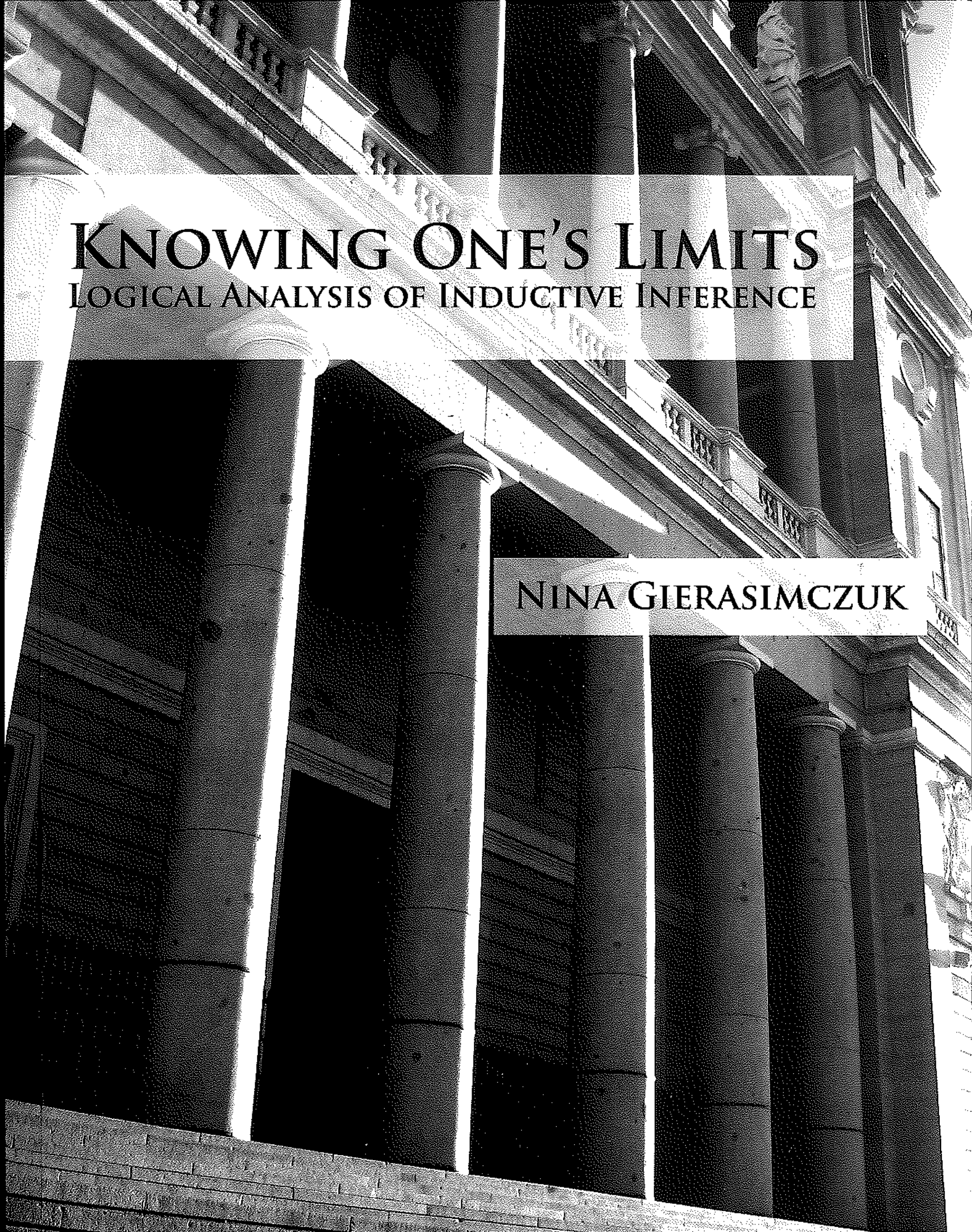

Knowing One's Limits

Logical Analysis of Inductive Inference

Nina Gierasimczuk

Knowing One's Limits

Logical Analysis of Inductive Inference

Nina Gierasimczuk

Knowing One's Limits

Logical Analysis of Inductive Inference

ILLC Dissertation Series DS-2010-11

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 904
1098 XH Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

Knowing One's Limits

Logical Analysis of Inductive Inference

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op vrijdag 17 december 2010, te 14.00 uur

door

Nina Gierasimczuk

geboren te Brussel, België.

Promotiecommissie:

Promotores:
Prof. dr. J. F. A. K. van Benthem
Prof. dr. D. H. J. de Jongh

Overige leden:
Prof. dr. P. W. Adriaans
Prof. dr. R. Clark
Prof. dr. V. F. Hendricks
Prof. dr. R. Scha
Dr. S. Smets
Prof. dr. R. Verbrugge

Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam
Science Park 904
1098 XH Amsterdam

Cover design by Nina Gierasimczuk.
Printed and bound by Ipskamp Drukkers.

ISBN: 978–90–5776–220–8

Contents

Acknowledgments

I am very grateful to my supervisors for their help and guidance. In particular, I would like to thank Prof. Dick de Jongh for showing me the value of patience and skepticism in scientific research, and for encouraging my interests in formal learning theory. To Prof. Johan van Benthem I am especially grateful for providing me with interesting challenges and opportunities, and for indicating the importance of the balance between setting boundaries and pursuing new ideas.

I am indebted to my co-authors. For the fruitful and inspiring collaboration, and all that I have gained through it I would like to thank: Dr. Alexandru Baltag, Dr. Cédric Dégremont, Prof. Dick de Jongh, Lena Kurzen, Dr. Sonja Smets, Dr. Jakub Szymanik, and Fernando Velázquez-Quesada.

There are also others that contributed to the shape of this book through discussions about various formal and philosophical issues, among them: Dr. Joel Uckelman (he also patiently proofread the dissertation), Umberto Grandi, Prof. Rineke Verbrugge, Yurii Khomski, Dr. Joanna Golińska-Pilarek, and Theodora Achourioti.

