make the best use of them, for instance avoiding redundant and irrelevant derivation of consequences). The correspondence (we can say the "mirroring") between an input sentence and a sentence that can be derived from the knowledge base is considered by Kowalski only a limiting case. Of course the agent might also generate its own hypothetical inputs, as in the case of abduction, induction, and theory formation.

The conceptual framework above, that is derived from a computationally–oriented logic approach that strongly contrasts with the traditional one in terms of model theory, is extremely interesting. It stresses the attention on the flowing interplay between internal and external representations/statements, so *epistemologically* establishing the importance of the agent–based character of cognition. In the following subsection I will illustrate that an analogous perspective is convenient also for depicting human beings' cognition so far as we are interested in studying its essential distributed dynamics.

2.2 Distributed Cognition in Organic Agents: External and Internal Representations

Even if we can say that a large portion of the complex environment of a thinking agent is internal, it is widely recognized that "human" cognitive systems are composed by distributed cognition among people and some "external" objects and technical artifacts (cf. for example Hutchins [14] and Norman [19]). It is the case of the human use of the construction of external diagrams in geometrical reasoning, useful to make observation and experiment to transform one cognitive state into another for example to discover new properties and theorems. Or the case of the use of the external representations based on the ordinary numeration system that eliminates some of the hardest parts of the addition or the difficult computations in multiplication when mentally performed. Mind is limited, both from a computational and an informational point of view: the act of delegating some aspects of cognition becomes necessary. In is in this sense that we can say that cognition is essentially multimodal.[3]

In addition, we can say that, adopting this perspective, we can give an account of the complexity of the whole human cognitive systems as the result of a complex interplay and *coevolution* of states of mind, body, and external environments suitably endowed with cognitive significance. An "agent-based" view aims at analyzing the features of "real" human thinking agents by recognizing the fact that a being-like-us agent functions "at two levels" and "in two ways". I define the two levels as *explicit* and *implicit* thinking. *Agent-based* perspective in logic has the power of recognizing the importance of both levels.

We maintain that representations are external and internal. We can say that

- *external representations* are formed by external materials that re-express (through reification) concepts and problems that are already present in the mind or concepts and problems that do not have a *natural home* in the brain;
- *internalized representations* are internal re-projections, a kind of recapitulations, (learning) of external representations in terms of neural patterns of activation in the brain. They can sometimes be "internally" manipulated like external objects and can originate new internal reconstructed representations through the neural activity of *transformation* and *integration*.

This process explains why human beings seem to perform both computations of a *connectionist* type such as the ones involving representations as

- (I Level) *patterns of neural activation* that arise as the result of the interaction between body and environment (and suitably shaped by the evolution and the individual history): pattern completion or image recognition,
  and computations that use representations as
- (II Level) *derived combinatorial syntax and semantics* dynamically shaped by the various external representations and reasoning devices found or constructed in the environment (for example geometrical diagrams); they are neurologically represented contingently as patterns of neural activations that "sometimes" tend to become stabilized structures and to fix and so *to permanently belong to the I Level* above.

The I Level originates those *sensations* (they constitute a kind of "face" we think the world has), that provide room for the II Level to reflect the structure of the environment, and, most important, that can follow the computations suggested by these external structures. It is clear we can now conclude that the growth of the brain and especially the synaptic and dendritic growth are profoundly determined by the environment.

When the fixation is reached the patterns of neural activation no longer need a direct stimulus from the environment for their construction. In a certain sense they can be viewed as *fixed internal records* of *external structures* that *can exist* also in the absence of such external structures. These patterns of neural activation that constitute the I Level Representations always keep record of the experience that generated them and, thus, always carry the II Level Representation associated to them, even if in a different form, the form of *memory* and not the form of a vivid sensorial experience. Now, the human agent, via neural mechanisms, can retrieve these II Level Representations and use them as *internal* representations or use parts of them to construct new internal representations very different from the ones stored in memory (cf. also [10]).

Human beings delegate cognitive features to external representations because in many problem solving situations the internal computation would be impossible or it would involve a very great effort because of human mind's limited capacity. First a kind of alienation is performed, second a recapitulation is accomplished at the neuronal level by re-representing internally that which was "discovered" outside. Consequently only later on we perform cognitive operations on the structure of data that synaptic patterns have "picked up" in an analogical way from the environment. We can maintain that internal representations used in cognitive processes like many events of *meaning creation* have a deep origin in the experience lived in the environment.

I think there are two kinds of artifacts that play the role of *external objects* (representations) active in this process of disembodiment of the mind: *creative* and *mimetic*. Mimetic external representations mirror concepts and problems that are already represented

[3] Thagard [20, 21] observes, that abductive inference can be visual as well as verbal, and consequently acknowledges the sentential, model–based, and manipulative nature of abduction we will illustrate below. Moreover, both data and hypotheses can be visually represented:

> For example, when I see a scratch along the side of my car, I can generate the mental image of grocery cart sliding into the car and producing the scratch. In this case both the target (the scratch) and the hypothesis (the collision) are visually represented. [. . .] It is an interesting question whether hypotheses can be represented using all sensory modalities. For vision the answer is obvious, as images and diagrams can clearly be used to represent events and structures that have causal effects [21].

Indeed hypotheses can be also represented using other sensory modalities.

in the brain and need to be enhanced, solved, further complicated, etc. so they sometimes can creatively give rise to new concepts and meanings, playing the role of creative representations.[4] Inductive and abductive *ideal agents* are mimetic artifacts in the sense I have just illustrated.

2.3 Internal, External, and Hybrid Inducers and Abducers

From the perspective I have illustrated in the previous section the expansion of the inductive and abductive minds typical of organic agents is in the meantime a continuous process of *externalization* of the minds themselves into the *material world* around them. In this regard the evolution of the mind is inextricably linked with the evolution of many kinds of large, integrated, material cognitive systems, like logical and computational systems. In the following I will illustrate some features of this extraordinary interplay between human brains and the *ideal cognitive systems* they make. We acknowledge that material artifacts like for example *inductive and abductive logical and computational agents* are tools for thoughts as is language: tools for exploring, expanding, and manipulating our own minds.

The two ways mentioned above are the *external way* and the *internal way*. In fact inductive and abductive logical and computational systems can be seen as external representations and tools expressed through artificial (in part mathematical) and ordinary language and the use of suitable artifacts. These ideal systems not only mirror and mimic the internal ways of inferring of the *being-like-us* reasoners we have illustrated above; they can also play a *creative* role. The activities of externalizing play a central role not just in mirroring the internal ways of thinking but also in finding room for concepts and new ways of inferring which cannot be found internally "in the mind".

In summary, organic agents like human beings are hasty generalizers and more or less naive inducers and abducers but are also the creators of sophisticated external cognitive representations that for example provide *demonstrative/deductive* and *computational* representations of those reasoning performances. The interplay between these "external" tools and the already "internalized" templates of reasoning certainly realizes a continuous improvement of the internal templates themselves but also expresses the centrality of the hybrid exploitation of both levels in reasoning.

Let us consider the case of abduction, I have indicated above that abduction appears to be a formal fallacy that can be recognized from the classical logic point of view: the fallacy of affirming the consequent. However, from the point of view of both everyday and scientific knowledge, abduction is an important kind of inference used to explain facts and invent hypotheses and theories [18].

Abduction is the process of *inferring* certain facts, laws and hypothesis that render some sentences plausible or explain/discover some eventually new phenomenon or observation. I have maintained elsewhere that, from the epistemological perspective, abduction has two main meanings: 1) abduction that only generate plausible hypotheses (selective or creative) and 2) abduction considered as "inference to the best explanation", which also evaluates hypotheses. I have introduced in [18] the concept of *theoretical abduction* as a form of neural and basically internal processing. I maintain that there are two kinds of theoretical abduction, "sentential", related to logic and to verbal/symbolic inferences, and "model-based", related to the exploitation of models such as diagrams, pictures, etc. Theoretical abduction certainly illustrates much of what is important in creative abductive reasoning, in humans and in computational programs, but fails to account for many cases of explanations occurring in science when the exploitation of external environment is crucial. It fails to account for those cases in which there is a kind of "discovering through doing", cases in which new and still unexpressed information is codified by means of manipulations of some external objects I have called *epistemic mediators* [18]. The concept of *manipulative abduction* (see below) captures a large part of hypothetical cognition where the role of action is central, and where the features of this action are often implicit and hard to be elicited. We can conclude, following Thagard [21] that abduction is a cognitive processes constitutively "multimodal" (cf. above footnote 3).

Abduction is of fundamental importance in many agent-based reasoning situations like scientific explanation, scientific discovery, and moral deliberation. We can furnish another reason that stresses the fruitfulness of abduction in agent-based reasoning: it is a powerful inferential process able to govern inconsistencies. For example, in the case of the formation of scientific theories epistemologists have recognized the role played by inconsistencies and anomalies that violate the paradigm–induced expectations derived from previously established conceptual frameworks. Logicians have in turn shown that inconsistencies generated by anomalies are difficult to be managed in deductive situations: they are unexpected facts that the rules of classical logic are not able to explain.

Hence, we can outline two different ways of thinking of abduction: 1) from the point of view of classical logic, abduction is a formal fallacy, not truth preserving; 2) from the point of view of epistemology, abduction is an important kind of reasoning able to discover new hypotheses and give explanation to scientific facts

In delineating the structure of a new agent-based perspective of logic Gabbay and Woods state that logic has to be considered an "account of how thinking agents reason and argue" [8, p. 1]. Their idea is that logic has to be defined as the disciplined description of the behavior of real–life of logical agents. Logic has to be thought of as an *agent-based logic*. From this viewpoint, abduction can be rendered as that kind of logical reasoning in which the fact of not being truth preserving (but *ignorance–preserving*, as they contend) has to live together with the fact that it is fruitfully used by real logical agents. In this framework induction is seen as *probability–enhancing* and deduction as *truth–preserving*.

To conclude, the use of abduction is good for at least two reasons. Abduction is not only a simple formal fallacy, but also a specific case of ignorance–preserving reasoning that can be fruitfully *idealized* in theoretical logical agents; on the applicative side, abduction is a good process able provide new hypothesis and govern inconsistencies.

At this point I hope it is clear that organic agents are spontaneous inducers and abducers and that they also construct logical and computational systems both able to *mimic* human inductions and abductions and to *create* new "rational" ways of inducing and abducing. These systems are in turn used by organic agents: they consequently have to be seen as *hybrid reasoners*. In the following section I will illustrate how what I called manipulative abduction can furnish a perfect example of this hybridity of human reasoning.

3 Manipulative Abduction and Hybrid Reasoning

I have introduced the concept of *manipulative abduction* - contrasted with theoretical abduction [18] - to illustrate situations where we are thinking through doing and not only, in a pragmatic sense, about do-

[4] Following this perspective it is at this point evident that the "mind" transcends the boundary of the individual and includes parts of that individual's environment.

ing.

First of all manipulative abduction is generally related to the suitable exploitation of external tools like logical and computational systems/agents[5] to the aim of generating desired hypotheses.

Second, in the case of the formation of scientific hypotheses the idea of manipulative abduction goes beyond the well–known role of experiments as capable of forming new scientific laws by means of the results (nature's answers to the investigator's question) they present, or of merely playing a predictive role (in confirmation and in falsification). Manipulative abduction refers to an extra-theoretical behavior that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (theoretical) practices.

In this sense the existence of this kind of extra-theoretical cognitive behavior is also testified by the many everyday situations in which humans are perfectly able to perform very efficacious (and habitual) tasks without the immediate possibility of realizing their conceptual explanation. In some cases the conceptual account for doing these things was at one point present in the memory, but now has deteriorated, and it is necessary to reproduce it, in other cases the account has to be constructed for the first time, like in creative settings of manipulative abduction in science.

Consequently we face with at least two cases of manipulative abduction.

1. The first one refers to the exploitation of external logical and computational abductive – but also inductive – systems/agents to form hybrid and multimodal representations and ways of inferring in organic agents. Doing this they are able to enhance their "rational" performances (see below subsection 3.1).
2. The second case refers to the role of manipulative abduction at the level of scientific experiment and of the so–called *thinking through doing* that in turn can improve our knowledge of induction, and its distinction from abduction: manipulative abduction can be considered as a kind of basis for further meaningful inductive generalizations .

Further preliminary observations have to be anticipated to favor the comprehension of the second case. Hutchins [14] illustrates the case of a navigation instructor that for 3 years performed an automatized task involving a complicated set of plotting manipulations and procedures. The insight concerning the conceptual relationships between relative and geographic motion came to him suddenly "as lay in his bunk one night". This example explains that many forms of learning can be represented as the result of the capability of giving conceptual and theoretical details to already automatized manipulative executions. The instructor does not discover anything new from the point of view of the objective knowledge about the involved skill, however, we can say that his conceptual awareness is new from the local perspective of his individuality.

In this kind of action-based abduction the suggested hypotheses are inherently ambiguous until articulated into configurations of real or imagined entities (images, models or concrete apparatus and instruments). In these cases only by experimenting we can discriminate between possibilities: they are articulated behaviorally and concretely by manipulations and then, increasingly, by words and pictures.

Gooding [11] refers to this kind of concrete manipulative reasoning when he illustrates the role in science of the so-called "construals" that embody tacit inferences in procedures that are often apparatus and machine based. They belong to the pre-verbal context of ostensive operations, that are practical, situational, and often made with help of words, visualizations, or concrete artifacts. The embodiment is of course an expert manipulation of objects in a highly constrained experimental environment, and is directed by abductive movements that imply the strategic application of old and new *templates* of behavior mainly connected with extra-theoretical components, for instance emotional, esthetical, ethical, and economic.

The hypothetical character of construals is clear: they can be developed to examine further chances, or discarded; they are provisional creative organization of experience and some of them become in their turn hypothetical *interpretations* of experience, that is more theory-oriented, their reference is gradually stabilized in terms of established observational practices. Step by step the new interpretation – that at the beginning is completely "practice-laden" – relates to more "theoretical" modes of understanding (narrative, visual, diagrammatic, symbolic, conceptual, simulative), closer to the constructive effects of theoretical abduction.

When the reference is stabilized the effects of incommensurability with other established observations can become evident. But it is just the construal of certain phenomena that can be shared by the sustainers of rival theories. Gooding [11] shows how Davy and Faraday could see the same attractive and repulsive actions at work in the phenomena they respectively produced; their discourse and practice as to the role of their construals of phenomena clearly demonstrate they did not inhabit different, incommensurable worlds in some cases. Moreover, the experience is constructed, reconstructed, and distributed across a social network of negotiations among the different scientists by means of construals.

These construals aim at arriving to a shared understanding overcoming all conceptual conflicts. As I said above they constitute a provisional creative organization of experience: when they become in their turn hypothetical interpretations of experience, that is more theory-oriented, their reference is gradually stabilized in terms of established and shared observational practices that also exhibit a cumulative character. It is in this way that scientists are able to communicate the new and unexpected information acquired by experiment and action.

3.1 Organic Hybrid Reasoners and External Semiotic Anchors

In the perspective we have illustrated above in section 2, resorting to the distinction between internal and external inducers and abducers a novel perspective on external ideal logical agents can be envisaged.

Starting from the low–level inferential performances of the kid's hasty generalization that is a strategic success and a cognitive failure human beings arrived to the externalization of "theoretical" inductive and abductive agents as *ideal agents*, logical and computational. It is in this way that *merely successful strategies* are replaced with *successful strategies* that also tell the "more precise truth" about things. These external representations can be usefully re-represented in our brains (if this is useful and possible), and they can originate new improved organic (mentally internal) ways of inferring or suitably exploited in a hybrid manipulative interplay, as I have said above.

From this perspective human beings are hardwired for survival and for truth alike so best inductive and abductive strategies can be built and made explicit, through self-correction and re-consideration (since for example the time of the inductive Mill's methods). Furthermore human beings are agents that can cognitively behave as *hybrid*

[5] I am referring here to systems – abstract or practical/computational – that are explicitly able to theoretically perform abductions and inductions in themselves thanks to their own knowledge bases, rules, and devices.

agents that exploit in reasoning both internal representations and externalized representations and tools, but also the mixture of the two.

Let 's consider the example of the externalization of some inferential skills in logical demonstrative systems, like for example the ones that are at the basis of logic programming.[6] They present interesting cognitive features (cf. also Longo [17]) which I believe deserve to be further analyzed and which can further develop the distinction above between theoretical and practical agents:

1. *symbolic*: they activate and semiotically "anchor" meanings in material communicative and intersubjective *mediators* in the framework of the phylogenetic, ontogenetic, and cultural reality of the human being and its language. It can be hypothesized these logical agents originated in embodied cognition, gestures, and manipulations of the environment we share with some mammals but also non mammal animals (cf. the case of monkeys' knots and pigeons' categorization, in [12]).[7]
2. *abstract*: they are based on a *maximal independence* regarding sensory modality; they strongly stabilize experience and common categorization. The maximality is especially important: it refers to their practical and historical invariance and stability;
3. *rigorous*: the rigor of proof is reached through a difficult practical experience. For instance, in the case of mathematics and logic, as the maximal place for convincing and sharable reasoning. Rigor lies in the stability of proofs and in the fact they can be iterated. Following this perspective mathematics is the best example of maximal stability and conceptual invariance. Logic is in turn a set of proof invariants, a set of structures that are preserved from one proof to another or which are preserved by proof transformations. As the externalization and result of a distilled praxis, the praxis of proof, it is made of maximally stable regularities;
4. I also say that a *maximization of memorylessness*[8] "variably" characterizes demonstrative reasoning. This is particularly tangible in the case of the vast idealization of classical logic and related approaches. The inferences described by classical logic do not yield sensitive information – so to say – about their real past life in human agents' use, contrarily to the "conceptual" – narrative – descriptions of human non-demonstrative processes, which variously involve "historical", "contextual", and "heuristic" memories. Indeed many thinking behaviors in human agents – for examples abductive inferences, especially in their generative part – are context-dependent. As already noted their *stories* vary with the multiple propositional relations the human agent finds in her environment and which she is able to take into account, and with various cognitive reasons to change her mind or to think in a different way, and with multiple motivations to deploy various tactics of argument. In this perspective Gabbay and Woods say:

   > Good reasoning is always good in relation to a goal or an agenda which may be tacit. [. . .] Reasoning validly is never *itself* a goal of good reasoning; otherwise one could always achieve it simply by repeating a premiss as conclusion, or by entering a new premiss that contradicts one already present. [. . .] It is that the reasoning actually performed by individual agents is sufficiently reliable not to kill them. It is reasoning that precludes neither security not prosperity. This is a fact of fundamental importance. It helps establish the fallibilist position that it is not unreasonable to pursue modes of reasoning that are known to be imperfect [8, pp. 19-20].

   As we have already illustrated in section 2.3 human agents, as practical agents, are hasty inducers and bad predictors, unlike ideal (logical and computational) agents. In conclusion, we can say abductive inferences in human agents have a memory, a story: consequently, an abductive ideal logical agent has to variably weaken many of the aspects of classical logic and to overcome the relative demonstrative limitations.