I would like to thank all members of faculty and staff, associates, and friends of the Institute of Logic, Language and Computation for creating a greatly exceptional academic and social environment.

To my closest family: my parents, Dr. Iwona and Dariusz, my sisters, Natalia and Marta and my godfather, Ryszard Peryt: I am grateful for recognizing my scientific interests, and for all other forms of appreciation and support. And thank you, Jakub, for being a great companion!

Nina Gierasimczuk
Amsterdam, November 2010

Part I

Setting and Motivation

Chapter 1

Introduction

This book is about change. Change of mind, revision of beliefs, formation of conjectures, and strategies for learning. We compare two major paradigms of formal epistemology that deal with the dynamics of informational states: formal learning theory and dynamic epistemic logic. Formal learning theory gives a computational framework for investigating the process of conjecture change (see, e.g., Jain, Osherson, Royer, & Sharma, 1999). With its central notion of identification in the limit (Gold, 1967), it provides direct implications for the analysis of language acquisition (see, e.g., Angluin & Smith, 1983) and scientific discovery (see, e.g., Kelly, 1996). On the other hand, directions that explicitly involve notions of knowledge and belief have been developed in the area of philosophical logic. After Hintikka (1962) established a precise language to discuss epistemic states, the need of formalizing dynamics of knowledge emerged. The belief-revision AGM framework (Alchourrón, Gärdenfors, & Makinson, 1985) constitutes an attempt to talk about the dynamics of epistemic states. Belief-revision policies thus explained have been successfully modeled in dynamic epistemic logic (see Van Benthem, 2007), which investigates the change in the context of multi-agent systems. Recent attempts to accommodate *iterated* knowledge and belief change is where epistemic logic meets learning theory.

Although the two paradigms are interested in similar and interrelated questions, the communication between formal learning theory and dynamic epistemic logic is difficult, mostly because of the differences in their methodologies. Learning theory is concerned with the global process of convergence in the context of computability. Belief-revision focuses on single steps of revision and constructive manners of obtaining new states, and the perspective here is more logic- and language-oriented.

Learning theory has been formed as an attempt to formalize and understand the process of language acquisition. In accordance with his nativist theory of language and his mathematical approach to linguistics, Chomsky (1965) proposed the existence of what he called a *language acquisition device*, a module that humans are born with, an 'innate facility' for acquiring language. This turned out to be only a step away from the formal definition of language learners as functions, that on ever larger and larger finite samples of a language keep outputting conjectures—grammars (supposedly) corresponding to the language in question. The generalization of this concept in the context of computability theory has taken the learners to be number-theoretic functions that on finite samples of a recursive set output indices that encode Turing machines, in an attempt to find an index of a machine that generates the set. In analogy to a child, who on the basis of finite samples learns to creatively use a language, by inferring an appropriate set of rules, learning functions are supposed to stabilize on a value that encodes a finite set of rules for generating the language.

Learning theory poses computational constraints. Learning functions are most often identified with computational devices, and this leads to assuming their recursivity. There are at least three mutually related reasons why learning theory has been developed in this direction. One comes from cognitive science: Church's Thesis in its psychological version; one is practical: the need of implementing learning algorithms; and finally there is a theoretical one: limiting recursion is in itself a mathematically interesting subject for logic and theoretical computer science.

Church's Thesis says that the human mind can only deal with computable problems. This statement underlies the very popular view about the analogy between minds and Turing machines (for an extensive discussion see Szymanik, 2009). This assumption is compatible with investigations into the implementations of learning procedures as effective algorithms. For similar reasons also the structures that are being learned are often considered to be computable—indeed, they are handled by minds which compute, or by algorithms. However restrictive these computability conditions might seem, learning remains a phenomenon of high complexity. Identification in the limit (Gold, 1967), the classical definition of successful learning, requires that the conjectures of learning functions, after some initial mind-changes, stabilize on the correct hypothesis. This exceeds computable resources, in fact it is an uncomputable, recursive in the limit, condition: there is a step k such that for all steps $n > k$ the computable learning function L outputs the correct hypothesis. Therefore, the question whether a structure is learnable falls outside the range of computable problems. Classes of sets for which such learning functions exist, i.e., learnable classes, constitute the domain of limiting recursion theory, an autonomous topic of research in theoretical computer science.

Summing up, the motivation of language acquisition initially directed learning considerations towards a recursive framework, with agents represented as certain type of number-theoretic functions. The discipline has been restricted to the functions that satisfy the limiting conditions of convergence on certain data structures. One might say that the domain has been taken over by successful, ultimately *reliable* functions (for learning theory in terms of reliability see, e.g., Kelly, 1998a). The observation that reliability is the feature that distinguishes successful learning functions from other possible mind-change policies led to relaxing the recursive paradigm. Learning theory has been re-interpreted as the framework for analyzing the procedural aspects of science, and became a study of information flow and general inquiry. This resulted in the treatment of formal learning theory as the mathematical embodiment of a normative epistemology.[1] In philosophy of science and general epistemology there is no need to assume that theory change is governed by a *computable* function. Immediately after dropping the heavy machinery of computability, learning theory linked to the problematics of knowledge and belief revision (see, e.g., Hendricks, 1995; Jain et al., 1999; Kelly, 1996), with attempts to plug the ready-to-use framework of successful convergence into the considerations of iterated belief-revision.

On the other side, a logical approach to belief-revision has been proposed in the so-called AGM framework (Alchourrón et al., 1985), where the beliefs of an agent are represented as a logically closed set of sentences of a particular language. A (new) belief-representing sentence gets introduced to the set and causes a belief change, which often leads to the necessity of removals to keep the beliefs consistent. AGM theory provides a set of axioms that put some rationality constraints on such revisions and allow the evaluation of various belief-revision policies. Presently, a very promising direction of combining the belief-revision framework with modal logics of knowledge and belief gives us a way to investigate revisions in a more linguistically-detached way. In this thesis we will look at these problems from a recently developed perspective of dynamic epistemic logic.

The framework of dynamic epistemic logic comprises a family of logics of explicit informational actions and corresponding knowledge and belief changes in agents. The information flow consisting of update actions performed in a stepwise manner can be defined as transformations of models. Those transformations can be studied and analyzed explicitly by combining techniques from epistemic, doxastic, and dynamic logic. Being logics, dynamic epistemic systems come with a semantics, but also with syntax: a formal language and a proof theory. Interestingly, like in learning theory, one of the sources is natural language and communication, but others include epistemology, and theories of agency in computer science (in particular Baltag, Moss, & Solecki, 1998; Gerbrandy, 1999a, developed basic update mechanisms that will be used in this thesis). By now many authors see dynamic epistemic logic as a general theory of social information- and preference-driven agency, which has led to growing links with temporal logics, game theory, and other formal theories of interaction (see Van Benthem, 2010). All these more recent themes will return at places in this dissertation.

This thesis brings learning theory and dynamic epistemic logic together on two levels. The first link is *semantic*. We combine local update mechanisms of dynamic epistemic logic, that constitute constructive step-by-step changes of current epistemic states, with the long-term temporal modeling offered by

[1] For the characteristics and history of this line of research see, e.g., the Stanford Encyclopedia of Philosophy entry *Formal Learning Theory* by Oliver Schulte.

learning theory. In terms of benefits of the paradigms, learning theory receives the fine-structure of well-motivated local learning actions[2], and dynamic epistemic logic gets a long-term 'horizon' which it missed (this approach is developed in Chapters 3 and 4). The second link is *syntactic*. Dynamic epistemic logic has its syntax and proof theory, learning theory does not. We show how basic notions of learning theory can be given simple perspicuous qualitative formulations in dynamic epistemic languages (the syntactic link is developed in Chapter 5). In the long run this perspective offers a chance of generic reasoning calculi about inductive learning.

***

The content of this thesis is organized in three parts. Let us give a brief overview of the chapters.

In Part I we introduce the setting and the motivation of the thesis. Chapter 2 gives mathematical preliminaries to the basic frameworks of formal learning theory and logics of knowledge and belief. Chapter 3 is intended to methodologically compare the two frameworks and provide a conceptual 'warming up' for the next part.

Part II is concerned with generally understood definability notions: expressing learnability conditions in the language of epistemic and doxastic logic. Chapter 4 gives a dynamic epistemic logic account of iterated belief-revision. By reinterpreting belief-revision policies as learning methods, we evaluate update, lexicographic and minimal upgrades with respect to their reliability on different kinds of incoming information. We are mainly concerned with identifiability in the limit. In the first part we restrict ourselves to learning from sound and complete streams of positive data. We show that learning methods based on belief revision via conditioning (update) and lexicographic revision are universal, i.e., provided certain prior conditions, those methods are as powerful as identification in the limit. We show that in some cases, these priors cannot be modeled using standard belief-revision models (as based on well-founded preorders), but only using generalized models (as simple preorders). Furthermore, we draw conclusions about the existence of tension between conservatism and learning power by showing that the very popular, most 'conservative' belief-revision method fails to be universal. In the second part we turn to the case of learning from both positive and negative data, and we draw conclusions about iterated belief revision governed by such streams. This enriched framework allows us to consider the occurrence of erroneous information. Provided that errors occur finitely often and are always eventually corrected we show that the lexicographic revision method is still reliable, but more conservative methods fail.

[2] One approach to learning theory, *learning by erasing* (see Section 2.1), uses update-like actions of hypotheses deletion.

In Chapter 5 we are again concerned with learnability properties analyzed in the context of epistemic and doxastic logic. We study both finite identification and identification in the limit. We represent the initial uncertainty of the learner as an epistemic model and characterize the conditions of the emergence of irrevocable knowledge in epistemic and dynamic epistemic logic. Then, we move to the case of identifiability in the limit and we give a doxastic logic characterization of the conditions required for converging to true stable belief. Following recent results on the correspondence between dynamic epistemic and temporal epistemic logics, we also give a characterization of learnability in terms of temporal protocols. We use the fact that the identification of sets can be performed by means of epistemic update. In the general context of learnability of protocols we characterize finite and limiting identification in an epistemic temporal and doxastic temporal language. Our temporal logic based approach to inductive inference gives a straightforward framework for analyzing various domains of learning on a common ground.

Part III consists of concrete case studies developing the general bridge that we built further, while also adding new themes. In Chapter 6 we are concerned with the problem of obtaining and using minimal samples of information that allow reaching certainty (i.e., allow finite identification). With the notion of eliminative power of incoming information, we analyze the computational complexity of finding such minimal samples. The problem of finding minimal-sized samples turns out to be NP-complete. Moreover, in the general case, we show that if we assume learners to be recursive, there are situations in which full certainty can be obtained in a computable way, but it cannot be computably realized by the learner at the first possible moment, i.e., as soon as the objective ambiguity between possibilities disappears. We also investigate different types of preset learners, that are tailored to use the knowledge of such minimal samples in their identification procedure. Differences in computational complexity between reaching certainty and reaching it in the optimal way give a motivation for explicitly introducing a new agent, a teacher, and provide a computational analysis of teachability.

In Chapter 7 we abstract away from the cooperativeness of the learner and the teacher, the property that is uniformly assumed in learning theory. We investigate the interaction between them in a particular kind of supervision learning games based on sabotage games. We are interested in the complexity of teaching, which we interpret in a similar way as in Chapter 6. Assuming the global perspective of the teacher, we treat the teachability problem as deciding whether the learning process can possibly be successful. We interpret learning as a game and hence we identify learnability and teachability with the existence of winning strategies in those games. In this context, we analyze different learning and teaching attitudes, varying the level of the teacher's helpfulness and the learner's willingness to learn. We use sabotage modal logic to reason about these games and, in particular, we identify formulae of the language that characterize the existence of winning strategies in each of the scenarios. We provide the complexity results for the related model-checking problems. They support the intuition that the cooperation of agents facilitates learning. Additionally, we observe the asymmetric nature of the moves of the two players and investigate a version without strict alternation of moves.