I think that a great contribution given to logic by Gabbay is the creation of the *labelled deductive systems* (and their application to the logic of abduction), where data is structured and labelled and different insertion policies can be formulated [7, 9]. The labelled deductive systems fulfill the request of weakening the rigidity of classical logic but also of many non standard logics strictly related to it, opening a new era in logic: the attention to the role of *meta-levels* – for instance in the logic of abduction – formalizes the flexibility and "historicity" of many kinds of human thinking which are meaningful in certain application areas they address. Gabbay and Woods' conclusion about psychologism is clear and leads to a new conception of logic:

> If [. . .] it is legitimate to regard logic as furnishing formal models of certain aspects of the cognitive behavior of logical agents, then not only do psychological considerations have a defensible place, they cannot reasonably be excluded [8, p. 2].[9]

We can conclude by stressing the fact that human non-demonstrative inferential processes of induction and abduction are more and more externalized and objectified at least in three ways:

1. through Turing's Universal Practical Computing Machines we can have running programs – often based on logic – that are able to mimic – and enhance – "the actions of a human computer very closely" [23], and so - amazingly - also those human agents' "actions" that correspond to the complicated inferential performances like abduction (cf. the whole area of artificial intelligence);
2. human non-demonstrative processes are more and more externalized and made available in forms of explicit narratives and learnable templates of behavior (cf. also the study of fallacies as important tools of the human "kit" that provides evolutionary advantages, in this sense a fallacy of the affirming the consequent –

[6] A survey on perspectives in logic programming about induction and abduction is given in Flach and Kakas [6], who also furnish a useful classical perspective on integration of abduction and induction. The following distinction is introduced between explanation and generalization:

- *explanation*: hypothesis does not refer to observables - already in the case of selective abduction, moreover, abduction also creates new hypotheses too;

- *generalization* - it is the introduction of a genuinely new hypothesis that in turn can entail additional observable information on unobserved individual, extending the theory T.

Imagine we have a new abductive theory $T' = T \cup H$ constructed by induction: an inductive extension of a theory can be viewed as a set of abductive extensions of the original theory T.

[7] Cf. also the cognitive analysis of the origin of the mathematical continuous line as a pre-conceptual invariant of three cognitive practices [22], and of the numeric line [3, 5, 1].

[8] I derive this expression from Leyton [16] that introduces a very interesting new geometry where forms are no longer memoryless like in classical approaches such as the Euclidean and the Kleinian in terms of groups of transformations.

[9] An analogous example of the new modeling flexibility of recent logic is represented by the work in the dynamic logics of reasoning of van Benthem [24]. This logic offers a distinction between inferences that are dependent on short term representations and those that depend on long-term memory, which involves the processing of representations of greater abstraction. In this way it is possible to formally and flexibly reproduce the interplay that occurs in human agents' thinking both at the level of short-term memory – more inclined to be damaged by inconsistencies – and at the level of the long-term memory, where inconsistencies can be inert.

which depicts abduction in classical logic – is better than nothing [25]).[10]

3. new demonstrative systems – ideal logical agents – are created able to model in a deductive way many non-demonstrative thinking processes, like abduction, analogy, creativity, spatial and visual reasoning, etc.[11]

4 Conclusion

I have described inductive and abductive reasoning in the light of the agent–based framework to the aim of clarifying their fallacious character and the role of their related ideal systems (logical and computational). In this perspective I have analyzed some inductive and abductive ways of reasoning that in the light of classical and informal logic are defined *fallacies*, showing the fact they can realize a kind of strategic "rationality". After having illustrated the distinction between internal and external representations in the tradition of both *logic programming* and *distributed reasoning*, I have described some important aspects of *manipulative abduction*. It can be interpreted as a form of practical reasoning a better understanding of which furnishes a description of human beings as *hybrid reasoners* to the extend that they are users of ideal and computational agents, for example devoted to perform sophisticated inductions and abductions.

References

[1] B. Butterworth, *The Mathematical Brain*, MacMillan, New York, 1999.

[2] C. Cellucci, 'Mathematical discourse vs. mathematical intuition', in *Mathematical Reasoning and Heuristics*, eds., C. Cellucci and D. Gillies, pp. 138–166, London, (2005). King's College Publications.

[3] G. Châtelet, *Les enjeux du mobile*, Seuil, Paris, 1993. English transl. by R. Shore and M. Zagha, *Figuring Space: Philosophy, Mathematics, and Physics*, Kluwer Academic Publishers, Dordrecht, 2000.

[4] A. Clark, *Natural-Born Cyborgs. Minds,Technologies, and the Future of Human Intelligence*, Oxford University Press, Oxford and New York, 2003.

[5] S. Dehaene, *The Number Sense*, Oxford University Press, Oxford, 1997.

[6] P. Flach and A. Kakas, eds. *Abductive and Inductive Reasoning: Essays on Their Relation and Integration*, Dordrecht, 2000. Kluwer Academic Publishers.

[7] D.M. Gabbay, 'Abduction in labelled deductive systems', in *Handook of Defeasible Reasoning and Uncertainty Management Systems*, eds., D.M. Gabbay and R. Kruse, pp. 99–153, Dordrecht, (2002). Kluwer Academic Publishers.

[8] D.M. Gabbay and J. Woods, *The Reach of Abduction*, North-Holland, Amsterdam, 2005. Volume 2 of *A Practical Logic of Cognitive Systems*.

[9] D.M. Gabbay and J. Woods, 'A formal model of abduction', in *Abduction and Creative Inferences in Science*, ed., L. Magnani, (2006). Special Issue of the *Logic Journal of IGPL*.

[10] A. Gatti and L. Magnani, 'On the representational role of the environment and on the cognitive nature of manipulations', in *Computing, Philosophy, and Cognition*, ed., L. Magnani, pp. 227–242, London, (2006).

[11] D. Gooding, *Experiment and the Making of Meaning*, Kluwer, Dordrecht, 1990.

[12] P. Grialou, G. Longo, and M. Okada, eds. *Images and Reasoning*, Tokyo, 2005. Keio University.

[13] H. H. Hansen, 'The straw thing of fallacy theory: the standard definition of "fallacy"', *Argumentation*, **16(2)**, 133–155, (2002).

[14] E. Hutchins, *Cognition in the Wild*, MIT Press, Cambridge, MA, 1995.

[15] R. Kowalski, 'Logic without Model Theory', in *What is a Logical System?*, ed., D. M. Gabbay, 35–71, Oxford University Press, (1994).

[16] M. Leyton, *A Generative Theory of Shape*, Springer, Berlin, 2001.

[17] G. Longo, 'The cognitive foundations of mathematics: human gestures in proofs and mathematical incompleteness of formalisms', in *Images and Reasoning*, eds., P. Grialou, G. Longo, and M. Okada, pp. 105–134, Tokyo, (2005). Keio University.

[18] L. Magnani, *Abduction, Reason, and Science. Processes of Discovery and Explanation*, Kluwer Academic/Plenum Publishers, New York, 2001.

[19] D.A. Norman, *Things that Make Us Smart. Defending Human Attributes in the Age of the Machine*, Addison-Wesley, Reading, MA, 1993.

[20] P. Thagard, 'How does the brain form hypotheses? Towards a neurologically realistic computational model of explanation', in *Symposium "Generating explanatory hypotheses: mind, computer, brain, and world"*, eds., P. Thagard, P. Langley, L. Magnani, and C. Shunn, Stresa, Italy, (2005). Cognitive Science Society, CD-Rom. Proceedings of the 27th International Cognitive Science Conference.

[21] P. Thagard, 'Abductive inference: from philosophical analysis to neural mechanisms', in *Inductive Reasoning: Cognitive, Mathematical, and Neuroscientific Approaches*, eds., A. Feeney and E. Heit, Cambridge, (2006). Cambridge University Press. Forthcoming.

[22] B. Theissier, 'Protomathematics, perception and the meaning of mathematical objects', in *Images and Reasoning*, eds., P. Grialou, G. Longo, and M. Okada, pp. 135–45, Tokyo, (2005). Keio University.

[23] A.M. Turing, 'Computing machinery and intelligence', *Mind*, **49**, 433–460, (1950).

[24] J. van Benthem, *Exploring Logical Dynamics*, CSLI Publications, Stanford, CA, 1996.

[25] J. Woods, *The Death of Argument*, Kluwer Academic Publishers, Dordrecht, 2004.

[26] J. Woods, 'Epistemic bubbles', in *We Will Show Them: Essays in Honour of Dov Gabbay*, eds., S. Artemov, H. Barringer, A. Garcez, L. Lamb, and J. Woods, volume II, 731–774, College Publications, London, (2005).

[10] Cf. also Gabbay and Woods [8, pp. 33-36].

[11] A skeptical conclusion about the superiority of demonstrative over non-demonstrative reasoning is provided by the following philosophical argumentation of Cellucci [2] I agree with, which seems to emphasize the role of *ignorance preservation* in logic: "To know whether an argument is demonstrative one must know whether its premises are true. But knowing whether they are true is generally impossible", as Gödel teaches. So they have the same status of the premises of non-demonstrative reasoning. Moreover: demonstrative reasoning cannot be more cogent than the premises from which its starts; the justification of deductive inferences in any absolute sense is impossible, they can be justified as much, or as little, as non-deductive - ampliative - inferences. Also checking soundness is a problem.

Abduction, Preduction and the fallible way of modelling nature

some epistemological consequences for the philosophy of physics [1]

Andrés Rivadulla[2]

1 Introduction

Since the very beginning of the methodology of science 2400 years ago, philosophers have been trying different ways of scientific discovery. Abduction and induction belong to the best known ones. Abduction was Plato's *ars inveniendi*, whereas induction was Aristotle's method for the discovery of the principles of science. Plato's way was abductive, because it was conceived of to propose hypotheses (geometrical models) intended to save the appearances presented by movements of the planets, as observed from the Earth. Aristotle's inductive method on his side has caused a big trouble in the history of Western philosophy. It was known that inductive inferences could be false. But, until the consolidation of the hypothetic-deductive method in the contemporary philosophy of science after Einstein and Popper, there was available no alternative method of scientific discovery that could replace it. Thus it was assumed that induction was a fallible way of dealing with Nature. So was conceived of abduction as well, since as Peirce (C.P. 2.776 and 2.777) acknowledged, abducted hypotheses are frequently wrong.

However my main contribution will be to point to a way of reasoning, very common in theoretical physics, but which has not yet attracted the attention of the philosophers of science. I call it *preduction*[3]. It consists in a form or reasoning that starts from first principles, *methodologically* postulated as premises of the inferential procedure. These premises can proceed from different theories. Preductive reasoning differs from abduction in that the hypotheses are not suggested by data, but constructed on the basis of the available theoretical background. Thus it depends more on the theoretical framework than on the empirical data, and it is an implementation of the hypothetical-deductive method. But it should not be confused with the axiomatic-deductive method. Preduction provides indeed the method by which most theoretical models are postulated in science. But since preduced models depend on the available theoretical background, and this cannot be known to be true, restrictions do frequently occur in the domain of their intended applications. Thus preduction only offers a fallible way of dealing with Nature as well.

In the following section I present some cases of study of both extensions and restrictions in the fields of Newtonian mechanics and classical statistical mechanics. My point is that the existence of both extensions and restrictions of the application domain of a preduced theoretical model can be taken as an argument on behalf of an antirealist viewpoint in the philosophy of physics: Unless one wants to immunize a theoretical construct against potential falsifiers, the extension of the domain of intended applications of a theoretical model cannot be used to claim either its approximation to the truth nor its probability to be true; moreover since a priori we have no reason to be suspicious about what does not count as one of the intended applications of a model, any a posteriori commitment to restrict its application domain cannot impel us to claim that it has been falsified. In other words: when a theoretical model has been preduced, its domain of intended applications is completely open. In scientific methodology there is no algorithm for the postulation of theoretical models, and both abductive and preductive reasoning are allowed. Both provide us with hypotheses that serve as premises for further inferences and empirical predictions. Scientific *ars inveniendi* is not submitted to rules. As a consequence, we cannot foresee how many phenomena will in the future be considered to belong to the application domain of the postulated theoretical constructs, nor how many will not or will have to be removed from it. It is therefore reasonable to assume that theoretical models are nothing but instruments intended merely to deal predictively with Nature. Inference to the best explanation goes without saying. But it does not mean inference to the true, or approximately true, or probably true explanation. For theory is not the space of truth.

2 *Domain revisions* in theoretical physics

As a particular form of the hypothetical-deductive method, preduction provides, on the basis of previously accepted theoretical constructs, the means of the postulation of further theoretical models. The confirmation or the empirical rejection of these models allows us to talk respectively about the domain extension or the domain restriction of theoretical models. Following Theo Kuipers (2006) domain extension and domain restriction are the two forms of the revision of the domain of intented applications of a theoretical construct. In the following I present some examples of domain extension and domain restriction in the methodology of physics as part of an argument intended to support an anti-realist viewpoint in the philosophy of physics[4].

2.1 *Domain extensions* in classical physics

Example 1 - Extensions of the domain of intended applications of the *celestial Newtonian model*

[1] This paper is part of a research on *Theoretical Models in Physics* supported by the Spanish Ministry of Education and Science.

[2] Universidad Complutense, Facultad de Filosofía, Dpto. de Lógica y Filosofía de la Ciencia, E-28040 Madrid, email: arivadulla@filos.ucm.es

[3] As I have been informed by two anonymous referees of this paper, this word has been used already by Jum Arima and by Allen Courtney and Norman Foo (in a different sense) in the domain of artificial intelligence. I thank both referees for further helpful criticisms on this paper.

[4] Cfr. also Rivadulla (2006).

Beside the so-called *paradigmatic* intended applications, *unexpected* applications of Newtonian mechanics are: the computation of star and planet masses, the existence of collapsed stars as well as the stability of stars, the light deflection by the sun, the critical density of the Universe, etc.

Example 2 - Extension of the domain of intended applications of *classical statistical mechanics*: the *Jeans' mass limit model* for star formation

James Jeans (1877-1946) investigated the conditions under which a molecular cloud composed of N molecules would collapse to form a star. The *Theorem of Equipartition of Energy* of classical statistical mechanics claims that the kinetic energy of the cloud is

$$E_c = \frac{3}{2} N k_B T$$

where k_B is Boltzmann's constant, or

$$E_c = \frac{3}{2} \frac{M}{\mu m_H} k_B T$$

expressing N in terms of the average molecular weight and hydrogen mass.

On the other hand, the *virial theorem*, applied to systems composed by many objects, claims that the average potential gravitational energy of the constituent objects is two times their average kinetic energy. Since the expression of the potential gravitational energy is

$$V_g \approx -\frac{3}{5} G_N \frac{M^2}{R}$$

then,

$$2\frac{3}{2} \frac{M}{\mu m_H} k_B T = \frac{3}{5} G_N \frac{M^2}{R}$$

M and R denoting here the mass and the radius of the molecular cloud respectively.

Since in terms of the density $\rho_0 = \frac{4}{3} \frac{M}{\pi R^3}$ of the cloud, assumed to be constant,

$R = \left(\frac{3M}{4\pi\rho_0}\right)^{1/3}$, we obtain Jeans' critical mass value:

$$M_J = \left(\frac{5 k_B T}{G_N \mu m_H}\right)^{3/2} \left(\frac{3}{4\pi\rho_0}\right)^{1/2}$$

to be overcome in order that the collapse takes place, i.e. $M > M_J$.

2.2 Domain restrictions in classical physics

Example 1 - Restriction of the domain of intended applications of the *Newtonian model*: The *Kelvin-Helmholtz gravitational collapse model*

What is the source of the energy of stars? According to Arthur Eddington (1930, p. 289), Helmholtz-Kelvin's gravitational contraction hypothesis

> *Supposes that the [star energy] supply is maintained by the conversion of gravitational energy into heat owing to the gradual contraction of the star.*

This idea was put forward by Hermann von Helmholtz in a lecture given in Königsberg on February the 7th, 1854, in occasion of the 50th anniversary of Immanuel Kant's death. Twelve years later Lord Kelvin retook this idea in "On the Age of the Sun's Heat". According to Kelvin (1903, pp. 493-494) Helmholtz's *meteoric theory*

> *Consists in supposing the sun and his heat to have originated in a coalition of smaller bodies, falling together by mutual gravitation, and generating, as they must do according to the great law demonstrated by Joule, an exact equivalent of heat for the motion lost in collision.*

He claims

> *That some form of the meteoric theory is certainly the true.*

In order to analyse the viability of this hypothesis I resort to A. Ostlie & D. Carroll (1996, p. 329): Since the total mechanical energy of a star in equilibrium is

$$E \approx -\frac{3}{10} G_N \frac{M^2}{R}$$

(i.e., the half of its potential energy, according to the *virial theorem*) in the case of our Sun the amount of gravitational energy liberated during his 'collapse' until today would be

$$E_g \approx 1.1 \times 10^{48} erg.$$

Assuming a constant *luminosity* during the Sun's whole life - given that luminosity is power, i.e. energy per time unity - the Sun's age should be

$$t = \frac{E_g}{l} \approx 10^7 \text{ years.}$$

This age is bizarrely short. As Eddington (1930, p. 290) claims

> *Biological, geological, physical and astronomical arguments all lead to the conclusion that this age is much too low and that the time-scale given by the contraction hypothesis must somehow be extended.*

Example 2 - Restriction of the domain of intended applications of classical statistical mechanics: The *Rayleigh-Jeans radiation model for the black body*

In 1900 Lord Rayleigh and James Jeans applied the *Theorem of Equipartition of Energy* of classical statistical mechanics, according to which the average kinetic energy of a particle of a system in thermodynamic equilibrium is $\frac{1}{2} k_B T$, to the calculation of the energy density of a gas of photons inside a receptacle in thermal equilibrium which contains N_ν stationary electromagnetic waves with frequencies in the interval $\nu, \nu + d\nu$ and average energy $k_B T$, resulting of the sum of the corresponding energies of the electric and magnetic fields. Thus the total value of the energy would be $E = N k_B T$. Since the energy density is independent of the geometry and material constitution of the receptacle, and it can be deduced that the number of *radiation modes* is $N_\nu = \frac{8\pi\nu^2}{c^3}$, then it suffices to multiply N_ν by $k_B T$ in order to determine that the energy density emitted by a black body is:

$$E(\nu, T) = \frac{8\pi\nu^2}{c^3} k_B T$$

which is known as *Rayleigh-Jeans radiation law*.

The problem with this expression is that integrating over all frequencies:

$$E = \frac{8\pi k_B T}{c^3} \int_0^\infty \nu^2 d\nu = \infty$$

thus contradicting experience.

Rayleigh-Jeans radiation law failure is due to the application of classical statistical mechanics to the domain of photon gases, which seems to be the proper application domain of quantum statistical mechanics of Bose-Einstein, from which Max Planck's radiation law mathematically follows. Indeed, the extension of the application domain of classical statistical mechanics to the study of radiation leads to the *ultraviolet catastrophe*, which is how Paul Ehrenfest called *Rayleigh-Jeans radiation law* failure.

3 Conclusion: Some reasons on behalf of an antirrealist viewpoint in the philosophy of physics

Any successful application of a scientific hypothesis does not have any repercussions on its truth or on its probability. The actual extension of the application domain of a theoretical model maintains the doors open to their empirical rejection or to their application restriction to further phenomena.

Anyhow, the restriction cases of the application domain of classical physics shown above do not commit to the revision of the theory, i. e.: from a domain restriction does not follow ipso facto the empirical refutation of the theoretical model. It merely points to the fact that not every previously accepted hypothesis can be successfully applied to any possible novel question posed either by Nature or by science.

Only a posteriori can we recognize the inapplicability of a hypothesis to a given domain. A priori we cannot suspect about what does not count as one of its intended applications.

Thus we have reached following conclusion: Neither does domain extension verify, approximate to the truth or increase the truth probability of a theoretical model, nor does restriction refute it ipso facto. I see in this double fact a good reason to take theoretical models merely as tools to deal predictively with Nature.

Any case although restriction does not amount to empirical refutation, it would be philosophically uninteresting to pursue immunization strategies leading to a complete determination of the application domain of theoretical constructs.

References

Eddington, A. (1930), The *Internal Constitution of the Stars*, University Press, Cambridge

Kelvin, Lord (1903): "On the Age of the Sun's Heat", *Macmillan's Magazine*, March 1862. Reprinted as Appendix E of his *Treatise on Natural Philosophy*, University Press, 1867, 1903, Cambridge.

Kuipers, T. (2006): "Theories Looking for Domains. Fact or Fiction?. Structuralist Truth Approximation by Revision of the Domain of Intended Applications", MBR, Pavia, 2004. In L. Magnani (ed.), *Model-Based Reasoning in Science and Engineering* (forthcoming).

Ostlie, A. & D. Carroll (1996), *An Introduction to Modern Astrophysics*, Addison-Wesley, Reading, Mass.

Rivadulla, A. (2006): "The Role of Theoretical Models in the Methodology of Physics". In L. Magnani (ed.), *Model-Based Reasoning in Science and Engineering* (forthcoming).

Abstraction, Induction and Abduction in Scientific Modelling

Demetris Portides[1]

1 Introduction

The development of scientific knowledge consists in two major components. The first component involves the construction of the calculus of a theory, that I choose to refer to as 'theory formulation', and the second involves the attempt to relate this calculus to experimental reports, that I choose to refer to as 'theory application'. Distinguishing the two is, in my view, important and useful both epistemologically and methodologically.

Philosophers of science, notably [10, 1, 6, 5, 8, 3, 4], have explicitly recognised that theory formulation involves the conceptual processes of abstraction and idealisation. Suppes' view is couched in the jargon of the Semantic Conception of scientific theories, but without committing to the latter we could still make use of his general idea, which could be spelled out as follows. Assuming that we begin with the universe of discourse, by selecting a small number of variables and parameters abstracted from the phenomena we are able to formulate what we generally refer to as the general laws of a theory. For example, in classical mechanics we select position and momentum and establish a relation amongst the two variables, which we call Newton's second law or Hamilton's equations. By abstracting a set of parameters we thus create a sub-domain of the universe of discourse, which we call the domain of a scientific theory. Thus, Newton's laws signify a conceptual object of study that we call the domain of classical mechanics. Similarly Maxwell's equations signify the domain of classical electromagnetism, the Schrödinger equation signifies the domain of quantum theory, and so forth. Scientific domains, viewed from this perspective, are clearly distinct from physical domains, which they could represent only if they are expanded by or integrated with other conceptual resources (see [9]. Hence theory formulation abstracts a scientific domain from the universe of discourse and thus groups together different phenomena based on the particular aspects dictated by the particular domain.

In all the above general laws something is left unspecified: the force function in Newton's 2nd law, the electric and magnetic field vectors in Maxwell's equations, and the Hamiltonian operator in the Schrödinger equation. Scientific methodology demands that these are specified in order to establish a link between the assertions of the theory and physical systems. The theory application component enters in the process of specifying those elements of scientific theories that need to be filled-out if the theoretical assertions are to be linked to empirical phenomena, such as force functions, electric and magnetic field vectors, Hamiltonian operators, etc. The aim of these specifications are not to extend the theoretical assertions all the way to phenomena but it is to construct a model that resembles as many of the features of its target physical system. My aim in this paper is to suggest a meta-algorithm that captures the ways by which we specify force functions, Hamiltonian operators, etc. To be more precise, my attempt is to establish a logical framework (i.e. to rationally reconstruct) that captures the ways by which scientific models are constructed for the representation of physical systems.

The process of specification can be understood to involve two distinct aspects, both of which, each in its own way, play a crucial role in improving the accuracy or the representational capacity of the model. The first aspect concerns the question of how the degree of resemblance of a model to its target physical system is increased. This aspect comprises in the amalgamation inside the model of different descriptions about different aspects of the physical system, so that a more detailed and refined representation of the former is achieved. Let me refer to this aspect as the process of concretisation (or de-idealisation). The second aspect involves discovering (or inventing) the different descriptions that enter in the process of concretisation. It is, I claim, in the latter aspect of model construction that induction and abduction are vital.

2 A Reconstruction of Modelling Processes

In trying to use the theoretical assertions to model physical systems, we usually start from a highly abstract description of an ideal-type, which we attempt to concretise by reintroducing into the description all the abstracted features. Concretisation may involve a careful study of the physical system in question and of all its peculiarities and it is something that often takes an entire scientific research program to achieve (e.g. the structure of the nucleus research program). What is important in my discussion is the question of how the theory-dictated 'primary' description of a physical system is supplemented by what within the theory is considered of 'secondary' importance. Concretisation is involved at three levels, firstly in distinguishing what factors are necessary for achieving an acceptable representation of the physical system, I refer to these as the primary factors of the theoretical description. Secondly, what factors are required in bringing every individual primary term of the description closer to reality, *as if it functions alone*. And thirdly what is required in bringing closer to reality the interacting terms, thus compensating for the assumption that the separate terms are disjoint and autonomous. The logical schema I want to suggest, to capture this thought process, is a multi-dimensional improvement of Nowak's 1980 account [8]. Nowak's idealisation account was meant to capture the logic of theories in the social sciences and economics. I believe that the complexities involved in the physical sciences, especially in the application of Quantum Mechanics, require the multi-dimensional more generalized account that I urge, and that

[1] Department of Classics and Philosophy, University of Cyprus, Nicosia, Cyprus, email: portides@ucy.ac.cy

could be formulated as follows:

$T^{\alpha\beta}:\ If\ R(x)\ and\ S_{11}(x)=0,\ldots,S_{\alpha\beta}(x)=0,$
$and\ if\ P_{m1}(x),\ldots,P_{mn}(x)\ act\ on\ the\ physical\ system$
$autonomously\ from\ P_{k1}(x),\ldots,P_{kl}(x),\ then$
$H(x)=f_1(P_{11}(x),\ldots,P_{1\gamma}(x))+f_2(P_{21}(x),\ldots,P_{2\xi}(x))+$
$\ldots+f_\delta(P_{\delta 1}(x),\ldots,P_{\delta\epsilon}(x))$

The statement $T^{\alpha\beta}$ says that in a realistic description $R(x)$ of a physical system we abstract in two distinct ways. Firstly we abstract by categorising the factors of influence into primary, $P's$, and secondary, $S's$, and by subtracting all the secondary factors of influence from our initial theoretical description (i.e. by assuming that they do not act on the system in question). Secondly we abstract by grouping the primary factors into separate terms, $f_i's$, each of which is assumed to act autonomously in the physical system, and by categorising the secondary factors into their corresponding groups. Each f_i represents a mathematical function of different primary and secondary factors of influence, and the subscripts (indices) are only meant to state distinctions between different factors and groupings among factors. For instance, f_1 is a function conceptually distinct from f_2 because the influencing factors of which it is a function are assumed to act autonomously on the physical system from the respective factors of which f_2 is a function. Also, each P_{ij} (or S_{ij}) are indexed so that the modelling assumption that each factor of influence can be described distinctly from other factors is captured in the logical schema. The first index in the primary (and secondary) factors refers to the grouping to which the factor belongs and the second index is its name. The overall model description is represented by H, which is the sum of mathematical terms each of which is functionally related only to different primary factors of influence. The step-by-step process of concretisation of our hypothesis, that would improve the representational capacity of our model, involves the gradual addition of the secondary factors related with each and every one of the individual primary terms. A first step concretisation would be the following:

$T^{\alpha\beta-1}:\ If\ R(x)\ and\ S_{11}(x)=0,\ldots,S_{\alpha\beta-1}(x)=0,$
$and\ S_{\alpha\beta}(x)\neq 0,and\ if\ P_{m1}(x),\ldots,P_{mn}(x)$
$act\ on\ the\ physical\ system\ autonomously\ from$
$P_{k1}(x),\ldots,P_{kl}(x),\ then\ H(x)=f_1(P_{11}(x),\ldots,P_{1\gamma}(x))+$
$\ldots+g_{\alpha\beta-1}[f_\alpha(P_{\alpha 1}(x),\ldots,P_{\alpha\eta}(x)),\ h_{\alpha\beta}(S_{\alpha\beta}(x))]+$
$\ldots+f_\delta(P_{\delta 1}(x),\ldots,P_{\delta\epsilon}(x))$

Where, I have added the influence of just one secondary factor ($S_{\alpha\beta}$) in just one of the g_{ij} terms (namely, $g_{\alpha\beta-1}$). The g_{ij} terms are simply new names to the grouping-function that is altered by the introduction of one secondary function of influence, the first index i signifies the name of the grouping and the second index j signifies the number of factors introduced into the particular grouping. The h_{ij} terms are the names of the mathematical expressions through which the secondary factors of influence are represented. The addition of just one secondary factor of influence into the logical schema goes only to show that concretisation factors are added only to individual primary terms, it does not portray the actual practice in science, where concretisation factors may be added simultaneously or after significant theoretical and experimental developments. It must be noted that this logical schema allows for the regrouping of the terms in a description, as well as for the introduction of new terms as correction factors or as addenda. In other words, it allows for radical improvements to representational models in a particular physical domain that usually come about after a breakthrough is accomplished. A final concretised assertion would have the following form:

$T^{00}:\ If\ R(x)\ and\ S_{11}(x)\neq 0,\ldots,S_{\alpha\beta}(x)\neq 0,$
$and\ if\ P_{m1}(x),\ldots,P_{mn}(x)\ act\ on\ the\ physical\ system$
$autonomously\ from\ P_{k1}(x),\ldots,P_{kl}(x),\ then\ H(x)=$
$g_{10}[f_1(P_{11}(x),\ldots,P_{1\gamma}(x)),\ h_{11}(S_{11}(x)),\ldots,h_{1\theta}(S_{1\theta}(x))]+$
$g_{20}[f_2(P_{21}(x),\ldots,P_{2\psi}(x)),\ h_{21}(S_{21}(x)),\ldots,h_{2\chi}(S_{2\chi}(x))]$
$+\ldots+$
$g_{\delta 0}[f_\delta(P_{\delta 1}(x),\ldots,P_{\delta\epsilon}(x)),\ h_{\delta 1}(S_{\delta 1}(x)),\ldots,h_{\delta\phi}(S_{\delta\phi}(x))]$

The final statement T^{00} says that in a theoretical description of a physical system, in which all known factors of influence that were initially abstracted from the realistic description R are now reintroduced, we have an expression that breaks down the impact of all influencing factors into several terms each of which is assumed to act autonomously in the physical system. I believe that this account captures well the construction process of many applications of Classical and Quantum Mechanics. It also sheds some light on how representational models relate to the theory (a task that is beyond the scope of the present work). Moreover, it explicates one other important element of scientific model construction. Each different term of the description carries its own separate, and frequently independent, assumptions, which is a much more accurate understanding of scientific practice than regarding all as assumptions bound to the overall model description.

The claim I want to urge is that inductive and abductive procedures are operative in discovering (or inventing) how each term in the overall description is to be represented. That is to say, that we need either inductive or abductive arguments in order to justify the introduction of the individual terms $P'_{ij}s$ and $S'_{ij}s$ in the logical schema above, but that such arguments on their own do not justify the overall model description H, the latter is something that is determined by the process of abstraction/idealisation and its converse process of concretisation. In other words, induction and abduction are processes that piggyback on the processes of abstraction/idealisation and concretisation. I will proceed to briefly sketch two examples that can help visualize the above modelling process and distinguish its two aspects.

3 Scientific Modelling from the Viewpoint of the Concretisation Logical Schema

The simple pendulum is probably one of the most successful scientific representations in the history of science. To model the actual pendulum apparatus we start by assuming a mass-point bob supported by a massless inextensible cord of length l performing infinitesimal oscillations about an equilibrium point. Thus the equation of motion of the simple harmonic oscillator can be used as the starting point for modelling a real pendulum and thus attempting to measure the acceleration due to the Earth's gravitational field: $\theta''+(g/l)\theta=0$. But the idealised assumptions underlying this model equation, do not describe how the apparatus is in the world but they dictate an ideal description of the apparatus. Hence it is obvious to physicists that if a reasonably accurate representation is demanded, the various influencing factors of the pendulum motion must be incorporated into the model. This is not something peculiar to the pendulum but it is a demand that is present in the majority of cases of modelling physical systems.

In the pendulum example a reasonably accurate representational model would involve the following influencing factors: (i) finite amplitude, (ii) finite radius of bob, (iii) mass of ring, (iv) mass of

cap, (v) mass of cap screw, (vi) mass of wire, (vii) flexibility of wire, (viii) rotation of bob, (ix) double pendulum, (x) buoyancy, (xi) linear damping, (xii) quadratic damping, (xiii) decay of finite amplitude, (xiv) added mass, (xv) stretching of wire, (xvi) motion of support. To increase the degree of resemblance of the model to the pendulum apparatus mathematical descriptions of these factors are introduced into the model equation in a cumulative manner. Hence the aspects of modelling that were discerned above, i.e. concretisation, induction, and abduction, are clearly discerned in the pendulum case. To identify these influencing factors and to decide how they must be introduced into the model is a clear demonstration of what I have labelled the process of concretisation. In fact the above logical schema applies to the model of the pendulum in its most abstract and idealised form as follows:

$T^{\alpha\beta}: \; If \; R(x) \; and \; S_{11}(x) = 0, \ldots, S_{1\beta}(x) = 0, \; then$
$H(x) = f_1(P_{11}(x)$

In this simple form the schema suggests that only one primary factor of influence is identified (that of the linear restoring force due to gravity), and all secondary factors of influence are corrections to the influence of gravity. Where $H(x) = f_1(P_{11}(x)$ is a metalinguistic description of the Newtonian equation of motion of the simple harmonic oscillator $\theta'' + (g/l)\theta = 0$, that is meant to model the pendulum at a high degree of idealization and abstraction.

To discover what descriptions must be used for each of the secondary influencing factors is a clear demonstration of either an induction or an abduction process (The modelling details of the real pendulum apparatus can be found in [7]). Here is a case of an abductive procedure in determining how the air resistance acts on the oscillating system (pendulum bob and wire) to cause the amplitude to decrease with time and to increase the period. The Reynolds number for each component of the system determines the law of force for that component. The drag force is hence expressed in terms of a dimensionless drag coefficient, which is a function of the Reynolds number. In the pendulum case it can be argued abductively that a quadratic force law should apply for the pendulum bob, whereas a linear force law should apply for the pendulum wire (both of these are clearly inferences to the best explanation). Hence, it makes sense to establish a damping force which is a combination of linear and quadratic velocity terms: $F = b\,|v| + cv^2$. To determine the physical damping constants b and c the work-energy theorem is employed, an appropriate velocity function $v = f(\theta_0, t)$ is assumed, and under the assumption of conservation of energy they are matched to experimental results. The final expression of the effect of air damping is introduced into the equation of motion of the model.

Here is a case based on an inductive procedure in determining how the length of the pendulum is increased by stretching of the wire due to the weight of the bob. By Hooke's law (which, being an empirical law, could be claimed that it is arrived at inductively) when the pendulum is suspended in a static position the increase is $\Delta l = mgl_0/ES$, where S is the cross-sectional area and E is the elastic modulus. The dynamic stretching when the pendulum is oscillating is due to the apparent centrifugal and Coriolis forces acting on the bob during the motion. This feature is modelled by analogy with the spring-pendulum system to the near stiff limit. When these features are introduced into the model equation it gives rise to a system of coupled equations of motion.

A more complicated modelling example is that of the nuclear unified model used in the representation of the nuclear structure [2]. The unified model is based on a highly complex hypothesis about the nature of the nucleus, which expresses our conception of the nuclear structure as it has been shaped by the successes and failures of predecessor models. The hypothesis asserts that the nucleus is a complex system of a collection of particles that exhibit some form of independent nucleon motion, but that this motion is constrained by a slow collective motion of a core of nucleons, and that the two modes of motion interact with each other. In addition it asserts that the collective mode of motion is constituted by three distinct kinds of motion (vibration, rotation and giant resonance), two of which demonstrate an interaction mode. These ideas are expressed in the formalism of Quantum Mechanics in terms of the Hamiltonian operator of the unified model that is used in the Schrdinger equation for the nucleus. This Hamiltonian operator takes the following form: $H_{TOT} = H_{SP} + H_{COL} + H_{INT}$. Where H_{SP} is the single-particle Hamiltonian term, H_{INT} is the interaction mode Hamiltonian term, and the collective Hamiltonian is divided into four distinct modes of motion: $H_{COL} = H_{ROT} + H_{VIB} + H_{ROT-VIB} + H_{GR}$. Each of these terms are, of course, constituted by complex expressions that represent the various factors involved in each particular mode of nuclear motion. Since there are six primary terms, the above logical schema applies to the unified model in its most abstract and idealised form as follows:

$T^{\alpha\beta}: \; If \; R(x) \; and \; S_{11}(x) = 0, \ldots, S_{\alpha\beta}(x) = 0,$
$and \; if \; P_{m1}(x), \ldots, P_{mn}(x) \; act \; on \; the \; physical \; system$
$autonomously \; from \; P_{k1}(x), \ldots, P_{kl}(x), \; then$
$H(x) = f_1(P_{11}(x), \ldots, P_{1\gamma}(x)) + f_2(P_{21}(x), \ldots, P_{2\xi}(x)) +$
$\ldots + f_6(P_{61}(x), \ldots, P_{6\epsilon}(x))$

Where $H(x) = f_1(P_{11}(x), \ldots, P_{1\gamma}(x)) + f_2(P_{21}(x), \ldots, P_{2\xi}(x)) + \ldots + f_6(P_{61}(x), \ldots, P_{6\epsilon}(x))$ is a metalinguistic description of the total Hamiltonian operator of the unified model of nuclear structure, i.e. $H_{TOT} = H_{SP} + H_{ROT} + H_{VIB} + H_{ROT-VIB} + H_{GR} + H_{INT}$.

The unified model is an example that demonstrates two fundamental elements of model construction in the application of quantum mechanics. Firstly, in the case of the unified model the hypothesis of the model is not asserted in a highly abstract form. It involves many of the significant features of the nuclear structure that are present in our description of the physical system. Nevertheless, in specifying a Hamiltonian we abstract by dividing these features into three separate terms, as if their contribution to the behaviour of the nucleus is distinct and autonomous. This procedure is very frequent in modelling in physics, but we must recognise that it is only a conceptual division. The three terms in the unified model Hamiltonian are not meant to act disjointedly nor to represent separately, we impel the division by abstracting. The abstraction involved is the foundation of the counterfactual assertion, implied by the Hamiltonian, that the overall nuclear motion is $as \; if$ it receives contributions from distinct and autonomous modes of motion. This way by which abstraction is used in our modelling is reflected in the above logical schema of model construction.