Finally, in Chapter 8 we consider another type of inductive inference that consists of iterated epistemic reasoning in multi-agent scenarios. We generalize the Muddy Children puzzle to treat arbitrary quantifiers in Father's announcement. Each child in the puzzle is viewed as a scientist who tries to inductively decide a hypothesis. The interconnection with other scientists can influence the discovery in a positive way. We characterize the property that makes quantifier announcements relevant in an epistemic context. In particular, we show what makes them prone to the occurrence of iteration of epistemic reasoning. The most immediate contribution to dynamic epistemic logic is a concise, linear representation of the epistemic situation of the Muddy Children. Moreover, we give a characterization of the solvability of the Muddy Children puzzle and a uniform way of deciding how many steps of iterated epistemic reasoning are needed for reaching the solution. This explicit, step by step analysis brings us closer to investigating the internal complexity of epistemic problems that the agents are facing and allows a comparison with computational complexity results from the domain of natural language quantifier processing.

Chapter 9 concludes the thesis by giving an overview of results and open questions.

As the reader may have observed from the above overview, the topics of this dissertation are drawn mainly from the domain of logic and theoretical computer science, at points reaching out to game theory and cognitive science. The approach is highly interdisciplinary. Even though the author's goal was to make this thesis self-contained, the reader is still assumed to be acquainted with basics of mathematical logic, computability and complexity theory.

Sources of the chapters

Chapter 3 is based on:

Gierasimczuk, N. (2009). Bridging learning theory and dynamic epistemic logic. *Synthese*, *169*(2), 371–384.

Gierasimczuk, N. (2009). Learning by erasing in dynamic epistemic logic. In *LATA'09: Proceedings of 3rd International Conference on Language and Automata Theory and Applications*, vol. 5457 of *LNCS*. (pp. 362–373). Springer.

Chapter 4 is based on:

Baltag, A., Gierasimczuk, N., & Smets, S. (2010). Truth tracking and belief revision. Manuscript. Presented at NASSLLI'10, Bloomington.

Chapter 5 is based on:

Dégremont, C., & Gierasimczuk, N. (2009). Can doxastic agents learn? On the temporal structure of learning. In X. He, J. F. Horty, & E. Pacuit (Eds.) *LORI'09: Proceedings of 2nd International Workshop on Logic, Rationality, and Interaction*, vol. 5834 of *LNCS*, (pp. 90–104). Springer.

Dégremont, C., & Gierasimczuk, N. (2010). Finite identification from the viewpoint of epistemic update. To appear in *Information and Computation*.

Chapter 6 is based on:

Gierasimczuk, N., & de Jongh, D. (2010). On the minimality of definite finite tell-tale sets in finite identification of languages. *The Yearbook of Logic and Interactive Rationality*, (pp. 26–41). Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Chapter 7 is based on:

Gierasimczuk, N., Kurzen, L., & Velázquez-Quesada, F. R. (2009). Learning and teaching as a game: A sabotage approach. In X. He, J. F. Horty, & E. Pacuit (Eds.) *LORI'09: Proceedings of 2nd International Workshop on Logic, Rationality, and Interaction*, vol. 5834 of *LNCS*, (pp. 119–132). Springer.

Gierasimczuk, N., Kurzen, L., & Velázquez-Quesada, F. R. (2010). Games for learning: A sabotage approach. Submitted to *Logic Journal of the Interest Group in Pure and Applied Logic*.

Chapter 8 is based on:

Gierasimczuk, N., & Szymanik, J. (2010). Muddy Children Playground: Number Triangle, Internal Complexity, and Quantifiers. Presented at *Logic, Rationality and Intelligent Interaction Workshop*, ESSLLI'10, Copenhagen.

Chapter 2

Mathematical Prerequisites

This chapter gathers background information on the two major paradigms discussed and linked in this thesis. First, preliminaries of formal learning theory are given. Then we discuss the basics of dynamic epistemic logic approaches to information and belief change. In both cases the existing literature varies in basic notions and notation. The decisions taken in this chapter should be viewed as defining the framework and laying the grounds for this thesis, rather than restricting the paradigms or indicating a general preference. For exhaustive overviews and references the reader is advised to consult respectively (Jain et al., 1999) and (Van Ditmarsch, Van der Hoek, & Kooi, 2007).

2.1 Learning Theory

Learning theory is concerned with sequences of outputs of recursive functions, focusing on those that stabilize on an appropriate value (Gold, 1967; Putnam, 1965; Solomonoff, 1964a,b). As mentioned in the introduction, the general motivation here is the possibility of inferring general conclusions from partial, inductively given information, as in the case of language learning (inferring grammars from sentences) and scientific inquiry (drawing general conclusions from partial experiments). These processes can be thought of as games between Scientist (Learner) and Nature (Teacher). At the start there is a class of possible worlds, or a class of hypotheses. It is assumed that both Scientist and Nature know what those possibilities are, i.e., they both have access to the initial class. Nature chooses one of those possible worlds to be the actual one. Scientist's aim is to guess which one it is. He receives information about the world in an inductive manner. The stream of data is infinite and contains only and all the elements from the chosen reality. Each time Scientist receives a piece of information he answers with one of the hypotheses from the initial class. We say that Scientist identifies Nature's choice in the limit if after some finite number of guesses his answers stabilize on a correct hypothesis. Moreover, it is required that the same is true for all the possible worlds from the initial class, i.e., regardless of which element from the class is chosen by Nature to be true, Scientist can identify it in the limit. In what follows, the possibilities are taken to be sets of integers, and they will be often called *languages*.