Secondly, the individual Hamiltonian terms of the model are not constructed in identical ways. The H_{SP} term is modelled by using the principles of Quantum Mechanics from the outset in a systematic manner, i.e. by using a stock model of the theory and postulating ways by which to concretise the abstractions involved. The collective motion terms, however, differ significantly in the method of construction. In fact the collective terms are first set up as if the system behaves in accordance to classical mechanics and at some appropriate stage its parameters are quantized, i.e. the classical functions

are converted to quantum mechanical operators. This is a standard procedure in phenomenological modelling in quantum mechanics, which deserves its own analysis. But for the purposes of this work we must discern that in such cases no stock model of Quantum Mechanics is used, and no theoretical justification exists for the quantization of classical variables. In other words, part of the Hamiltonian of the unified model is in fact semi-classical. This gives rise to questions concerning the construction of representation models that are not outright products of quantum theory alone. This aspect of modelling, which is so common in the application of Quantum Mechanics, is also reflected in the above logical schema of model construction, since there is no restriction that the f_i's and the g_{ij}'s must be dictated by theory.

Abductive reasoning enters in the construction of the unified model in two levels. The first is in reaching the conclusion that although the individual motion term and the collective motion term are constructed in significantly different ways (i.e. the first by using quantum mechanical principles from the outset, and the second by semi-classical processes) the best way to achieve an explanation of the nuclear properties is by employing both terms in a unified Hamiltonian. The second is in reaching the conclusion of what contributes to each particular term of the Hamiltonian, i.e. in establishing the best possible description of each term that would most accurately represent the different modes of motion of the nucleus.

4 Conclusion

The logical schema of the concretisation process, I suggest, captures most of the elements of theory application. But most importantly, what underlies this way of looking at theory application is that inductive and abductive inferences are mainly present in determining specific factors that influence the behaviour of physical systems, and not in determining general unifying theories. Grouping these factors together in order to reach a theoretical representation of a target physical system is a process that is primarily guided by the abstraction and concretisation processes. This is, in my view, a more precise characterisation of scientific practice, and in particular 'theory application'. The importance of induction and abduction could be best understood if these processes are seen as operating together with the process of concretisation, and the logical schema above serves as a meta-algorithm for understanding how all three processes operate together in our attempt to construct representations of phenomena.

REFERENCES

[1] N. D. Cartwright, *Nature's Capacities and their Measurement*, Clarendon Press, Oxford, 1989.

[2] J. M. Eisenberg and W. Greiner, *Nuclear Theory: Nuclear Models, Vol. 1*, North-Holland, Amsterdam, 1970.

[3] R. N. Giere, *Explaining Science: A Cognitive Approach*, The University of Chicago Press, Chicago, 1988.

[4] R. Laymon, 'Idealisation and the testing of theories by experimentation', in *Observation, Experiment, and Hypothesis in Modern Physical Science*, eds., P. Achinstein and O. Hannaway, MIT Press, Massachusetts, (1985).

[5] E. McMullin, 'Galilean idealisation', *Studies in History and Philosophy of Science*, **16**, 247–273, (1985).

[6] M. C. Morrison, 'Models as autonomous agents', in *Models as Mediators*, eds., M. S. Morgan and M. Morrison, Cambridge University Press, (1999).

[7] R. A. Nelson and M. G. Olsson, 'The pendulum- rich physics from a simple system', *American Journal of Physics*, **54(2)**, 112–121, (1986).

[8] L. Nowak, *The Structure of Idealization*, Reidel Publishing Company, Dordrecht, 1980.

[9] D. Shapere, *Reason and the Search for Knowledge*, Reidel, Dordrecht, 1984.

[10] F. Suppe, *The Semantic Conception of Theories and Scientific Realism*, University of Illinois Press, Urbana, 1989.

Disjunctive Bottom Set and Its Computation

Wenjin Lu and **Ross King** [1]

Abstract. This paper presents the disjunctive bottom set and discusses its computation. Different from existing extensions of the bottom set, such as the kernel[6], which is a set of hypotheses, the disjunctive bottom set is the weakest minimal single hypothesis in the whole hypothesis space. It happens that the disjunctive bottom set can be characterized in terms of minimal models. As minimal models can be computed in polynomial space complexity, so can the disjunctive bottom set. A flexible ILP framework based on the disjunctive bottom set is also outlined. The framework shares the low space complexity of the disjunctive bottom set. Another novelty of the framework is that it leaves an opening via hypothesis selection function to integrate more advanced hypothesis selection mechanisms.

1 Introduction

Inverse Entailment(IE) [4] is one of the most important inference mechanisms in inductive logic programming (ILP). It is an inverse process of deductive reasoning. More formally, given background knowledge B and an example E and $B \not\models E$, IE will work out a set of rules H such that

$$B \wedge H \models E$$

In practice, a typical framework for implementing IE consists of the following modules:

1. **Bottom set generation**: The bottom set of E under B, is defined as a specific (ground) clause set whose negation is derivable from $B \wedge \bar{E}$.
2. **Bottom set generalisation**: This will construct a clause theory H such that every clause in the bottom set is θ-subsumed by a clause in H.
3. **Hypothesis selection**: Biases are used for the selection of a specific hypothesis in the hypothesis space.

As an inverse process of deductive reasoning, inductive reasoning is intrinsically a multi-solution process. Given B and E, however, the hypothesis H that can be found with IE mainly depends on the bottom set.

Example 1 *Given background knowledge B and an example E as follows,*

$$\begin{aligned} B \;=\; & \{\, b \rightarrow a, \\ & \;\; c \wedge d \rightarrow a, \\ & \;\; e \rightarrow c, \\ & \;\; f \rightarrow c, \\ & \;\; g \rightarrow d, \\ & \;\; h \rightarrow d \,\} \\ E \;=\; & \;\; a \end{aligned}$$

the possible (minimal) inductive hypotheses could be:

$$\begin{aligned} &(1)\quad \{a\}, \\ &(2)\quad \{b\}, \\ &(3)\quad \{c,\ d\}, \\ &(4)\quad \{e,\ g\}, \\ &(5)\quad \{e,\ h\}, \\ &(6)\quad \{f,\ g\}, \\ &(7)\quad \{f,\ h\}, \\ &(8)\quad \{a \vee b\}, \\ &(9)\quad \{a \vee b \vee c \vee e \vee f,\ \ a \vee b \vee d \vee g \vee h\} \end{aligned}$$

Depending on the selection of bottom set, existing ILP systems may deliver different solutions. As it is limited to single Horn clause hypotheses, the Progol family takes (8) as a bottom set and may deliver (1) or (2) as hypotheses. The HAIL system allows hypotheses consisting of many Horn clauses and takes (1), (2), ..., (7) all together as the bottom set. It may deliver hypothesis from (1) to (7) but not (8) and (9). Hypothesis (9), however, does possess some desirable properties as a bottom set:

- it is a minimal hypothesis in a sense that no proper subset of (9) is a hypothesis.
- it is the weakest hypothesis in a sense that it is subsumed by other hypotheses.
- it is complete in a sense that all other hypothesis can be obtained from (9) by selecting some literals from each clause in it. Therefore it represents the multi-solution in a compact way.

This observation has led us to introduce the concept of the disjunctive bottom set which is defined as the weakest minimal ground hypothesis[2] for given background knowledge B and an example E. In addition to the properties listed above, the disjunctive bottom set also has the following advantages:

- it can be characterised by the minimal models of a simple duality transformation of B and E.
- With some restriction on the syntax of B, the disjunctive bottom set can be computed in polynomial space complexity as minimal model computation can do so.

The rest of the paper is organized as follows. After introducing some preliminaries in the next section, in section 3, we present the disjunctive bottom set. Section 4 discusses the issues of computing the disjunctive bottom set. The comparison with related work is presented in 5. We conclude the paper in section 6 by discussing some future work

2 Preliminaries and Background

In this section, based on the assumption of familiarity with first order logic and logic programming [3], we give a brief review on the

[1] Department of Computer Science, University of Wales, Aberystwyth, Ceredigion, SY23 3DB, Wales, UK, e-mail:{wwl, rdk}@aber.ac.uk

[2] see definitions in section 3

inverse entailment and its variants.

Given a first order language $\mathcal{L}$, here are the necessary notation and terminology. A positive literal is an atom and a negative literal is the negation of an atom. A ground literal is a literal without variables. We denote $HB(\mathcal{L})$ the Herbrand base of $\mathcal{L}$, the set of all ground atoms formed from $\mathcal{L}$. The disjunctive Herbrand base, denoted as $dHB(\mathcal{L})$, is the set of all (finite) positive ground disjunctions formed from the elements of the Herbrand base $HB(\mathcal{L})$. The set of all ground literals of $\mathcal{L}$ is denoted by $GL(\mathcal{L})$. A clause is a disjunction of literals where all variables in the clause are (implicitly) universally quantified. Conventionally, a clause is also represented as a set of literals which means a disjunction of the literals in the set. In logic programming setting, a clause C is written as

$$B_1 \wedge ... \wedge B_n \rightarrow A_1 \vee ... \vee A_m$$

where $m, n \geq 0$ and A_i, B_i are atoms. A Horn clause is a clause containing at most one positive literal, that is, $m \leq 1$. A (Horn) clausal theory is a conjunction of (Horn) clauses. Given C as above, $\overline{C} = (B_1 \wedge ... \wedge B_n \wedge \neg A_1 \wedge ... \wedge \neg A_m)\sigma$ is called the complement of C, where σ is a Skolemising substitution for C.

Given a clausal theory B, an (Herbrand) interpretation of B is a subset of Herbrand base. Given an interpretation I, a ground clause $C = B_1 \wedge ... \wedge B_k \rightarrow A_1 \vee ... \vee A_l$ is true in the I iff $\{B_1, ..., B_k\} \subseteq I$ implies $\{A_1, ..., A_l\} \cap I \neq \emptyset$, denoted as $I \models C$. I is a model of B iff all clauses in B are true in I. A model M of B is minimal model iff there is no model M_1 of B such that $M_1 \subset M$. The set of all minimal models of B is denoted by $\mathcal{MM}(B)$.

The central task of ILP is to find a hypothesis H from given background knowledge B and examples E such that

$$B \wedge H \models E$$

where H, B and E are all finite clausal theories. Inverse Entailment fulfills this task by so-called bottom generalisation, which is, in turn, based on bottom set [4]. The following definitions and notations are taken from [9] with B and E are limited to a Horn theory and a Horn clause, respectively.

Definition 1 (Muggleton's Bottom Set) *Let B be a Horn theory and E be a Horn clause. Then the bottom Set of B and E is the clause*

$$bot(B, E) = \{L \mid L \in GL(\mathcal{L}) \text{ and } B \wedge \overline{E} \models \neg L\}$$

We denote $bot^+(B, E)$ the set of atoms in $bot(B, E)$ and $bot^-(B, E)$ the set of atoms whose negation is in $bot(B, E)$. With the notation, we have

$$bot(B, E) \equiv \bigwedge bot^-(B, E) \rightarrow \bigvee bot^+(B, E)$$

Definition 2 (Bottom Generalisation) *Let B be a Horn theory and E be a Horn clause. A Horn clause H is said to be derivable by bottom generalization from B and E iff H θ-subsumes $Bot(B, E)$.*

For computational purpose, Bottom set has been rephrased in [9] in terms of deductive and abductive reasoning. In the following, without loss of generality, we assume that example E is a ground atom, as in the case E is a Horn clause, normalisation process can be applied [3].

[3] Given a Horn theory B and Horn clause $E = a_1 \wedge ... \wedge a_n \rightarrow b$, $\mathcal{B} = B \wedge a_1\sigma \wedge ... \wedge a_m\sigma$ and $\epsilon = b\sigma$ is called a normalisation of B and E, where σ is a Skolemising substitution for E [6]

Proposition 1 *Given Horn theory B and ground atom E with $B \not\models E$. Then*

$$\begin{aligned} bot^-(B, E) = &\quad \{a \mid a \in HB(\mathcal{L}) \text{ and } B \models a\} \\ bot^+(B, E) = &\quad \{b \mid b \in HB(\mathcal{L}) \text{ and } B \wedge \{b\} \models E\} \end{aligned}$$

The interesting point with this reformulation is that it explicitly reveals the relationship between inductive logic programming and abductive logic programming, that is, $bot^+(B, E)$ can be generated by employing an abductive procedure to abduce all single atom hypotheses (assuming that all atoms are abducible). As indicated in [6], however, Muggleton's bottom set is incomplete due to its restriction to single clause hypotheses. This has led to a further generalisation of the bottom set by allowing abductive hypotheses with multiple atoms [6, 7], which provides a semantic underpinning to a larger hypothesis space than that computed using Muggleton's bottom set.

Definition 3 (Kernel, Kernel Generalisation) *Let B be a Horn theory and E a ground atom with $B \not\models E$. Then the Kernel of B and E, written as $\mathcal{K}er(B, E)$, is the formula defined as follows:*

$$\mathcal{K}er(B, E) \equiv \bigwedge \mathcal{K}er^-(B, E) \rightarrow \bigvee \mathcal{K}er^+(B, E)$$

where

$$\begin{aligned} \mathcal{K}er^-(B, E) = &\quad \{a \mid a \in HB(\mathcal{L}) \text{ and } B \models a\} \\ \mathcal{K}er^+(B, E) = &\quad \{\Delta \mid \Delta \subseteq HB(\mathcal{L}) \text{ and } B \wedge \Delta \models E\} \end{aligned}$$

A Horn theory H is said to be derivable by Kernel Generalisation iff $H \models \mathcal{K}er(B, E)$.

It has been shown that kernel generalisation is sound in the sense that give B and E as above, for any Horn theory H, $H \models \mathcal{K}er(B, E)$ only if $B \wedge H \models E$.

3 Disjunctive Bottom Set

This section presents the formal definition of the disjunctive bottom set. After taking a further look at the Muggleton's bottom set, we show that for a given background knowledge B and a ground atom E such that $B \not\models E$, there exist an unique weakest hypothesis H such that $B \vee H \models E$. Naturally, the disjunctive bottom set is then defined to be the weakest hypothesis. We start with the following simple facts.

Proposition 2 *Let B be a Horn theory and E be a ground atom. Then for $C = c_1 \vee ... \vee c_n \in dHB$, $B \wedge C \models E$ iff $B \wedge c_i \models E$ for all $i = 1, ..., n$.*

Proposition 3 *Let B be a Horn theory and E a ground atom with $B \not\models E$. For any $H \in dHB$, if $B \wedge H \models E$, then $H \models \bigvee bot^+(B, E)$.*

Proposition 2 and proposition 3 together show that $bot^+(B, E)$ is nothing but the weakest positive ground hypothesis consisting of single clause for B and E. For example, the hypothesis (8) in example 1. Considering the fact that Muggleton's bottom set is incomplete due to this limitation, by the above propositions, it would be natural to select the weakest ground hypothesis in the whole hypothesis space as a bottom set. This is exactly the idea behind the definition of the disjunctive bottom set. In the following we give a formal account of "the weakest" ground hypothesis.

Definition 4 (Positive Ground Hypothesis) *Let B be a Horn theory and E be a ground atom where $B \not\models E$. A positive ground hypothesis of B and E is a set of positive ground clauses of the form*

$$PH = \{\mathcal{D}_i \,|\mathcal{D}_i \in dHBi = 0, 1, ..., m\}$$

satisfying

$$B \wedge \mathcal{D}_1 \wedge ... \wedge \mathcal{D}_m \models E$$

A positive ground hypothesis PH is called minimal if there is no positive ground hypothesis PH' such that $PH' \subset PH$.

In the following, a clausal theory S is said to clausally subsume a clausal theory T, written as $S \sqsupseteq T$, if every clause in T is θ-subsumed by at least one clause in S. If $S \sqsupseteq T$, then we say T is weaker than S.

Definition 5 (Weakest positive ground hypothesis) *Let PH be a minimal positive ground hypothesis of a Horn theory B and a ground atom E where $B \not\models E$. PH is called weakest iff there is no minimal positive ground hypothesis PH' of B and E such that $PH \sqsupseteq PH'$ and $PH \neq PH'$.*

The following lemma shows that the weakest positive ground hypothesis, if any, is unique.

Lemma 1 (Uniqueness of weakest positive ground hypothesis) *Let B be a Horn theory and E be a ground atom where $B \not\models E$. If both H_1 and H_2 are weakest positive ground hypotheses, then $H = H'$*[4]*.*

Proof: Let $H = H_1 \vee H_2$, then $B \wedge H \models E$. Convert H into a conjunctive normal form (CNF) and remove all clauses which are subsumed by others. Let the resulting CNF be H_c, then H_c is a positive ground hypothesis and is weaker than H_1 and H_2. But H_1 and H_2 both are weakest, we have $H_c \sqsupseteq H_i$ (i = 1, 2). As H_1, H_2 and H_c are all positive ground, we have $H_1 = H_c = H_2$.

For a given Horn theory B and an example E satisfying $B \not\models E$, we still need to show the existence of the weakest positive ground hypothesis. To fulfill this task, we borrow the approach and results from [8] which discusses the duality for goal-driven query processing in disjunctive deductive databases. The interesting point for us is that it shows that the weakest minimal hypothesis can be obtained by computing the minimal models of a duality transformation of B and E. The following result taken from [8] has been tailored and rephrased according to our needs. A more general version and its proof can be found in [8].

Definition 6 (Dual clause [8]) *Let $C = B_1 \wedge ... \wedge B_k \rightarrow A_1 \vee ... \vee A_l$ be a clause, the dual clause of C, denoted by C^d, is a clause of the form*

$$C^d = A_1 \wedge ... \wedge A_l \rightarrow B_1 \vee ... \vee B_m$$

The dual of a set of clauses S is the set S^d of duals of each of the members of S.

Theorem 1 ([8]) *Let B be a Horn theory and E be a ground atom. Let $B^d_E = B^d \cup \{E\}$. If $\mathcal{MM}(B^d_E)$ is non empty, then*

- *$B \not\models E$*
- *E becomes derivable from the updated clause theory B' achieved by adding to B the set of clauses S such that $S \sqsupseteq \mathcal{MM}(B^d_E)$.*
- *$S = \mathcal{MM}(B^d_E)$ is the minimal and weakest such set that can be added to B to guarantee the derivablility of E from B'.*

The following corollary clarifies the relationship between minimal models and positive ground disjunctive hypotheses.

Corollary 1 (Existence of weakest positive ground hypothesis) *Let B be a Horn theory and E be a ground atom with $B \not\models E$. Then $S = \mathcal{MM}(B^d \cup \{E\})$ is the weakest minimal positive ground hypothesis.*

Example 2 *Let B and E be as in example 1, then*

$$B^d = \{ \begin{array}{l} a \rightarrow b, \\ a \rightarrow c \vee d, \\ c \rightarrow e, \\ c \rightarrow f, \\ d \rightarrow g, \\ d \rightarrow h\} \end{array}$$

$\mathcal{MM}(B^d \cup \{a\}) = \{\{a, b, c, e, f\}, \{a, b, d, g, h\}\}$. *As $bot^-(B, E) = \emptyset$, we have*

$$B \cup \{a \vee b \vee c \vee \vee e \vee f,\ a \vee b \vee d \vee \vee g \vee h\} \models a$$

With lemma 1 and corollary 1, we have the following theorem.

Theorem 2 *Let B be a Horn theory and E be a ground atom with $B \not\models E$. Then there exists an unique weakest minimal positive ground hypothesis.*

With these results, we are now in a position to present the definition of the disjunctive bottom set.

Definition 7 (Disjunctive Bottom Set) *Let B be a Horn theory and E a ground atom with $B \not\models E$. Let WPH be the weakest minimal hypothesis of B and E. The disjunctive bottom set of B and E is a clausal theory of the form*

$$dBot(B, E) = \{bot^-(B, E) \rightarrow D | D \in WPH\}$$

Example 3 *Let B and E be as in example 1, By example 2 and $bot^-(B, E) = \emptyset$, we have*

$$dBot(B, E) = \{a \vee b \vee c \vee e \vee f,\ a \vee b \vee d \vee g \vee h\}$$

In the following, we show that the disjunctive bottom set is a real extension of bottom set (theorem 3). The next lemma follows the fact that for any derivation $\mathcal{D}$ of $\neg a$ from $B \wedge \neg E$, we have a derivation $\mathcal{D}^d$ of a from $B^d \wedge E$ obtained by replacing each C in the $\mathcal{D}$ with C^d which is a clause in $B^d \wedge E$.

Lemma 2 *Let B be a Horn theory and E be ground atom with $B \not\models E$. Then for any ground atom a, $B \wedge \neg E \models \neg a$ iff $B^d \wedge E \models a$.*

Theorem 3 *Let B be a Horn theory and E be a Horn clause. Then $bot(B, E) \sqsupseteq dBot(B, E)$.*

Proof: By lemma 2, for any $a \in bot^+(B, E)$, $B^d \wedge E \models a$. That is, a is true in every minimal model of $B^d \wedge E$. Therefore for every minimal model $M \in \mathcal{MM}(B^d \wedge E)$, $bot^+(B, E) \subseteq M$. Thus the theorem follows the definitions of the bottom set and the disjunctive bottom set.

[4] here we read a ground hypothesis as a set of clauses, which, in turn, are sets of ground atoms.

4 On Computation of the Disjunctive Bottom Set

In this section we discuss the issues of computing the disjunctive bottom set. By theorem 1 and the definition of the disjunctive bottom set, for given background knowledge B and an example E where $B \not\models E$, the computation of $dBot(B, E)$ turns out to be the generation of minimal models of $B^d \cup E$.

Minimal model computation has been intensively studied in the community of disjunctive logic programming and theorem proving. Many minimal model generation approaches have been proposed in the literature [5, 1]. Among them the methods based on hyper tableaux seem to offer a promising basis for minimal model reasoning [5, 1]. The hyper tableau calculus combines the idea from hyper resolution and from analytic tableaux. When applied to minimal model generation, hyper tableaux are defined as a special kind of literal trees. The tree is generated in such a way, that in any step an open branch is a candidate for a partial model.

While it is true that there is a lot of algorithms for minimal model generation, however, many of them were defined for ground theories or theories with restricted syntax. One such a restriction is range restriction clauses defined below.

Definition 8 (Range restricted clause [1]) *A clause is said to be range restricted if every variable occurring in a positive literal also appears in a negative literal. A clause theory is range restricted if every clause in it is range restricted.*

As discussed in [1], for a non-range restricted clausal theory, a range-restricted transformation can be applied to it to produce a range-restricted clauses theory [1].

Specifically, for a range restricted clause theory, there exist minimal model generation procedures with a polynomial space complexity. One such procedure is reported in [5]. The basic idea is to generate models with a hyper tableau proof procedure and to include an additional test for ruling out those branches in the tableau that do not represent minimal models. This groundedness test is done *locally*, i.e. there is no need to compare a branch with other branches computed previously; hence there is no need to store models. In the following discussion, we will rely on this fact and assume that the minimal model generation procedure provides an API $next\_minimal\_model(B)$, which takes a range-restricted clause theory B and always returns the next minimal model if any without repeating.

Next, under the assumption that for a given background Horn theory B, B^d is range-restricted, we outline an ILP framework based on the disjunctive bottom set. To make the framework more flexible, we introduce the concept of a hypothesis selection function, which will be used to select a ground hypothesis from the disjunctive bottom set.

Definition 9 (Hypothesis selection function) *A hypothesis selection function is a mapping*

$$f : 2^{HB} \rightarrow 2^{HB}$$

such that

- $f(\emptyset) = \emptyset$
- *if* $M \neq \emptyset$, *then* $f(M) \neq \emptyset$ *and* $f(M) \subseteq M$

f *is called a Horn hypothesis selection function if* $f(M)$ *contains only one atom.*

Algorithm 1 presents a computational procedure to compute inductive hypotheses. The basic idea behind the procedure is as follows: for a given Horn theory B, a ground atom E, as B^d is assumed to be range-restricted, a minimal model generation procedure can be applied to generate all minimal models of $B^d \cup \{E\}$. For each minimal model, apply hypothesis selection function to produce a partial hypothesis. This partial hypothesis is then generalised by a hypothesis generalising procedure. Once the algorithm terminates, it will produce an inductive hypothesis for E.

Algorithm 1 : Computing Inductive hypotheses

Input: A Horn Theory B,
A ground atom E,
A hypothesis selection function $\mathcal{F}$
Output: A hypothesis $\mathcal{H}$
begin
  $\mathcal{H} = \emptyset$
  repeat
    $M = next\_minimal\_model(B^d_E)$
    if $M \neq$ "no"
      let H be a generalisation of $bot(B, E)^- \rightarrow \mathcal{F}(M)$
      $\mathcal{H} = \mathcal{H} \cup \{H\}$
  until $M =$ "no"
  return $\mathcal{H}$
end

The following theorem shows that algorithm 1 is sound and complete.

Theorem 4 (Soundness and Completeness) *Let B be a Horn theory and E a ground atom. Then a clause $\mathcal{H}$ is a hypothesis of E given B iff there exists a hypothesis selection function f such that $\mathcal{H}$ is the output of algorithm 1 with the input of the Horn theory B, the ground atom E and the hypothesis selection function f.*

5 Related work

The work presented here has been influenced by several existing work. The bottom set was first introduced in [4]. As rephrased in [9, 6, 7], given a background knowledge B and ground atom E, the bottom set $bot(B, E)$ can be represented in two parts, $bot^-(B, E)$ and $bot^+(B, E)$, where $bot^-(B, E)$ is the least Herbrand model of B and $bot^+(B, E)$ is the set of atoms abducible from B and E. By proposition 2 and 3, $bot^+(B, E)$ is nothing but the weakest *single* clause which is an hypothesis for E given B. In this sense, the disjunctive bottom set is a natural extension of the bottom set as it is the weakest set of clauses which, altogether, form a hypothesis for E given B.

The disjunctive bottom set has been much inspired by the kernel set approach [6], which is a generalisation on bottom set. Given B and E, the kernel can be represented as

$$\mathcal{K}er(B, E) \equiv \bigwedge \mathcal{K}er^-(B, E) \rightarrow \bigvee \mathcal{K}er^+(B, E)$$

where

$$\begin{aligned} \mathcal{K}er^-(B, E) = &\quad \{a \mid a \in HB(\mathcal{L}) \text{ and } B \models a\} \\ \mathcal{K}er^+(B, E) = &\quad \{\Delta \mid \Delta \subseteq HB(\mathcal{L}) \text{ and } B \wedge \Delta \models E\} \end{aligned}$$

As Kernel is a complete extension of the bottom set, it it is not surprise that, the disjunctive bottom set and the Kernel are semantically

equivalent in a sense that they represented each other in a dual way. More precisely we have the following result.

Theorem 5 (Duality of the disjunctive bottom set and the Kernel) *Let B be a Horn theory and E be a ground atom where $B \not\models E$. Then*

$$\bigvee \mathcal{K}er^{+}(B, E) \leftrightarrow \bigwedge \mathcal{MM}(B^{d} \cup \{E\})$$

Proof: Let $WPH = \bigwedge \mathcal{MM}(B^{d} \cup \{E\})$, then WPH is a ground clause theory consisting of only positive ground clauses. Let Δ be a model of WPH, then Δ subsumes WPH. As WPH is the weakest hypothesis, we have $B \wedge \Delta \models E$, therefore, $\Delta \in \mathcal{K}er^{+}(B, E)$. That is, Δ is a model of $\bigvee \mathcal{K}er^{+}(B, E)$.

Now let δ be a model of $\bigvee \mathcal{K}er^{+}(B, E)$, then δ must be a hypothesis of E under B. As WPH is the weakest hypothesis, we have δ subsumes WPH. Therefore δ is a model of WPH. This completes our proof.

While it is true that the disjunctive bottom set and the kernel are semantically equivalent, the differences between the two are also clear. The $\mathcal{K}er^{+}(B, E)$ is defined as a set of hypotheses consisting of ground atoms as each Δ is a hypothesis. The disjunctive bottom set is a single hypothesis. The difference in representation has also an impact on their implementation. The kernel set approach has its implementation based on abductive reasoning and the ILP framework presented here will be implemented on top of minimal model reasoning and share the advantage of lower space complexity.

Another interesting ILP framework is CF-induction [2]. It is also sound and complete for finding hypotheses from full clausal theories, and can be used for inducing not only definite clauses but also non-Horn clauses and integrity constraints. The big difference between CF-induction and our framework is the way in which the hypotheses are computed. CF-induction computes hypotheses using a resolution method via consequence finding. Our framework is based on minimal model generation. Another difference is in dealing with bias. While it is modelled in CF-induction by production field, inductive bias can be represented in our framework via more general hypothesis selection function.

6 Conclusions and Future Work

This paper presents the disjunctive bottom set which is a natural extension of Muggleton's bottom set. Different from existing extensions, the disjunctive bottom set is the weakest minimal hypothesis and can be represented by the minimal models of a duality transformation of background knowledge B and an example E. In addition, the disjunctive bottom set can be computed in polynomial space complexity. An ILP framework based on the disjunctive bottom set is also outlined. The main novelty of the new framework is its low space complexity. In addition the hypothesis selection function in the framework leaves an opening to integrate more advanced hypothesis selection mechanism in hypothesis construction.

A lot of work remains to do. Firstly we will prototype the framework for experiment and compare the results with existing work. The other point we want to exploit further is to cooperate statistical methods into the hypothesis selection. An interesting application area will be bioinformatics, where ILP has shown great success.

Acknowledgements.We thank the anonymous reviewers for their informative and valuable comments. Also many thanks to Emma Byrne for her comments on the improvement of the paper. This work was partially funded by the European Union IST Programme, contract no. FP6-516169 IQ project and BBSRC Project BLID.

REFERENCES

[1] François Bry and Adnan Yahya, 'Positive unit hyperresolution tableaux and their application to minimal model generation', *J. Autom. Reason.*, **25**(1), 35–82, (2000).

[2] Katsumi Inoue, 'Induction as consequence finding', *Machine Learning*, **55**(2), 109–135, (2004).

[3] J.W. Lloyd, *Foundations of Logic Programming*, Springer-Verlag, Berlin, 1987. Second edition.

[4] S. Muggleton, 'Inverse entailment and Progol', *New Generation Computing, Special issue on Inductive Logic Programming*, **13**(3-4), 245–286, (1995).

[5] Ilkka Niemela, 'A tableau calculus for minimal model reasoning', in *Analytic Tableaux and Related Methods*, pp. 278–294, (1996).

[6] O. Ray, K. Broda, and A.Russo, 'Hybrid abductive inductive learning: a generalisation of progol', in *13th International Conference on Inductive Logic Programming*, eds., T. Horváth and A. Yamamoto, volume 2835 of *Lecture Notes in AI*, pp. 311–328. Springer Verlag, (2003).

[7] O. Ray, K. Broda, and A. Russo, 'Generalised Kernel Sets for Inverse Entailment', in *Proceedings of the 20th International Conference on Logic Programming*, eds., B. Demoen and V. Lifschitz, volume 3132 of *Lecture Notes in Computer Science*, pp. 165–179. Springer Verlag, (2004).

[8] Adnan H. Yahya, 'Duality for goal-driven query processing in disjunctive deductive databases', *Journal of Automated Reasoning*, **28**(1), 1–34, (2002).

[9] A. Yamamoto, 'Which hypotheses can be found with inverse entailment?', in *Proceedings of the 7th International Workshop on Inductive Logic Programming*, eds., S. Džeroski and N. Lavrač, volume 1297, pp. 296–308. Springer-Verlag, (1997).

Abduction, Induction, and the Logic of Scientific Knowledge Development

Peter Flach[1] and **Antonis Kakas**[2] and **Oliver Ray**[3]

Abstract. In this paper we outline some recent developments in the study of abduction and induction and their role in scientific modelling and knowledge refinement. We also describe a central challenge that appears to be emerging from this study: namely, the problem of developing practical approaches for exploiting abduction and induction, of formally characterising the limitations of such approaches, and of identifying the classes of real-world problems to which they can be usefully applied.

1 Modelling Scientific Theories

Modelling a scientific domain is a continuous process of observing and understanding phenomena according to some currently available model, and using this understanding to improve the original domain model. In this process one starts with a relatively simple model which gets further improved and expanded as the process is iterated. At any given stage of its development, the current model is very likely to be *incomplete*. The task then is to use the information given to us by experimental observations to improve and possibly complete this description. The development of our theories is driven by the observations and the need for these theories to conform to the observations. This point of view forms the basis of many formal theories of scientific discovery [22, 7, 15] [4] in the sense that the development of a scientific theory is considered to be an incremental process of refinement strongly guided by the empirical observations.

Considering a logical approach to this problem of incremental development of a scientific model, philosophers of science have recognized the need to introduce new *synthetic* forms of reasoning, alongside with the analytical reasoning form of deduction. Drawing on Aristotle's syllogistic logic, Charles Sanders Peirce [6, 21] distinguished between *abduction* and *induction*, and studied their respective role in the development of scientific theories. More recently, several authors have studied abduction and induction from the perspective of Artificial Intelligence and Cognitive Science. [8, 11, 17, 4]. In particular, one recent volume [4] is devoted to the problem of comparing these two forms of reasoning and investigating their possible unification or integration for the purposes of Artificial Intelligence.

Given a theory T describing our current (incomplete) model of the scientific domain under investigation, and a set of observations described by the sentences, O, abduction and induction are employed in the process of incorporating the new information contained in the observations into the current theory. They both synthesize new knowledge, H, that extends the current model to $T \cup H$, such that (1) $T \cup H \models O$, and (2) $T \cup H$ is consistent (where $\models$ denotes the deductive entailment relation of the formal logic used in the representation of our theory and consistency refers also to the corresponding notion in this logic). The particular choice of the underlying logic depends on the problem or phenomena that we are trying to model. In many cases this is based on first-order predicate calculus, as for example in the approach of Theory Completion in [20]. But other logics can be used, e.g. the non-monotonic logics of Default Logic [25] or Logic Programming with Negation-as-Failure [1, 16] when the modelling of our problem requires this level of expressivity.

Given this single formal definition of these two forms of reasoning, how can they be distinguished and why should we need to do so? One way to distinguish them is to consider the extent to which we allow the new knowledge H, to complement the current theory T. Abduction typically assumes that we can identify two distinct sets of predicates: *observable* predicates and *abducible* predicates. This reflects the assumption that our model T has reached a sufficient level of comprehension of the domain such that all the incompleteness of the model can be isolated in its abducible predicates. The observable predicates are assumed to be completely defined (in T) in terms of the abducible predicates and other background auxiliary predicates; any incompleteness in their representation comes from the incompleteness in the abducible predicates. Furthermore, the empirical observations of the domain that we are trying to model are described using observable predicates only (typically as ground atomic facts).

The abducible predicates describe underlying (theoretical) relations in our model that are not observable directly but can, through the model T, bring about observable consequences. Having isolated the incompleteness of our model in the abducible predicates, abductive reasoning generates *explanations* in terms of these predicates for understanding, according to the model, the specific observations that we have of our scientific domain. Such explanations generate knowledge that is specific to the particular state or scenario of the world pertaining to the observations explained and to the given model T from which they were generated. Adding an explanation to the theory then allows us to predict further observable information but this new predictive power is restricted to come only through the already given knowledge in our theory that defines the observable predicates. Note that the form of the abductive hypothesis depends heavily on the particular theory T at hand, and the way that we have chosen to represent the domain.

On the other hand, induction typically generates knowledge in the form of new general rules that can provide – either directly, or indirectly through the current theory T that they extend – new interrela-

[1] Department of Computer Science, University of Bristol, United Kingdom, email: Peter.Flach@bristol.ac.uk
[2] Department of Computer Science, University of Cyprus, Nicosia, Cyprus, email:antonis@cs.ucy.ac.cy
[3] Department of Compputing, Imperial College, London, United Kingdom, email:or@doc.ic.ac.uk
[4] References in this short paper are only indicative of the subject matter and are necessarily incomplete.

tionships between the predicates of our theory that can include the observable predicates and even in some cases new predicates. The inductive hypothesis thus introduces new, hitherto unknown, links between the relations that we are studying, thus allowing new predictions on the observable predicates that would have been impossible to obtain from the original theory under any abductive extension.

The role of an inductive hypothesis, H, is to extend the existing theory T to a new theory $T' = T \cup H$, rather than reason with T under the set of assumptions H as is the case for abduction. Hence T is replaced by T' to become a new theory with which we can subsequently reason, either deductively or abductively, to extract information from it. In effect, H provides the link between observables and non-observables that was missing or incomplete in the original theory T. This is particularly evident from the fact that induction can be performed starting from an empty given theory T, using just the set of observations. The observations specify incomplete (usually extensional) knowledge about the observable predicates, which we aim to *generalise* into new knowledge.

Indeed, from one point of view (e.g. as applied in Machine Learning) the essential aspect of induction seems to be the kind of sample-to-population inference exemplified by the following schema, usually called (categorical) inductive generalisation:

All objects in the sample satisfy $P(x)$;
therefore, all objects in the population satisfy $P(x)$.

In contrast, the generalising effect of abduction, if at all present, is much more limited. With the given current theory T we implicitly restrict the generalising power of abduction as we require that the basic model of our domain remains that of T. The existence of this theory separates two levels of generalisation: (a) that contained in the theory and (b) new generalisations that are not given by the theory. Through abduction we can only have the first level, while induction aims for a stronger and genuinely new generalising effect on the observable predicates. Whereas the purpose of abduction is to extend the theory with an explanation and then reason with it, thus enabling the generalising potential of the given theory T, in induction the purpose is to extend the given theory to a new theory, which can provide new possible observable consequences.

2 Integrating Abduction and Induction

This complementarity of abduction and induction suggests a basis for their integration within the context of theory formation. A *cycle of integration* of abduction and induction [3] emerges that is suitable for our task of incremental scientific modelling. Abduction is used to transform the observations to information on the abducible predicates. Then induction takes this as input and tries to generalize this information to general rules for the abducible predicates now treating these as observable predicates for its own purposes. The cycle can then be repeated by adding the learned information on the abducibles back in the model as new partial information on the incomplete abducible predicates. This will affect the abductive explanations of new observations to be used again in a subsequent phase of induction. Hence through this cycle of integration the abductive explanations of the observations are added to the theory, not in the (simple) form in which they have been generated, but in a generalized form given by a process of induction on these (Figure 1).