Let $U \subseteq \mathbb{N}$ be an infinite recursive set; we call any $S \subseteq U$ a language. In the general case, we will be interested in indexed families of recursive languages, i.e., classes $\mathcal{C}$ for which a computable function $f : \mathbb{N} \times U \to \{0, 1\}$ exists that uniformly decides $\mathcal{C}$, i.e.,

$$f(i, w) = \begin{cases} 1 & \text{if } w \in S_i, \\ 0 & \text{if } w \notin S_i. \end{cases}$$

In large parts of this thesis we will also consider $\mathcal{C}$ to be $\{S_1, S_2, \ldots, S_n\}$, a finite class of finite sets, in which case we will use $I_\mathcal{C}$ for the set containing indices of sets in $\mathcal{C}$, i.e., $I_\mathcal{C} = \{1, \ldots, n\}$. We will often refer to the setting in which the possible realities are taken to be sets using the terms *language learning* or *set learning*.

The global input for Scientist is given as an infinite stream of data. In learning theory, such streams are often called *texts* (positive presentations).[1]

Definition 2.1.1. *By a* text (positive presentation) ε *of* S *we mean an infinite sequence of elements from* S *enumerating all and only the elements from* S *(allowing repetitions).*

Definition 2.1.2. *We will use the following notation:*

- ε_n *is the* n*-th element of* ε*;*
- $\varepsilon{\restriction}n$ *is the sequence* $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$*;*
- $\mathrm{set}(\varepsilon)$ *is the set of elements that occur in* ε*;*
- *Let* U^* *be the set of all finite sequences over* U*. If* $\alpha, \beta \in U^*$*, then by* $\alpha \sqsubset \beta$ *we mean that* α *is a proper initial segment of* β*.*
- L *is a learning function—a recursive map from finite data sequences to indices of hypotheses,* $L : U^* \to \mathbb{N}$*. We will sometimes take the learning function to be* $L : U^* \to \mathbb{N} \cup \{\uparrow\}$*. Then the function is allowed to refrain from giving a natural number answer, in which case the output is marked by* $\uparrow$*, but the function remains recursive.*[2] *We sometimes relax the condition of recursivity of* L *to discuss some cases of non-effective finite identifiability.*

[1] We will be mainly concerned with sequences of positive information, *texts*. They are sometimes also known under the name of *environments* (see, e.g., Jain et al., 1999). The type of information that, besides positive, includes also negative information is usually called an *informant*.

[2] The symbol $\uparrow$ in the context of learning functions should not be read as a calculation that does not stop.

- *Let* $T \subseteq \mathbb{N}$ *be a finite set. Then* $\hat{T}$ *is the finite sequence such that* $\mathrm{set}(\hat{T}) = T$ *and* $\mathrm{length}(\hat{T}) = |T|$*, where* $|\cdot|$ *stands for the cardinality of a set, and* $\hat{T}$ *enumerates the integers from* T *in increasing order.*

2.1.1 Finite Identification

Finite identifiability of a class of languages from positive data is defined by the following chain of conditions.

Definition 2.1.3. *A learning function* L*:*

1. *finitely identifies* $S_i \in \mathcal{C}$ *on* ε *iff, when inductively given* ε*, at some point* L *outputs a single value* i*;*
2. *finitely identifies* $S_i \in \mathcal{C}$ *iff it finitely identifies* S_i *on every* ε *for* S_i*;*
3. *finitely identifies* $\mathcal{C}$ *iff it finitely identifies every* $S_i \in \mathcal{C}$*.*

A class $\mathcal{C}$ *is finitely identifiable iff there is a learning function* L *that finitely identifies* $\mathcal{C}$*.*

Example 2.1.4. *Let* $\mathcal{C}_1 = \{S_i = \{0, i\} \mid i \in \mathbb{N} - \{0\}\}$*.* $\mathcal{C}_1$ *is finitely identifiable by the following function* $L : U^* \to \mathbb{N} \cup \{\uparrow\}$*:*

$$L(\varepsilon{\restriction}n) = \begin{cases} \uparrow & \text{if } \mathrm{set}(\varepsilon{\restriction}n) = \{0\} \text{ or } \exists k < n\ L(\varepsilon{\restriction}k) \neq \uparrow, \\ \max(\mathrm{set}(\varepsilon{\restriction}n)) & \text{otherwise.} \end{cases}$$

In other words, L *outputs the correct hypothesis as soon as it receives a number different than* 0*, and the procedure ends.*

To see how restrictive this notion is, we can consider a finite class of languages that is not finitely identifiable.

Example 2.1.5. *Let* $\mathcal{C}_2 = \{S_i = \{0, \ldots, i\} \mid i \in \{1, 2, 3\}\}$*.* $\mathcal{C}_2$ *is not finitely identifiable. To see that, assume that* $S_2 = \{0, \ldots, 2\}$ *is chosen to be the actual world. Then the learning function can never conclusively decide that* S_2 *is the actual language. For all it knows,* 3 *might appear in the future, so it has to leave the* S_3*-possibility open.*

A necessary and sufficient condition for finite identifiability has already been formulated in the literature (Lange & Zeugmann, 1992; Mukouchi, 1992).

Definition 2.1.6 (Mukouchi 1992). *A set* D_i *is a definite finite tell-tale set for* $S_i \in \mathcal{C}$ *if*

1. $D_i \subseteq S_i$*,*

2. D_i is finite, and

3. for any index j, if $D_i \subseteq S_j$ then $S_i = S_j$.

On the basis of this notion, finite identifiability can be then characterized in the following way.

Theorem 2.1.7 (Mukouchi 1992). *A class $\mathcal{C}$ is finitely identifiable from positive data iff there is an effective procedure $\mathcal{D} : \mathbb{N} \to \mathcal{P}^{<\omega}(\mathbb{N})$, given by $n \mapsto \mathcal{D}_n$, that on input i produces a definite finite tell-tale of S_i.*

In other words, each set in a finitely identifiable class contains a finite subset that distinguishes it from all other sets in the class. Moreover, for the effective identification it is required that there is a *recursive* procedure that provides such definite finite tell tale-set.