A simple example, adapted from [24], that illustrates this cycle of integration of abduction and induction is as follows. Suppose that our current model, T, contains the following rule and background facts:

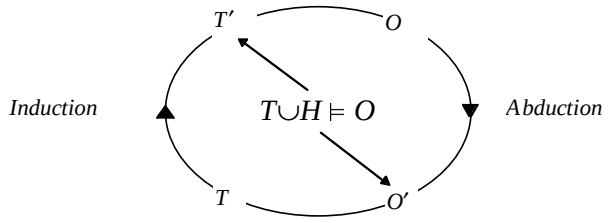

Figure 1. The cycle of abductive and inductive knowledge development: T is the current theory, O the observations triggering theory development, and H the new knowledge generated. On the left-hand side we have induction, its output feeding into the theory T for later use by abduction on the right; the abductive output in turn feeds into the observational data O for later use by induction, and so on.

$$sad(X) \textbf{ if } tired(X), poor(X).$$

$academic(oli).$	$academic(ale).$	$academic(kr).$
$student(oli).$	$lecturer(ale).$	$lecturer(kr).$
$tired(oli).$	$tired(ale).$	$tired(kr).$

Suppose also that our only observable predicate is *sad* and we are given the observations

$$O = \{sad(ale), sad(kr), not\ sad(oli)\}$$

Can we use these facts to improve our model? If we assume that the incompleteness resides in the predicate *poor*, then we can use abduction to explain the observations O via the explanation

$$E = \{poor(ale), poor(kr), not\, poor(oli)\}$$

Subsequently, treating this explanation as training data for inductive generalization we can generalize this to get the hypothesis:

$$H = \{poor(X) \textbf{ if } lecturer(X).\}$$

thus (partially) defining the abducible predicate *poor* when we extend our theory with this rule.

The combination of abduction and induction has recently been studied and deployed in several ways within the context of Inductive Logic programming (ILP). In particular, the widely used inference method of *Inverse Entailement* [20] can be seen as integrating abductive inference (which is used in the construction of the so-called "bottom clause") and inductive inference (which is used to generalize the bottom clause). This is realized in the ILP system Progol 5 and applied to several problems including the discovery of the function of genes in a network of metabolic pathways [14] and, more recently, to the study of enzyme inhibition in metabolic networks [26].

In [19] Theory Completion is realized in an ILP system called ALECTO, which integrates a first phase of *extraction* or *identification* case abduction [2] – to transform each training example into an abductive hypothesis – followed by a second phase of induction that generalizes these abductive hypotheses. It has been used to learn robot navigation control programs by completing the specific domain knowledge required, within a general theory of planning that the robot uses for its navigation [18].

Unlike most other machine learning approaches, frameworks that incorporate abductive reasoning capabilities can perform what is called *non-observation predicate learning* (non-OPL) [20] where the concept being learnt differs from that observed in the examples. This

ability is absolutely crucial in the applications cited above, where the concept of interest (e.g. enzyme inhibition) cannot be observed directly, but must be inferred indirectly from the observed data (i.e. metabolite concentrations).

The development of these initial frameworks for integrating abduction and induction in a cycle of knowledge refinemnt prompted the study of their *completeness* with respect to the general problem of finding consistent hypotheses H such that $T \cup H \models O$ for a given theory T and observations O. Progol was found to be semantically [27] and procedurally [24] incomplete and several new frameworks of integration of abduction and induction were later proposed, such as SOLDR [10], *Model Constraining Clauses* [5], *Abductive Concept Learning* (ACL) [13] and *Hybrid Abductive Inductive Learning* HAIL [24, 23].

In particular, HAIL has shown that one of the main reasons for the incompleteness of Progol is that it uses a very restricted form of abductive reasoning. Lifting some of these restrictions through the employment of methods from Abductive Logic Programming [12], HAIL has also enabled the theory and practice of Bottom Generalisation to be extended in order to allow the inference of multi-clause hypotheses in response to a single example while continuing to exploit the tried and tested mechanisms of language and search bias used in systems like Progol 5. In this way, HAIL has enlarged the class of real-world problems that are soluble in practice.

By contrast, theoretically complete inductive procedures have been proposed for full clausal logic. These include *Consequence Finding Induction* (CF-Induction) [9] and *Residue Hypotheses* [28]. CF-Induction is especially interesting as it provides some support for language bias and pruning. It also offers a useful framework to study the incompleteness of other systems such as Progol5 and HAIL, which can both be viewed as practically motivated restrictions of CF-Induction. Moreover, between the two extremes of systems like Progol and CF-Induction, there is a whole spectrum of opportunities where may lie the delicate balance between efficiency and generality that will be necessary to address real-world applications.

An exciting research agenda is therefore emerging that involves (i) exploring the inevitable tradeoffs between efficiency and generality, (ii) examining the potential utility of non-Horn learning systems by developing procedures for disjunctive and normal logic programs, (iii) studying the strengths and limitations of such approaches, (iv) identifying the class of problems to which they can be profitably applied, and (v) investigating the degree to which such methods are actually used in scientific methodology and everyday life. We believe the work presented at this workshop clearly shows that these challenges are beginning to be addressed and that the results may lead to important developments in fields ranging from the practice of Artificial intelligence and Machine Learning to scientific theory development in areas such as Systems Biology and Cognitive Science.

REFERENCES

[1] K.L. Clark, 'Negation as failure', in *Logic and Databases*, eds., H. Gallaire and J. Minker, Plenum Press, (1978).

[2] Y. Dimopoulos and A.C. Kakas, 'Abduction and Inductive Learning', in *Advances in Logic Programming*, ed., L. De Raedt, 144–171, IOS Press, (1996).

[3] P. Flach and A.C. Kakas, 'Abductive and inductive reasoning: Background and issues', in *Abductive and Inductive Reasoning*, eds., P. A. Flach and A. C. Kakas, Pure and Applied Logic, Kluwer, (2000).

[4] *Abductive and Inductive Reasoning*, eds., P. A. Flach and A. C. Kakas, Pure and Applied Logic, Kluwer, 2000.

[5] K. Furukawa and T. Ozaki, 'On the Completion of Inverse Entailment for Mutual Recursion and its Application to Self Recursion', in *Proceedings of the Work-in-Progress Track of the 10th International Conference on Inductive Logic Programming*, volume 1866 of *Lecture Notes in Computer Science*, pp. 107–119. Springer Verlag, (2000).

[6] *Collected Papers of Charles Sanders Peirce*, eds., C. Harstshorne, P. Weiss, and A. Burks, Harvard University Press, 1958.

[7] C.G. Hempel, *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, Free Press, New York, 1965.

[8] J.H. Holland, K.J. Holyoak, R.E. Nisbett, and P.R. Thagard, *Induction: Processes of Inference, Learning and Discovery*, MIT Press, 1989.

[9] K. Inoue, 'Inverse entailment for full clausal theories', in *LICS-2001 Workshop on Logic and Learning*, (2001).

[10] K. Ito and A. Yamamoto, 'Finding hypotheses from examples by computing the least generlisation of bottom clauses', in *Proceedings of Discovery Science '98*, 303–314, Springer, (1998).

[11] *Abductive Inference: Computation, Philosophy, Technology*, eds., J.R. Josephson and S.G. Josephson, Cambridge University Press, 1994.

[12] A.C. Kakas, R.A. Kowalski, and F. Toni, 'Abductive Logic Programming', *Journal of Logic and Computation*, **2**(6), 719–770, (1992).

[13] A.C. Kakas and F. Riguzzi, 'Abductive concept learning', *New Generation Computing*, **18**, 243–294, (2000).

[14] R.D. King, K.E. Whelan, F.M. Jones, P.K.G. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, and S.G. Oliver, 'Functional genomic hypothesis generation and experimentation by a robot scientist', *Nature*, **427**, 247–252, (2004).

[15] Thomas S. Kuhn, *The structure of scientific revolutions*, University of Chicago Press, 1970.

[16] V. Lifschitz, 'Answer set programming and plan generation', *Artificial Intelligence*, **138**(1-2), 39–54, (2002).

[17] L. Magnani, *Abduction, Reason and Science*, Kluwer Academic/Plenum Publishers, 2001.

[18] S. Moyle, 'Using theory completion to learn a robot navigation control program', in *Proceedings of the 12th International Conference on Inductive Logic Programming*, pp. 182–197. Springer-Verlag, (2002).

[19] S. A. Moyle, *An investigation into Theory Completion techniques in Inductive Logic Programming*, Ph.D. dissertation, Oxford University Computing Laboratory, University of Oxford, 2000.

[20] S.H. Muggleton and C.H. Bryant, 'Theory completion using inverse entailment', in *Proc. of the 10th International Workshop on Inductive Logic Programming*, pp. 130–146, Berlin, (2000). Springer-Verlag.

[21] C.S. Peirce, *Essays in the Philosophy of Science*, Liberal Arts Press, 1957.

[22] K. Popper, *The Logic of Scientific Discovery*, Basic Books, New York, 1959.

[23] O. Ray, *Hybrid Abductive-Inductive Learning*, Ph.D. dissertation, Department of Computing, Imperial College London, UK, 2005.

[24] O. Ray, K. Broda, and A. Russo, 'Hybrid Abductive Inductive Learning: a Generalisation of Progol', in *13th International Conference on Inductive Logic Programming*, volume 2835 of *LNAI*, pp. 311–328. Springer Verlag, (2003).

[25] R. Reiter, 'A logic for default reasoning', *Artificial Intelligence*, **13**, 81–132, (1980).

[26] A. Tamaddoni-Nezhad, A. Kakas, S.H. Muggleton, and F. Pazos, 'Application of abductive ilp to learning metabolic network inhibition from temporal data', in *To appear in the Journal of Machine Learning*, (2005).

[27] A. Yamamoto, 'Which hypotheses can be found with inverse entailment?', in *Proceedings of the Seventh International Workshop on Inductive Logic Programming*, 296–308, Berlin, (1997). LNAI 1297.

[28] A. Yamamoto, 'Hypothesis finding based on upward refinement of residue hypotheses', *Theoretical Computer Science*, **298**, 5–19, (2003).

An Abduction framework for Handling Incompleteness in First-Order Learning

S. Ferilli and **F. Esposito** and **N. Di Mauro** and **T.M.A. Basile** and **M. Biba**[1]

Abstract. This paper presents the ILP incremental learning system INTHELEX, focusing on its abductive capability. It is based on an abductive proof procedure that aims at attacking the problem of incomplete information by hypothesizing likely facts that are not explicitly stated in the observations. The system implements a framework in which inductive and abductive inference been brought to cooperation, and its performance in experiments on both artificial and real-world dataset is encouraging.

1 INTRODUCTION

Most traditional Machine Learning approaches focus on inductive mechanisms in order to achieve the learning goal. In order to broaden the investigation and the applicability of machine learning schemes, it is necessary to move on to more expressive representations which require more complex inference mechanisms and strategies to work together, taking advantage of the benefits that each approach can bring. In particular, one of the problems of the traditional approach to predicate-learning is the partial relevance of the available evidence, that could be takled by abduction. The problem of integrating an abductive strategy in an inductive learner is made harder in the incremental setting, where hypothesize information is more difficult since the knowledge is not completely available at the beginning.

INTHELEX (INcremental THEory Learner from EXamples) [6] is an incremental learning system for the induction of hierarchical first-order logic theories from positive and negative examples, that works under the Object Identity (OI) assumption [15]. It learns simultaneously multiple concepts, possibly related to each other, and guarantees validity of the theories on all the processed examples. It uses feedback on performance to activate the theory revision phase on a previously generated version of the theory, but learning can also start from scratch. In the learning process, it exploits a previous version of the theory (if any), a graph describing the dependence relationships among concepts, and an historical memory of all the past examples that led to the current theory. Another peculiarity of the system is the integration of multistrategy operators that may help solve the theory revision problem. The purpose of *induction* is to infer regularities and laws (from a certain number of significant observations) that may be valid for the whole population. INTHELEX incorporates two inductive refinement operators, one for generalizing hypotheses that reject positive examples, and the other for specializing hypotheses that explain negative examples.

Deduction is exploited to fill observations with information that is not explicitly stated, but is implicit in their description. Indeed, since the system is able to handle a hierarchy of concepts, some combinations of predicates might identify higher level concepts that are worth adding to the descriptions in order to raise their semantic level. For this reason, the system exploits deduction to recognize such concepts and explicitly add them to the example description. The role of *abduction* in INTHELEX is helping to manage situations where not only the set of all observations is partially known, but each observation could also be incomplete. Indeed, it can be exploited both during theory generation and during theory checking to hypothesize facts that are not explicitly present in the observations. This prevents the refinement operators from being applied, as long as possible, leaving the theory unchanged. Lastly, *abstraction* removes superfluous details from the description of both the examples and the theory. The exploitation of abstraction in the system concerns the shift from the language in which the theory is described to a higher level one according to the framework proposed in [8].

Figure 1 graphically represents the architecture of the system, embodying the cooperation between the different multistrategy operators. In the typical information flow, every incoming example preliminarily undergoes a pre-processing step of abstraction, that eliminates uninteresting details according to the available operators provided in the abstraction theory. Then, the example is checked for correct explanation according to the current theory and the background knowledge, and it is stored in the examples repository. During the coverage (i.e., checking whether the observation is explained by the current theory) and saturation (i.e., identifying higher level concepts and explicitly adding them to the example description) steps, if abduction is enabled, an abductive derivation is used. Otherwise the normal deductive derivation is started to reach the same goal without hypothesizing unseen information. In case the derivation fails, a theory refinement is necessary, and thus the example is (abductively or deductively) saturated and the inductive engine is started in order

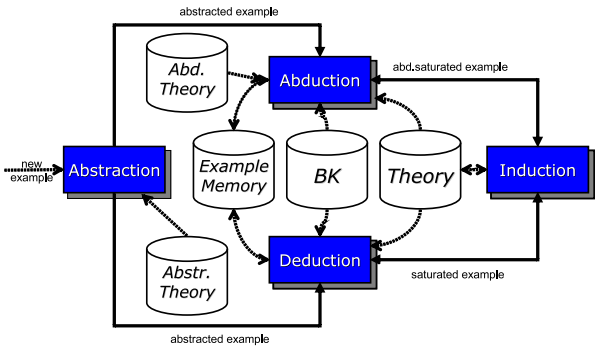

Figure 1. Architecture of the learning system

[1] Università di Bari, Italia, email: {ferilli,esposito,ndm,basile,biba}@di.uniba.it

to generalize/specialize the proper definitions, possibly using the abductive or deductive derivation whenever needed. Specifically, when a positive example is not covered, a revised theory is obtained in one of the following ways (listed by decreasing priority) such that completeness is restored: 1) replacing a clause in the theory with one of its generalizations; 2) adding a new clause to the theory; 3) adding a positive exception. When, on the other hand, a negative example is covered, a revised theory that restores consistency is reached by performing one of the following actions: 1) adding positive literals to clauses; 2) adding a negative literal to a clause; 3) adding a negative exception.

2 A FRAMEWORK FOR INTEGRATING INDUCTION AND ABDUCTION

Abduction, just like induction, has been recognized as a powerful mechanism for performing hypothetical reasoning in the presence of incomplete knowledge. Indeed, abduction is able to capture *default reasoning* as a form of reasoning which deals with incomplete information [9]. Moreover, abduction can model also *negation as failure* rule (NAF) [3], with simple transformations of logic programs into abductive theories. Thus, abduction gives a uniform way to deal with negation, incompleteness and integrity constraints [12]. The problem of Abduction, defined as *inference to the best explanation* according to a given domain theory, can be formalized as follows[4]: **Given** a theory T, some observations O and some constraints I; **Find** an explanation H such that: $T \cup H$ is consistent, $T \cup H$ satisfies I, $T \cup H \models O$. Candidate abductive explanations H should be described in terms of domain-specific predicates, referred to as *abducibles*, that are not (completely) defined in T, but contribute to the definition of other predicates. The integrity constraints I should provide indirect information about such incompleteness [9]. They can also be exploited to encode preference criteria for selecting the best explanation that may hold in this problem setting.

An abductive proof procedure can find explanations that make hypotheses (abductive assumptions) on the state of the world, possibly involving new abducible concepts. Indeed, when partial relevance is assumed, it could be the case that not only the set of all observations is partially known, but also any single observation may turn out to be incomplete. The procedure is generally goal-driven by the observations that it tries to explain. Preliminary, the top-level goal undergoes a transformation process that converts it into sub-goals. The theory and goals must be transformed into their *positive version*, by converting each literal ¬p into its positive version `not_p` (*default literals*). Moreover, to embed NAF in such a mechanism, it is necessary to add, for each predicate p, an integrity constraint stating that both p and its negation cannot hold at the same time. This provides a simple and unique modality for dealing with non-monotonic reasoning.

The classic algorithm for an abductive proof procedure [10] is analogous to standard SLD derivations, except that whenever a fact is not known or derivable to be true, before failing an attempt is made to check whether it can be abductively assumed to be true according to the given integrity constraints. Such a check is carried out by a consistency-check subroutine, ensuring that at least one condition of each constraint involving the hypothesized fact is (deductively or abductively) false. Each abductive assumption is considered as known in subsequent processing.

Abductive and Inductive operators address different forms of incompleteness in the theories. Specifically, abduction *extracts from the theory* a hypothesis which is considered to bear incompleteness with respect to some (abducible) predicates but is complete with respect to others. Moreover, the explanations constructed by abduction are specific to the situation of that observation. Hence abduction can be seen as a way to reason with incomplete information, rather than to complete knowledge [4].

Revise (T: theory; E: example; $M = M^+ \cup M^-$: historical memory);
$AbsE \leftarrow$ Abstract(E, $AbsT$)
if Derive($AbsE, T, D$) succeeds **then**
  $M \leftarrow M \cup \{AbsE \cup D\}$
else
  $M \leftarrow M \cup AbsE$; $SatE \leftarrow AbsE \cup$ Saturate($AbsE, T \cup BK$)
  if $AbsE$ is a positive example **then**
    Generalize($T, BK, SatE, M^-$)
  else
    Specialize($T, BK, SatE, M^+$)
  end if
end if

Derive (G: goal; T: theory; D: abduced literals);
if Abduction is ON at the current stage of processing **then**
  $D \leftarrow G$
  if *success* $\leftarrow$ Abduct($G, T \cup BK, AbdT, D$) succeeds **then**
    Add to D the abduced literals
  end if
else
  $D \leftarrow \emptyset$; *success* $\leftarrow$ Deduct($G, T \cup BK$)
end if
return *success*

Figure 2. Multistrategy Theory Revision in INTHELEX

Figure 2 summarizes the extension of the general schema of the inductive incremental learning system INTHELEX with an abductive proof procedure, derived from the classical one but properly modified to embed the Object Identity assumption. $M = M^+ \cup M^-$ represents the set of all positive and negative processed examples, E is the example currently examined, T represents the theory generated so far according to M. For simplicity, BK (the background knowledge), $AbsT$ (the abstraction theory) and $AbdT$ (the abduction theory), that must be provided by the user, are assumed to be fixed parameters (and hence are not present in the procedure headings). $AbsE$ and $SatE$ represent the example E after the abstraction and saturation phases, respectively; D is the set of literals (facts) returned by the abductive derivation when successfully applied to a goal G in theory T. Procedure *Derive* exploits abduction (through procedure *Abduct*) or deduction (through procedure *Deduct*), according to the specific settings for each step of the revision process, to prove a goal. It returns *true* or *false*, according to the success or failure of the proof procedure. *Saturate* is the procedure that returns all implicit information in the given example. *Generalize* and *Specialize* are the inductive operators used by the system to refine an incorrect theory. The resulting refinement is then implemented in the new version of the theory, and the procedure ends.

The system has been provided with an abductive proof procedure to help it in managing situations in which not only the set of all observations is partially known, but each observation could be incomplete too [6]. Specifically, abduction has been exploited to complete the observations in such a way that the corresponding examples are either covered (if positive) or ruled out (if negative) by the already generated theory, thus avoiding, whenever possible, the use of the generalization/specialization operators above mentioned to modify the

theory. The set of abduced literals for each observation is minimal, which ensures that the inductive operators use abducibles only when really needed. Since specific facts are not allowed in the learned theory, the abduced information is attached directly to the observation that generated it, so that the 'completed' examples obtained this way will be available for subsequent refinements of the theory. Such information will also be available to subsequent abductions, in order for them to preserve consistency among the whole set of abduced facts. To sum up, when a new observation is available, the abductive proof procedure is started, parameterized on the current theory, the example and the current set of past abductive assumptions. If the procedure succeeds, the resulting set of assumptions, that were necessary to correctly classify the observation, is added to the example description before storing it (of course, being it minimal by definition, if no assumption is needed for the correct classification, the example description is not affected). Otherwise the usual refinement procedure (generalization or specialization) is performed.

3 EXPERIMENTS

INTHELEX's abduction capability was tested on various domains, both toy and real-world ones. In the following we show the experiments aimed at assessing the quality of the results obtained by the exploitation of the abductive version of the system in handling incomplete data. INTHELEX has been provided with the abductive proof procedure [6] in order to complete the observations in such a way that the corresponding examples are correctly classified by the already generated theory, thus avoiding, whenever possible, the use of the operators to modify the theory.

Multiplexer. The "multiplexer" problem [14] aims at learning the definition of a 6-bits multiplexer. The dataset contains descriptions of all possible configurations of 6 bits, in which the first 2 bits represent the address of one of the subsequent 4 bits, that must be set at 1. Thus, if the bit addressed is actually 1 the example is positive, otherwise it is considered as negative for the target concept. Since a 6-bits multiplexer can assume $2^6 = 64$ possible configurations, the complete training set is made up of 64 examples, 32 positive and 32 negative. The representation language of the observations is the same as in [14]. Starting from scratch with the complete training set containing all the 64 possible configurations, the correct theory was learned in 1.38 secs, performing 12 theory revisions.

Successively, an incomplete dataset was obtained by corrupting 12 examples out of 64 so that only 3 bits out of 6 of the original configuration were specified. Both the examples to be corrupted and their bits to be neglected were randomly selected for 10 times. As described in [14], such an incomplete dataset was exploited for learning theories in two different ways: first using induction only, and then using induction supported by abduction. The theories obtained in the two cases were tested (without using abduction) on the uncorrupted dataset. Table 1 shows the system performance in the two cases, averaged on the 10 corrupted datasets, as regards the number of definitions in the learned theories, the performed theory revisions, the number of exceptions, runtime and predictive accuracy. The Abduction Theory provided to the system included all the predicates as abducibles, and integrity constraints meaning that "if the bit in position N is set to 0 it can't be set to 1, and *vice versa*.
INTHELEX was able to capture the correct definitions but applying less theory revisions, adding less exceptions and in less time with respect to induction alone, while not affecting the predictive accuracy.

Table 1. System performance on the Multiplexer dataset

	Def	Rev	Exceptions	Time (sec.)	Acc
W/o Abd	4.1	6.05	2.05	4.55	99.38
With Abd	4.1	5.55	0.4	4.36	99.22

Congressional Voting Records. The problem, as reported in [11], consists in classifying a Congressman as a democrat or a republican according to his votes on 16 issues. A certain amount of noise is present in the descriptions, in the form of unknown votes. Definitions for the class *democrat* were learned, exploiting first pure induction and then induction plus abduction, starting from the empty theory. The corresponding predictive accuracy was tested according to a 10-fold cross validation methodology, ensuring that each fold contained the same proportion of positive and negative examples. Table 2 shows the system performance on this dataset. It is possible to note that the use of abduction improves all evaluation parameters, except Runtime. This can be explained by taking into account the additional time needed to search for consistent abductive explanations due to the large number of integrity constraints in the abductive theory.

Table 2. System performance on the Congressional Voting Records dataset

	Def	Rev	Exceptions	Time (sec.)	Acc
W/o Abd	12.40	26.90	1.7	30.30	93.33
With Abd	10.10	19.20	0.80	41.36	96.8

Family Relationships. The experiment here described aims at investigating the abductive proof procedure behavior with respect to different degrees of incompleteness. In this case, we followed the same approach adopted by [11] on the same dataset [1]. Only examples about *father* were taken into account: the training set included 36 positive examples and 200 negative ones that were randomly generated. The examples description includes also all the known facts concerning the concepts other than *father* (i.e. *son*, *daugther*, *mother*, etc.), for a total of 742 literals. Progressive corruption of such a complete description was obtained by randomly eliminating facts from it: 100% (no incompleteness, 742 literals), 90% (668 literals), 80%, 70%, 60%, 50% and 40%. For each percentage, the dataset was corrupted in 5 different ways, thus obtaining 5 corresponding learning problems whose performance was averaged according to a 5-fold cross validation methodology, ensuring that each fold contained the same proportion of positive and negative examples. Comparing the performance with and without abduction

Table 3. System Performance on the Family dataset

		Rev/Def	Runtime	Accuracy
100%	noabd	1.6	52.25	99.58
	abd	1.2	47.13	100
90%	noabd	2.2	146.19	96.28
	abd	1.2	69.04	99.17
80%	noabd	2.3	190.12	96.27
	abd	1.2	70.35	100
70%	noabd	1.8	218.03	93.78
	abd	1.2	59.70	100
60%	noabd	1.7	287.57	92.13
	abd	0.5	448.82	100
50%	noabd	1.3	256.91	92.15
	abd	0.5	43.08	100
40%	noabd	1.2	871.51	90.9
	abd	0.5	24.32	98.75

on the corrupted datasets, the benefit becomes very evident with respect to all the parameters taken into account in Table 3. Abduction is able to preserve the theories from being refined (indeed, the number of revisions per clause dramatically decreases). Moreover, lower runtimes (except in one case) prove that the abductive procedure is also efficient. Finally, note that, in spite of the number of clauses being less when using abduction in all corrupted cases, predictive accuracy is always higher than the case without abduction.

Scientific Paper Domain. In the experiment concerning the induction of classification rules for a dataset of scientific paper documents belonging to one of 4 classes [5], the corruption consisted in eliminating 8% of the descriptors for each observation (made up of 112 facts on average (76 min-170 max)) contained in the tuning set. INTHELEX was applied first without exploiting its abductive procedure. Successively, the learning process was repeated, allowing the system to exploit its abductive capability and binary constraints made up of unary and binary predicates, i.e. of the form $(ic([a(X), b(X)], ic([c(X, Y), d(X, Y)])$.

Table 4 reports the system performance as to performed theory revisions, added definitions, predictive accuracy and runtime (secs.). Predictive accuracy and number of theory revisions improve when the abductive procedure is exploited. This means that the system was able to correctly complete the corrupted observations without applying the refinement procedure. As regards runtime, it increases because of the abductive procedure.

Table 4. System performance on the Scientific Papers Domain

	Rev	Clauses	Accuracy (%)	Runtime (sec.)
Without abd	7.72	4.09	96.24	5.16
With abd	5.58	3.18	99.32	40.05

Comparison. The proposed approach does not aim at completing the training data before the learning process starts. Thus, a comparison with systems that propose to overcome the problem of handling missing values by pre-processing the training data before the learning process starts (FOIL [13], LINUS [13], ASSISTANT [2]) would be unfair. Nevertheless, we compare our system to ACL1 [11] and mFOIL [13], the FOIL extension able to deal with incomplete data on the family and congressional votes datasets (the same exploited by [11] for the same purpose). Table 5 reveals that predictive accuracy on the family dataset for progressive corruption (which percentage is reported in the first row of the table) is almost the same as that obtained by the other systems, while on congressional voting INTHELEX turned out to be better with respect to the other systems.

Table 5. Comparison of Abduction on the Family dataset

	100	90	80	70	60	50	40
INTH.	1	99.17	1	1	1	1	98.75
ACL1	1	1	99.60	1	1	97.20	97.60
mFOIL	1	99.20	98.40	97.50	98.40	98.40	95.10

4 CONCLUSION

This paper presented the ILP incremental learning system INTHELEX, with specific focus on its abductive capability that allows it to takle the problem of relevance within a language bias, that is typical of many real-world domains. After presenting and discussing, an abductive proof procedure that aims at attacking the problem by hypothesizing likely facts that are not explicitly stated in the observations, a framework in which inductive and abductive inference been brought to cooperation, and its implementation in INTHELEX, that make it able to add unseen information that can be consistently hypothesized or deduced, have been mentioned.

The abductive proof procedure exploited in this work requires that an abductive theory for the specific application domain is available. In the current practice, it is in charge of the human expert to specify it, but it is not easy to single out and formally express such parameters. Of course quality, correctness and completeness in the formalization of such meta-information can affect the feasibility of the learning process. To overcome such a bottleneck, we also developed a procedure that can automatically generate such information starting from the same observations that are input to the learning process, thus making the learning system completely autonomous [7]. Actually, the abductive theories provided to INTHELEX for the experiments in Section 3 were automatically learned using our procedure.

REFERENCES

[1] H. Blockeel and L. De Raedt, 'Inductive database design', in *ISMIS96*, volume 1079 of *LNAI*, pp. 376–385. SV, (1996).

[2] B. Cestnik, I. Kononenko, and I. Bratko, 'Assistant 86: A knowledge-elicitation tool for sophisticated users.', in *EWSL*, pp. 31–45, (1987).

[3] K.L. Clark, 'Negation as failure', in *Logic and Databases*, eds., H. Gallaire and J. Minker, 293–322, Plenum Press, (1978).

[4] Y. Dimopoulos and A.C. Kakas, 'Abduction and learning', in *Advances in Inductive Logic Programming*, ed., L. De Raedt, 144–171, IOS Press, (1996).

[5] F. Esposito, D. Malerba, and F.A. Lisi, 'Machine learning for intelligent processing of printed documents', *Journal of Intelligent Information Systems*, **14**(2/3), 175–198, (2000).

[6] F. Esposito, G. Semeraro, N. Fanizzi, and S. Ferilli, 'Multistrategy Theory Revision: Induction and abduction in INTHELEX', *Machine Learning*, **38**(1/2), 133–156, (2000).

[7] S. Ferilli, T.M.A. Basile, N. Di Mauro, and F. Esposito, 'Automatic induction of abduction and abstraction theories from observations', in *ILP05*, eds., S. Kramer and B. Pfahringer, volume 3625 of *LNAI*, pp. 103–120. Springer Verlag, (2005).

[8] F. Giunchiglia and T. Walsh, 'A theory of abstraction', *Artificial Intelligence*, **57**(2/3), 323–389, (1992).

[9] A.C. Kakas, R. Kowalski, and F. Toni, 'Abductive logic programming', *Journal of Logic and Computation*, **2**(6), (1993). 718-770.

[10] A.C. Kakas and P. Mancarella, 'On the relation of truth maintenance and abduction', in *Proceedings of the 1st Pacific Rim International Conference on Artificial Intelligence*, Nagoya, Japan, (1990).

[11] A.C. Kakas and F. Riguzzi, 'Learning with abduction', *New Generation Computing*, **18**(3), 243, (May 2000).

[12] E. Lamma, P. Mello, M. Milano, F. Riguzzi, F. Esposito, S. Ferilli, and G. Semeraro, 'Cooperation of abduction and induction in logic programming.', in *Abductive and Inductive Reasoning: Essays on their Relation and Integration*, eds., A.C. Kakas and P. Flach, Kluwer, (2000).

[13] N. Lavrač and S. Džeroski, *Inductive Logic Programming: Techniques and Applications*, Ellis Horwood, 1994.

[14] Fabrizio Riguzzi, *Extensions of Logic Programming as Representation Languages for Machine Learning*, Ph.D. dissertation, University of Bologna, Novembre 1998.

[15] G. Semeraro, F. Esposito, D. Malerba, N. Fanizzi, and S. Ferilli, 'A logic framework for the incremental inductive synthesis of datalog theories', in *Logic Program Synthesis and Transformation*, ed., N. E. Fuchs, number 1463, pp. 300–321. Springer-Verlag, (1998).

Using Abduction for Induction of Normal Logic Programs

Oliver Ray [1]

Abstract. This paper proposes the approach of *eXtended Hybrid Abductive Inductive Learning* (XHAIL) for generalising positive and negative examples with respect to normal logic programs. A proof procedure is described that uses abduction to realise the abductive, deductive, and inductive phases which comprise this approach.

1 Introduction

Logic-based machine learning techniques have benefits over other approaches in terms of their ability to represent and utilise background knowledge and in terms of the expressivity and understandability of their hypotheses. Inductive Logic Programming (ILP) [13] is the branch of machine learning concerned with the generalisation of positive and negative examples with respect to prior knowledge expressed in a logic programming formalism. Recently, several ILP systems have been developed that also exploit techniques from Abductive Logic Programming (ALP) [6] to enable the learning of concepts different from those in the examples (e.g. Progol5 [12] and ALECTO [9]) and to allow more sophisticated inference under incomplete information (e.g. INTHELEX [4] and ACL [7]).

From a knowledge representation point of view, a key advantage of logic programming formalisms is their support for the *Negation-as-Failure* (NAF) operator. Indeed, NAF is used in most significant applications of Progol5 (such as learning the functions of genes [12]) and in most significant applications of ALECTO (such as learning robot control programs [9]). This reliance on NAF is significant given that, semantically, Progol5 and ALECTO are only defined for pure Horn clause theories. Moreover, as explained below, they are in fact unsound for programs with NAF in the sense that they can return hypotheses which do not entail all of the examples.

One difficulty of learning in the presence of NAF is the nonmonotonicity of this operator, which is essentially incompatible with the incremental methods used by most ILP systems. Take a theory with two clauses $p(X,1) \leftarrow q(X), not(r(X))$ and $p(X,2) \leftarrow r(X)$ and two examples $p(a,1)$ and $p(a,2)$. Given the mode declarations $modeh(1,q(+any))$ and $modeh(1,r(+any))$, Progol5 computes a hypothesis with two atoms $q(X)$ and $r(X)$. After picking the first example $p(a,1)$, Progol5 asserts the hypothesis $q(X)$ and retracts $p(a,1)$. In response to the second example $p(a,2)$, Progol5 asserts the hypothesis $r(X)$ and retracts $p(a,2)$. But, as it stands, Progol5 does not detect that the second hypothesis invalidates the first, so that only one of the two examples is finally covered.

Another difficulty faced by hybrid learners is the need to perform abduction through negation. For example, given a theory with two clauses $p(X) \leftarrow not(q(X))$ and $q(X) \leftarrow not(r(X))$ and two examples $p(1)$ and $p(2)$, we would like to compute the hypothesis $r(X)$. But, both the contrapositive method of Progol5 and the SOLD resolution of ALECTO – which perform the abductive reasoning of these systems – are unable to reason with negative literals and so cannot compute this hypothesis. By contrast, ALP techniques [6] are designed to handle negation, but can only return hypotheses that are sets of ground literals.

The integration of ALP and ILP techniques can potentially overcome the limitations of both these approaches. This is evidenced by the methodology of *Hybrid Abductive Inductive Learning* (HAIL) [14]. Compared to Progol5, the ability of HAIL's ALP procedure to compute multi-atom abductive hypotheses enables the inference of multi-clause hypotheses in response to a single example. Compared to ALECTO, the integrated integrity checks performed by HAIL's ALP procedure improves efficiency by detecting violations as soon as they arise. More importantly, the fact that HAIL incorporates a full ALP procedure greatly facilitates its extension from Horn clause theories to normal logic programs.

The nonmonotonicity arising from the use of NAF makes design of efficient generalisation procedures very difficult. Horn clause ILP procedures rely heavily upon the monotonicity of classical logic to support incremental learning techniques and efficient pruning mechanisms. Unfortunately, these strategies are not viable in formalisms that support NAF. Since a brute-force search of the entire hypothesis space is generally infeasible, a practical approach for restricting the search to some relevant subsets of the hypothesis space is clearly necessary. The present work suggests that HAIL can fulfil this role in much the same way as it does in the Horn clause case.

This paper introduces a generalisation of HAIL called *eXtended Hybrid Abductive Inductive Learning* (XHAIL) for logic programs with NAF. Like its predecessor, XHAIL is based on the construction and generalisation of a ground theory K called a *Kernel Set* [14]. The core procedure consists of three phases: first, the head atoms of K are obtained by an abductive procedure; then, the body literals of K are obtained by a deductive procedure; and, finally, K is generalised by an inductive procedure. A methodology is proposed that uses a standard ALP procedure to implement all three phases of the XHAIL approach. This methodology is then briefly illustrated on a small case study based on the Event Calculus (EC) [8].

2 eXtended Hybrid Abductive Inductive Learning

Given a background theory B and a set of (positive and negative) examples E, the task of ILP is to find a consistent hypothesis H that entails E relative to B. Symbolically, this requirement can be written $B \cup H \models E$. When B, H and E are Horn theories, $\models$ is the standard entailment relation of classical logic; but if B, H and E are logic programs with NAF, then an alternative logic programming

[1] Imperial College London, United Kingdom, email: or@doc.ic.ac.uk

semantics must be chosen. This paper adopts the *credulous partial stable model* semantics [6] so that $B \cup H \models E$ means the examples E are true in a partial stable model of the augmented program $B \cup H$.[2] The hypothesis H is usually restricted by some form of language and search bias. This paper utilises the well-known ILP techniques of *mode declarations* and *compression* [11].

2.1 The Covering Loop

The XHAIL methodology comprises two distinct levels. An outer *covering loop* performs the selection and normalisation of examples and invokes an inner *core procedure* to realise the construction and generalisation of Kernel Sets. The covering loop is parameterised by a selection function that, on each iteration, selects a subset of the remaining examples to be generalised. For example, selecting the first available example results in a behaviour that subsumes Progol5, while selecting all remaining examples results in a behaviour that subsumes ALECTO.[3] Of course, other policies could also be used that cluster examples in some other way.

Covering loops are already well documented. As described in [11] and [14], a general Horn clause example is dealt with by temporarily replacing all variables by fresh (Skolem) constants and transferring any body atoms in the resulting clause as ground facts to the background knowledge. The resulting ground atom is then generalised with respect to the augmented background knowledge. However, since covering approaches are only really useful in the monotonic case, the emphasis in this paper is on describing the core XHAIL procedure, which lifts the three-phase HAIL methodology from Horn clause theories to normal logic programs.

2.2 The Core Procedure

The inputs to the core procedure consist of a logic program B (background knowledge), a set of ground literals E (examples), and a set of mode declarations M (language bias) that specify a set of clauses L_M (hypothesis space). The output is a logic program $H \subseteq L_M$ (hypothesis) such that $B \cup H \models E$.[4] In addition, the core procedure attempts to minimise the number of literals in H.