2.1.2 Identification in the Limit

Let us consider again Example 2.1.5, i.e., take $\mathcal{C}_2 = \{S_i = \{0, \ldots, i\} \mid i \in \{1, 2, 3\}\}$, but now assume that S_2 is the actual language. Then Scientist cannot conclusively decide that it is the case. There is however a way to deal with this kind of uncertainty. Namely, if we allow Scientist to answer each time he gets a new piece of data, we can define the success of learning using the notion of *convergence* to the right answer. After seeing 0, 1 and 2 Learner can keep conjecturing S_2 indefinitely, because in fact 3 will never appear. This leads to the notion of identification in the limit.

Definition 2.1.8 (Identification in the limit (Gold, 1967)). *Learning function L:*

1. identifies $S_i \in \mathcal{C}$ in the limit on ε iff for co-finitely many m, $L(\varepsilon{\restriction}m) = i$;

2. identifies $S_i \in \mathcal{C}$ in the limit iff it identifies S_i in the limit on every ε for S_i;

3. identifies $\mathcal{C}$ in the limit iff it identifies in the limit every $S_i \in \mathcal{C}$.

A class $\mathcal{C}$ is identifiable in the limit iff there is a learning function that identifies $\mathcal{C}$ in the limit.

Below we give some examples of classes of languages which are identifiable in the limit. First let us consider an example of a finite class of finite sets.

Example 2.1.9. *Recall the class $\mathcal{C}_2$ from the previous example. $\mathcal{C}_2$ is identifiable in the limit by the following function $L : U^* \to \mathbb{N}$:*

$$L(\varepsilon{\restriction}n) = \max(\mathrm{set}(\varepsilon{\restriction}n)).$$

We can use the same learning function to identify an infinite class of finite sets.

Example 2.1.10. *Let $\mathcal{C}_3 = \{S_i \mid i \in \mathbb{N} - \{0\}\}$, where $S_n = \{1, \ldots, n\}$.*

The property of identification in the limit of the class $\mathcal{C}_3$ is lost when we enrich it with the set of all natural numbers.

Example 2.1.11. *Let $\mathcal{C}_4 = \{S_i \mid i \in \mathbb{N}\}$, where $S_0 = \mathbb{N}$ and for $n \geq 1$, $S_n = \{1, \ldots, n\}$. $\mathcal{C}_4$ is not identifiable in the limit. To show that this is the case, let us assume that there is a function L that identifies $\mathcal{C}_4$. We will construct a text, ε on which L fails: ε starts by enumerating $\mathbb{N}$ in order: $0, 1, 2, \ldots$, if arriving at a number k, L decides it is S_0, we start repeating k indefinitely. This means we will have a text for S_k. As soon as L decides it is S_k we continue with $k+1, k+2, \ldots$, so we get a text for S_0, etc. This shows that there is a text for a set from $\mathcal{C}_4$ on which L fails.*

We have already seen an infinite class of finite sets that is identifiable in the limit. The next example shows an infinite class of *infinite* sets that is identifiable in the limit.

Example 2.1.12. *Let $\mathcal{C}_5 = \{S_n \mid S_n = \mathbb{N} - \{n\}, n \in \mathbb{N}\}$. $\mathcal{C}_5$ is identifiable in the limit by the learning function $L : U^* \to \mathbb{N}$:*

$$L(\varepsilon{\restriction}n) = \min(\mathbb{N} - \mathrm{set}(\varepsilon{\restriction}n)).$$

A characterization of classes that are identifiable in the limit can be given in terms of *finite tell-tale sets*[3] (Angluin, 1980).

Definition 2.1.13 (Angluin 1980). *A set D_i is a finite tell-tale set for $S_i \in \mathcal{C}$ if*

1. $D_i \subseteq S_i$,

2. D_i is finite, and

3. for any index j, if $D_i \subseteq S_j$ then $S_j \not\subset S_i$.

Identifiability in the limit can be then characterized in the following way.

Theorem 2.1.14 (Angluin 1980). *An indexed family of recursive languages $\mathcal{C} = \{S_i \mid i \in \mathbb{N}\}$ is identifiable in the limit from positive data iff there is an effective procedure $\mathcal{D}$, that on input i enumerates all elements of a finite tell-tale set of S_i.*

In other words, each set in a class that is identifiable in the limit contains a finite subset that distinguishes it from all its subsets in the class. Moreover, for the effective identification it is required that there is a *recursive* procedure that enumerates such finite tell-tales.

[3] The notion of *definite* finite tell-tale set from Definition 2.1.6 in the previous section, is a modification and strengthening of the presently discussed, original notion of finite-tell tale set.

2.1.3 Other Paradigms

Learning by Erasing Learning by erasing (Lange, Wiehagen, & Zeugmann, 1996) is an epistemologically intuitive modification of the identification in the limit. It has not drawn much attention in the field of formal learning theory but for our purposes (a comparison with the approach of dynamic epistemic logic) it is interesting. Very often the cognitive process of converging to a correct conclusion consists of eliminating those possibilities that are falsified during the inductive inquiry. Accordingly, in the formal model the outputs of the learning function are negative, i.e., the function each time eliminates a hypothesis, instead of explicitly guessing one that is supposed to be correct. The difference between the definition of this approach and the usual identification is in the interpretation of the conjecture of the learning function. In learning by erasing one assumes an ordering of the initial hypothesis space isomorphic to the natural numbers. This allows one to interpret the actual positive guess of the learning-by-erasing function to be the least hypothesis (in the given ordering) not yet eliminated.

Let us give now the two definitions that shape the notion of learning by erasing.

Definition 2.1.15 (Function stabilization). *In learning by erasing we say that a function stabilizes to number k on environment ε iff for co-finitely many $n \in \mathbb{N}$:*

$$k = \min(\{\mathbb{N} - \{L(\varepsilon{\restriction}1), \ldots, L(\varepsilon{\restriction}n)\}\}).$$

Definition 2.1.16 (Learning by erasing (Lange et al., 1996)). *We say that a learning function L:*

1. *learns $S_i \in \mathcal{C}$ by erasing on ε iff L stabilizes to i on ε;*
2. *learns $S_i \in \mathcal{C}$ by erasing iff it learns S_i by erasing from every ε for S_i;*
3. *learns $\mathcal{C}$ by erasing iff it learns every $S_i \in \mathcal{C}$ by erasing.*

A class $\mathcal{C}$ is learnable by erasing iff there is a learning function that learns $\mathcal{C}$ by erasing.