Each hypothesis H is computed in three steps. The first two steps result in a ground logic program K that entails E with respect to B. The head atoms of K are computed by an abductive procedure that returns minimal abductive explanations of E whose atoms α_i are instances of M. The body literals of K are computed by a deductive procedure that uses M to compute a sequence of literals δ_i^j that are deductive consequences of B.

The theory K produced by the first two steps is then generalised in the third step by an inductive procedure which searches for a highly compressive hypothesis H that θ-subsumes K.[5] Of course, the non-monotonicity of the stable model semantics, makes it hard to design search procedures very much more efficient than a complete search of the θ-subsumption lattice bounded by K.

In this way, the theory K, which is called a *Kernel Set* of B and E, is a ground hypothesis that bounds the hypothesis space explored in the search for more general solutions. In effect, the Kernel Set acts as a filter by selecting some highly relevant set of head and body literals guided by B, E and M. By definition, the head atoms of K entail the examples and the body literals are entailed by the theory.

The intuition is essentially the same as the Horn case: namely that generalising a Kernel Set (or a Bottom Set, for that matter) is likely to produce better quality hypotheses than generalising some arbitrary theory (such as a set of random clauses, for example). But, even if this is true, it may be worth investigating the possibility that adding some random literals to K might result in further improvements.

XHAIL is based on the principle of exploiting efficient abductive methods to facilitate the computation of inductive hypotheses. But XHAIL takes this philosophy to a new extreme by using the same ALP procedure to implement the abductive, deductive and inductive phases of the proof procedure.

The ALP system used in this work is an enhanced implementation of the Kakas-Mancarella ALP procedure called *ProLogICA* [15]. The inputs are a program T (theory), a set of literals G (goals), and a set of predicates A (abducibles). Each output returned by the system consists of a substitution θ (answer) and a set of ground atoms Δ (explanation) with predicates in A such that $T \cup \Delta \models G\theta$.

The remainder of the this subsection briefly describes each of the three phases of the core XHAIL procedure and explains how they are implemented with ProLogICA by stating the theory, goals and abducibles in each case. Just like the HAIL approach, the deductive and inductive phases are applied to each explanation returned by the abductive phase in order to find the best overall hypothesis.[6]

Abductive Phase: The abductive phase of XHAIL must compute a set Δ of ground atoms that explain E with respect to B. Because each abduced atom will go in the head of a Kernel Set clause, the abducible predicates A are those predicates appearing in some head declaration of M.[7] The goals G and the theory T are simply the examples E and background knowledge B modulo two simple syntactic modifications. To ensure any type and schema requirements in the head declarations of M are respected, and to avoid potential complications caused by abducible predicates appearing in clause heads, each abducible a in A is associated with two fresh predicates denoted a' and a^*.[8] Each occurrence of a in B and E is replaced by a' and one clause is added to B of the form $a'(X_1, \ldots, X_n) \leftarrow a^*(X_1, \ldots, X_n), a(X_1, \ldots, X_n)$ where n is the arity of a, a' and a^*. For each head declaration m in M, one clause is added to B of the form $schema^*(m) \leftarrow type(m)$ where $schema(m)$ is the atom obtained by replacing each placemarker in m with a fresh variable, and $type(m)$ is the set of atoms of the form $t_i(X_i)$ where t_i is the type predicate in the placemarker that was replaced by the variable X_i.[9] Intuitively, the introduction of these clauses forces all of the abduced atoms to satisfy the language bias M at the ground level. As a result, the ALP procedure will return well-formed explanations Δ (see below) which can each be thought of as an atomic Kernel Set of B and E.

$$\Delta = \{\alpha_1, \ldots, \alpha_n\}$$

[2] This contrasts with the sceptical stable model semantics, which requires truth in all stable models, or the well-founded semantics, for example.

[3] For correctness, the former can only be used if the theory is negation free, while the latter can only be used if the examples are ground atoms.

[4] Integrity constraints are clauses in B with falsity $\bot$ in their head. Since $\bot$ is true in no models, satisfaction of the integrity constraints is implied.

[5] Clause C θ-subsumes D if $C\theta \subseteq D$ for some variable substitution θ. Program P θ-subsumes Q if each clause in P is θ-subsumed by one in Q.

[6] This paper does not discuss the many system parameters that bound the size of the computation and ensure finite termination.

[7] As defined in [11], mode declarations consist of head and body declarations each having a scheme with placemarker symbols and type predicates.

[8] Intuitively, a' acts as a non-abducible proxy for a, while a^* identifies the instances of a that satisfy the head declaration schemas.

[9] The technical details are formalised in [14] but are not especially important for the purposes of this paper.

Deductive Phase: The deductive phase of XHAIL must compute a maximally specific Kernel Set K of B and E with respect to M whose head atoms are the abducibles computed in the previous phase. This is achieved by saturating each head atom with a sequence of ground body literals entailed by B. The body literals are computed by finding the successful ground instances of the queries obtained by substituting a set of input terms into the placemarkers of the body declaration schemas. The process is identical to that used in HAIL and Progol5, except that XHAIL solves the deductive computations abductively by simply declaring an empty set of abducible predicates (so that only negative literals can actually be assumed).

This technique of using ALP to implement NAF was proposed by Eshgi and Kowalski [3] and has many advantages over standard Prolog. In particular, ALP correctly terminates on many problems involving recursion through negation as well as processing integrity constraints more efficiently and recording the negative literals used in a derivation. Saturating each head atom results in a maximally specific Kernel Set K (see below) of B and E that conforms to the language bias M at the ground level.[10] When querying a negative literal from a body declaration, it is necessary to use the type predicates to ground any variables in order to avoid floundering.

$$K = \left\{ \begin{array}{l} \alpha_1 \leftarrow \delta_1^1, \ldots, \delta_1^{m_1} \\ \vdots \\ \alpha_n \leftarrow \delta_n^1, \ldots, \delta_n^{m_n} \end{array} \right\}$$

Inductive Phase: The inductive phase of XHAIL must compute a compressive program H that θ-subsumes the Kernel Set of B and E returned in the previous phase. This process essentially amounts to replacing constants by variables and deleting as many literals from K as possible. Two simple syntactic transformations prepare the ALP system for this task through the introduction of two new predicates $try/3$ and $use/2$. First, all of the input and output terms in K are replaced by variables to give a program $K' \subseteq L_M$. Second, each body literal δ'^j_i in K' is replaced by the atom $try(i, j, [X_1, \ldots, X_k])$ where $[X_1, \ldots, X_n]$ is the list of all variables in the ith clause of K', and the two clauses $try(i, j, [X_1, \ldots, X_k]) \leftarrow not(use(i, j))$ and $try(i, j, [X_1, \ldots, X_k]) \leftarrow use(i, j), \delta'^j_i$ are added to K'. Applying an ALP procedure to the resulting theory $B \cup K'$ with the goal E and one abducible $use/2$ returns a set S of ground atoms of the form $use(i, j)$, which indicate that the corresponding literals δ'^j_i should be included in H and the others should be removed.[11] As the ALP system is biased to return minimal explanations, it is guaranteed to compute all maximally compressive hypotheses (in the sense of containing the fewest number of literals).

The intuition underlying this approach is that in order to use a head atom α'_i from K' in some derivation of E, the ALP procedure must solve each of the body atoms $try(i, 1, [X_1, \ldots, X_k]), \ldots, try(i, m_i, [X_1, \ldots, X_k])$. By the two rules added to K', each such atom can be solved in one of two ways: either by assuming $not(use(i, j))$ or by abducing $use(i, j)$ and solving δ'^j_i. The former case effectively ignores δ'^j_i as if it were not there, while the latter case solves δ'^j_i as if it were part of the clause. Once this decision is made, it can only be reconsidered upon backtracking. The list of variables ensures any bindings are correctly propagated through the clause. In this way, the ALP procedure records which atoms from K' should be included in H and which should not. This computed explanation is then used to select the best hypothesis H (see below) such that $B \cup H \models E$. [12]

$$H = \left\{ \begin{array}{l} a_1 \leftarrow d_1^1, \ldots, d_1^{q_1} \\ \vdots \\ a_p \leftarrow d_p^1, \ldots, d_p^{q_p} \end{array} \right\}$$

3 Learning Event Calculus Preconditions

This section illustrates the XHAIL procedure on an example problem simplified from [1]. Given an Event Calculus (EC) [8] description of some domain and a narrative of events, the task is to learn a set of rules stating when certain actions are impossible to perform. In this particular example, the domain concerns a pump operating in a mine. There are two actions $switchOff$ and $switchOn$ which, if they are successful, cause the predicate $pumpOn$ to change from $true$ to $false$ and vice versa. In addition, there are two predicates $water$ and $methane$ whose truth is controlled by the environment.

%— Domain Independent Axioms —%

holdsAt(F,T2) :- attempt(A,T1), initiates(A,F,T1), T1<T2, not impossible(A,T1), not clipped(T1,F,T2).

holdsAt(F,T2) :- initially(F), not clipped(0,F,T2).

holdsAt(F,T2) :- observed(F,T2).

clipped(T1,F,T2) :- attempt(A,T), terminates(A,F,T), T1=<T, T<T2, not impossible(A,T).

%— Domain Dependent Axioms —%

initiates(switchOn,pumpOn,T).

terminates(switchOff,pumpOn,T).

%— Narrative —%

attempt(switchOn,1). attempt(switchOn,2).
attempt(switchOff,3). attempt(switchOff,4).
observed(methane,1). observed(water,1).
observed(water,2). observed(water,3).

Figure 1. Theory (B)

As formalised in Figure 1 above, the background knowledge B contains the *domain independent* EC axioms which dictate how the truth of each fluent predicate changes over time in response to various actions. Intuitively, a fluent F is true at a time $T2$ if an action A was successfully attempted as some earlier time $T1$ which caused F to be true (i.e. $initiated$) and no intervening action happened in between that caused F to become false (i.e. $terminated$). Fluents can be declared as $initially$ true or can be $observed$ to be true.

[10] Technically, only those body literals δ_i^j should be added to K whose derivations do not assume the negation of any previous literal in Δ or K, as this ensures the existence of a partial stable model whereby $B \cup K \models E$.

[11] The transformation can be simplified by wrapping each body literal δ'^j_i in K' within a meta-predicate $try(i, j, \delta'^j_i)$ and adding just two clauses $try(X, Y, G) \leftarrow not(use(X, Y))$ and $try(X, Y, G) \leftarrow use(X, Y), G$.

[12] Strictly speaking, the search procedure may not respect the linking of input and output variables implied by the mode declarations if it can achieve greater compression by dropping redundant literals. Like type predicates and recalls, input and output variables are used in the construction of the Kernel Set but not in its generalisation if they would result in the computation less compressive hypotheses.

modeh(*,impossible(#action,+time)).

modeb(*,holdsAt(#fluent,+time)).

modeb(*,not holdsAt(#fluent,+time)).

Figure 2. Mode Declarations (M)

not holdsAt(pumpOn,1),

not holdsAt(pumpOn,2),

holdsAt(pumpOn,3),

holdsAt(pumpOn,4),

not holdsAt(pumpOn,5).

Figure 3. Examples (E)

In this EC axiomatisation, attempted actions only have successful outcomes if certain preconditions are satisfied: namely it is not $impossible$ to perform the action at that time. When this precondition is met, the *domain dependent* EC axioms state which fluents are affected by which actions. In this case, $switchOn$ initiates $pumpOn$ whereas $switchOff$ terminates it. The theory also contains narrative information giving the times at which certain actions were attempted and particular fluents were observed to hold.

The remaining inputs to XHAIL are the mode declarations M and the examples E formalised in Figures 2 and 3 above. The type predicates, which are not shown, simply declare the actions $switchOn, switchOff$, the fluents $methane, water, pumpOn$, and the time points $0, 1, 2, 3, 4, 5$. Given these inputs, the abductive phase of XHAIL returns just one abductive explanation Δ below, which entails the observations E when added to the theory B.

$$\Delta = \{\ \text{impossible(switchOff,3), impossible(switchOn,1)}\ \}$$

These atoms are saturated in the deductive phase to give the Kernel Set K below. The literals in the body of the first clause are the successful instances of the queries $holdsAt(X, 3)$ and $not(holdsAt(X, 3))$. (Note that, to avoid floundering, the latter query must be explicitly grounded using the type predicates. Alternatively, as explained in [10], this particular problem can be avoided by using the so-called *flip-clip* formulation of the Event Calculus.)

$$K = \left\{ \begin{array}{l} \text{impossible(switchOff,3) :- not holdsAt(methane,3),} \\ \quad \text{holdsAt(water,3),holdsAt(pumpOn,3).} \\ \text{impossible(switchOn,1) :- holdsAt(methane,1),} \\ \quad \text{holdsAt(water,1), not holdsAt(pumpOn,1).} \end{array} \right\}$$

This logic program is generalised in the inductive phase to give the hypothesis H below. After applying the necessary transformations, just one minimal hypothesis is computed $S = \{use(1, 2), use(2, 1)\}$, indicating that the second atom form the first clause and the first atom from the second clause are to appear in the hypothesis H. As required, it can be shown that the examples E are all satisfied in a stable model of the extended theory $B \cup H$.

$$H = \left\{ \begin{array}{l} \text{impossible(switchOff,X) :- holdsAt(water,X).} \\ \text{impossible(switchOn,X) :- holdsAt(methane,X).} \end{array} \right\}$$

These rules explain the failure of $perform(switchOn, 1)$ to ensure $holdsAt(pumpOn, 2)$ and of $perform(switchOff, 3)$ to ensure $not\ holdsAt(pumpOn, 4)$. They also explain the observed success of $holdsAt(pumpOn, 3)$ and $holdsAt(pumpOn, 4)$.

4 Conclusions, Related and Future Work

This paper presented an extension of HAIL from pure Horn clauses to normal logic programs. The XHAIL proof procedure for non-monotonic ILP was introduced and illustrated on a simple EC case study. It was then shown how ALP can be used to implement the abductive, deductive and inductive phases of the methodology. This achievement supports the hypothesis that abductive reasoning can be usefully exploited in inductive learning procedures. It remains to carry out a detailed comparison with related approaches for non-monotonic ILP, such as those proposed in [2, 5, 16]. Unlike these other approaches, XHAIL uses the Kernel Set to restrict the search to a relevant part of the hypothesis space. The limitations of XHAIL need to be studied more closely and it remains to validate the method on a more challenging case study.

ACKNOWLEDGEMENTS

I am grateful to Dalal Alrajeh, Krysia Broda, Antonis Kakas and Alessandra Russo for useful discussions relating to this work.

REFERENCES

[1] D. Alrajeh, A. Russo, and S. Uchitel, 'Inferring Operational Requirements from Scenarios and Goal Models Using Inductive Systems', in *Proceedings of the 5th International Workshop on Scenarios and State Machines: Models, Algorithms and Tools*, (2006).

[2] M. Bain and S. H. Muggleton, 'Non-monotonic learning', in *Inductive Logic Programming*, 145–161, Academic Press, (1992).

[3] K. Eshghi and R.A. Kowalski, 'Abduction compared with negation by failure', in *Proceedings of the 6th International Conference on Logic Programming*, pp. 234–254. MIT Press, (1989).

[4] F. Esposito, G. Semeraro, N. Fanizzi, and S. Ferilli, 'Multistrategy Theory Revision: Induction and Abduction in INTHELEX', *Machine Learning*, **38(1,2)**, 133–156, (2000).

[5] K. Inoue and Y. Kudoh, 'Learning extended logic programs', in *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, volume I, pp. 176–181. Morgan Kaufmann, (1997).

[6] A.C. Kakas, R.A. Kowalski, and F. Toni, 'Abductive Logic Programming', *Journal of Logic and Computation*, **2**(6), 719–770, (1992).

[7] A.C. Kakas and F. Riguzzi, 'Abductive concept learning', *New Generation Computing*, **18**(3), 243–294, (2000).

[8] R. Kowalski and M. Sergot, 'A logic-based calculus of events', *New Generation Computing*, **4**, 67–95, (1986).

[9] S. Moyle, *An investigation into theory completion techniques in inductive logic programming*, Ph.D. dissertation, Oxford University, 2003.

[10] S.A. Moyle, 'Using theory completion to learn a robot navigation control program', in *Proceedings of the 12th International Workshop on ILP*, volume 2583 of *LNAI*, pp. 182–197. Springer Verlag, (2002).

[11] S.H. Muggleton, 'Inverse Entailment and Progol', *New Generation Computing*, **13**(3-4), 245–286, (1995).

[12] S.H. Muggleton and C.H. Bryant, 'Theory Completion Using Inverse Entailment', in *Proceedings of the 10th International Conference on ILP*, volume 1866 of *LNCS*, 130–146, Springer Verlag, (2000).

[13] S.H. Muggleton and L. De Raedt, 'Inductive Logic Programming: Theory and Methods', *Logic Programming*, **19,20**, 629–679, (1994).

[14] O. Ray, *Hybrid Abductive-Inductive Learning*, Ph.D. dissertation, Imperial College London, 2005.

[15] O. Ray and A. Kakas, 'ProLogICA: a practical system for Abductive Logic Programming', in *Proceedings of the 11th International Workshop on Non-monotonic Reasoning*, (2006).

[16] C. Sakama, 'Induction from answer sets in nonmonotonic logic programs', *ACM Trans. on Computational Logic*, **6**(2), 203–231, (2005).

Abduction, Induction, and the Robot Scientist (invited talk abstract)

Ross King [1]

A Robot Scientist is a physically implemented computer/robotic system which utilizes techniques from artificial intelligence to carry out cycles of scientific experimentation. A central motivation for our work on the Robot Scientist project [2] is philosophical. We wish to better understand the nature of Science by building a computer/robot system that is capable of doing scientific research. This approach to the philosophy of science is analogous to the standard AI approach to the philosophy of mind: build and investigate artifacts that are empirically shown to have some of the attributes of the object of study. Our aim is to develop intelligent systems that do science. The key advantage of this approach to the philosophy of science is that it is objective: the Robot Scientist can be empirically judged to be capable of doing science or not. This approach differs fundamentally from most philosophy of science, which either studies science in the abstract, or is based on historical analysis.

Abduction and induction are integral to the Robot Scientist. We argue that for a number of the abstract concepts used by the Robot Scientist, their truth values cannot be physically verified in finite time. To reason about these abstract objects, from corresponding physical observations, therefore requires explicit inductions. The formation of hypotheses has traditionally been the hardest part of science to envisage automating. Indeed, many philosophers of science have openly expressed views that hypothesis formation could only be truly accomplished by humans. We argue that most hypothesis formation in modern biology is abductive, rather than inductive. What are hypothesised are factual relationships between objects, e.g. the gene *ypr060c* codes for enzyme *chorismate mutase*, gene *ypr060c* exists at location 675628-674858 (C) on chromosome 16. In our original Robot Scientist work we used Abductive Logic Programming to infer hypotheses. For efficiency reasons we are now using domain specialised techniques (bioinformatics). One way at looking at the bioinformatic technique of genome annotation is as abductive hypothesis generation on an enormous scale.

[1] Department of Computer Science, University of Wales, Aberystwyth, Ceredigion, SY23 3DB, Wales, UK, e-mail:rdk@aber.ac.uk

[2] R.D. King et al, 'Functional genomic hypothesis generation and experimentation by a robot scientist', *Nature*, **427**,247-252, (2004)