It is easy to observe that in this setting learnability heavily depends on the chosen *enumeration* of languages, since the positive conjecture of the learning function is interpreted as the minimal one that has not yet been eliminated.

Several types of learning by erasing have been proposed. They vary in the condition of which hypotheses the learning function is allowed to remove (for details and results on learning by erasing see Lange et al., 1996).

Function learning Let us now mention another paradigm of learning in the limit—function learning. This falls out of the scope of the language-learning paradigm, but the notion of identification is in its essence very similar. The success of learning is again defined in the limit as convergence to a correct hypothesis.

This time however we take possible realities to be total recursive functions. This can be made concrete in various ways. For instance, it has been considered as a way to model program synthesis in the context of learning and empirical inquiry (see, e.g., Jantke, 1979; Shapiro, 1998); in linguistics, the framework has been used to model language learning in the context of finding an appropriate assignment of deep syntactic structures to syntactic representations (for discussion see Wexler & Cullicover, 1980).

Since we consider a different type of structure here, we have to change the definition of text.[4]

Definition 2.1.17. *A text of a function, ε, is any infinite sequence over $\mathbb{N} \times \mathbb{N}$ (any infinite sequence of pairs of numbers), such that for each $x \in \mathbb{N}$ there is exactly one y such that (x, y) occurs in the sequence. In other words a text of a function g is any enumeration of the content of the graph of g.*

For text of functions we will use the notation introduced in Definition 2.1.2.

Let us take $\mathcal{C}_f$ to be a class of total recursive functions. For each $g \in \mathcal{C}_f$ we consider Turing machines φ_n which compute g. We take $I_g = \{n \mid \varphi_n \text{ computes } g\}$. Let us now assume that $g \in \mathcal{C}_f$ and ε is a text for g. We specify function identification in the limit by the following definition.

Definition 2.1.18 (Identification in the limit of functions). *We say that a learning function L:*

1. *identifies a function $g \in \mathcal{C}_f$ in the limit on ε iff for co-finitely many m, $L(\varepsilon{\restriction}m) = k$ and $k \in I_g$;*
2. *identifies $g \in \mathcal{C}$ in the limit iff it identifies g in the limit on every ε for g;*
3. *identifies $\mathcal{C}$ in the limit iff it identifies every $h \in \mathcal{C}$ in the limit.*

Function learning differs from language learning in many respects. One of the most important differences lies in the specific properties of possible realities—functions. Namely, environments of functions carry more information than streams of data defined for set learning. In an environment for a total function it is enough to examine a finite fragment of the environment to decide whether a given pair (n, m) is in the whole sequence. That is so because in some finite fragment we can find either (n, m) itself or some (n, m') with $m \neq m'$. In the latter case it follows that (n, m) is not in the sequence. In language learning it is impossible to conclude the non-existence of an element in an environment on the basis of finite examination. This allows the class of all recursive functions to be identifiable in the limit (see Jain et al., 1999). Let us also mention that totality of functions implies that for every n, there is an m, such that (n, m) is an element of an

[4] Similarly to the case of set learning, we take a text to be a positive presentation of a function. We are not concerned here with negative information at all.

environment. Therefore, it makes little difference to the learning if the function is enumerated in order $(g(0), g(1), \ldots)$. In that case learning is equivalent to the ability to guess the next value of the function after a certain time.

2.2 Logics of Knowledge and Belief

Modal logics of epistemic change are used to analyze the information flow in multi-agent systems (see, e.g., Baltag et al., 1998; Van Benthem, Van Eijck, & Kooi, 2006; Gerbrandy, 1999a,b). The approach of *dynamic epistemic logic*, DEL for short, (Plaza, 1989, see also Van Ditmarsch et al., 2007 for a handbook presentation) focuses on formalizing the principles of such epistemic changes.

2.2.1 Epistemic Logic

Let us begin with the notion of epistemic model. In what follows $\mathcal{A} = \{1, \ldots, n\}$ is a finite set of agents and PROP is a countable set of propositional letters.

Definition 2.2.1. *An* epistemic model $\mathcal{M}$ *based on a set of agents* $\mathcal{A}$ *is a triple:*

$$(W, (\sim_i)_{i \in \mathcal{A}}, V),$$

where $W \neq \emptyset$ *is a set of states, for each* $i \in \mathcal{A}$, $\sim_i$ *is a binary equivalence relation on* W, *and* $V : \text{PROP} \to \mathcal{P}(W)$ *is a valuation.*

A pair $(\mathcal{M}, w)$, *where* $\mathcal{M} = (W, (\sim_i)_{i \in \mathcal{A}}, V)$ *is an epistemic model and* $w \in W$, *will be called a* pointed epistemic model.

The information that agent i possesses in state w is denoted by

$$K_i[w] = \{v \in W \mid w \sim_i v\}.$$

It stands for all information within the uncertainty range of agent i with respect to w. Accordingly, the knowledge of agent i in state w consists of those statements that are true in all worlds he considers possible from state w. To explicitly talk about knowledge we will use the language of basic epistemic logic (see, e.g., Blackburn, Rijke, & Venema, 2001).

Definition 2.2.2 (Syntax of $\mathcal{L}_{\text{EL}}$). *The syntax of epistemic language* $\mathcal{L}_{\text{EL}}$ *is defined as follows:*

$$\varphi := \; p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi$$

where $p \in \text{PROP}$, $i \in \mathcal{A}$. *We will write* $\top$ *for* $p \vee \neg p$ *and* $\bot$ *for* $\neg\top$.

Definition 2.2.3 (Semantics of $\mathcal{L}_{\text{EL}}$). *We interpret* $\mathcal{L}_{\text{EL}}$ *in the states of epistemic models as follows.*

$$\begin{array}{lll}
\mathcal{M}, w \models p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, w \models \varphi \\
\mathcal{M}, w \models \varphi \vee \psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models K_i\varphi & \text{iff} & \text{for all } v \text{ such that } w \sim_i v \text{ we have } \mathcal{M}, v \models \varphi
\end{array}$$

Let us now provide an axiomatic system for epistemic logic EL (see, e.g., Blackburn et al., 2001).

PL	$\vdash \varphi$ if φ is a substitution instance of a tautology of propositional logic
Nec	if $\vdash \varphi$, then $\vdash K_i\varphi$
K	$\vdash K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$
T	$\vdash K_i\varphi \to \varphi$
4	$\vdash K_i\varphi \to K_iK_i\varphi$
5	$\vdash \neg K_i\varphi \to K_i\neg K_i\varphi$
MP	if $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$

Theorem 2.2.4. *The axiomatic system* EL *is complete with respect to the class of epistemic models.*

Epistemic Update

Epistemic models are static—they represent the informational state of an agent in temporal isolation. We will now make the setting more dynamic by assuming that agents observe some incoming data and are allowed to revise their informational states. We will consider *update* (see Van Benthem, 2007)—a policy that restricts models; each time a piece of data is encountered, it is assumed to be truthful and all worlds of the epistemic model that do not satisfy this new information are eliminated. The definition below formalizes the notion of update with a formula φ.

Definition 2.2.5. *The update of an epistemic model* $\mathcal{M} = (W, (\sim_i)_{i \in \mathcal{A}}, V)$ *with a formula* φ, *restricts* $\mathcal{M}$ *to those worlds that satisfy* φ, *formally* $\mathcal{M} \mid \varphi = \mathcal{M}' := (W', (\sim'_i)_{i \in \mathcal{A}}, V')$,

1. $W' = \{w \in W \mid w \models \varphi\}$;
2. *for each* $i \in \mathcal{A}$, $\sim'_i = \sim_i \upharpoonright W'$;
3. $V' = V \upharpoonright W'$.

Obviously, the incoming information that triggers update need not be propositional, not even purely linguistic. It can be any *event* that itself has an epistemic structure.[5] Below we consider a quite challenging case of an update with epistemic information.

[5] To consider changes caused by such arbitrary events, the notion of *event model* and *product update* has been introduced (Baltag et al., 1998). The former represents the epistemic content of an event, the latter stands for combining an epistemic model with an event model.

Muddy Children We want to devote some space to the classical logical puzzle which received a considerable amount of attention in dynamic epistemic logic (see, e.g., Van Ditmarsch et al., 2007; Gerbrandy, 1999a; Moses, Dolev, & Halpern, 1986). We discuss it here to give a flavor of complicated epistemic reasoning that can be successfully analyzed within DEL framework. We will return to the puzzle in the last chapter of this thesis, where we also propose a novel representation of this problem.

Example 2.2.6 (Muddy Children Puzzle). *The children, who were playing outside for a while, are called back in by their father. Some of them are dirty, in particular they have mud on their foreheads. The father decides to play with them and says:*

(1) *At least one of you has mud on your forehead.*

And immediately after, he asks:

(I) *Can you tell for sure whether or not you have mud on your forehead? If yes, step forward and announce your status.*

Each child can see the mud on others but cannot see his or her own forehead. Nothing happens. After that the father repeats **I**. *Still nothing. But after he repeats the question three times suddenly all children know whether or not they have mud on their forehead. How is that possible?*

The framework of dynamic epistemic logic allows a clear and comprehensive explanation of the underlying phenomena. Let us briefly explain the classical modeling. Assume there are three children, let us call them a, b and c, and assume that, in fact, all of them are muddy. We will take three propositional letters m_a, m_b and m_c that express that the corresponding child is muddy. The initial epistemic model of the situation is depicted in Figure 2.1.

In the model, possible worlds correspond to the 'distribution of mud' on children's foreheads, e.g., $m_a, \neg m_b, \neg m_c$ stands for a being muddy and b and c being clean. Two worlds are joined with an edge labeled with x, if the two worlds are in the uncertainty range of agent x (i.e., if agent x cannot distinguish between the two worlds). We drop the reflexive arrows for each state for clarity of the presentation. The boxed state stands for the actual world. Now, let us see what happens after the first announcement is made.

(1) At least one of you has mud on your forehead.

In propositional logic, this statement has the following form: (1') $m_a \vee m_b \vee m_c$. Since the children trust their father, they all eliminate world w_8 in which (1') is false: none of the children is muddy. In other words, they perform an update with formula (1'). The result is depicted in Figure 2.2.

Now the father asks for the first time:

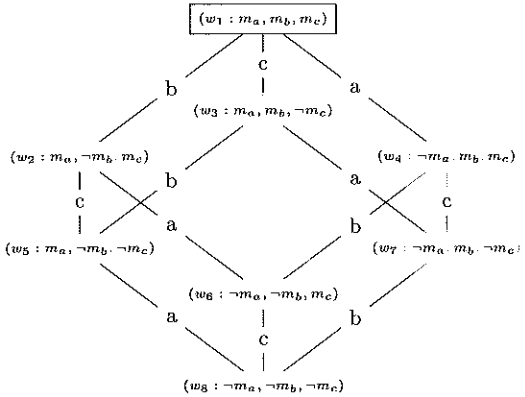
$(w_1 : m_a, m_b, m_c)$
c
b
a
$(w_3 : m_a, m_b, \neg m_c)$
$(w_2 : m_a, \neg m_b, m_c)$
$(w_4 : \neg m_a, m_b, m_c)$
b
a
c
c
a
b
$(w_5 : m_a, \neg m_b, \neg m_c)$
$(w_7 : \neg m_a, m_b, \neg m_c)$
$(w_6 : \neg m_a, \neg m_b, m_c)$
a
c
b
$(w_8 : \neg m_a, \neg m_b, \neg m_c)$

Figure 2.1: Initial epistemic model of the Muddy Children puzzle

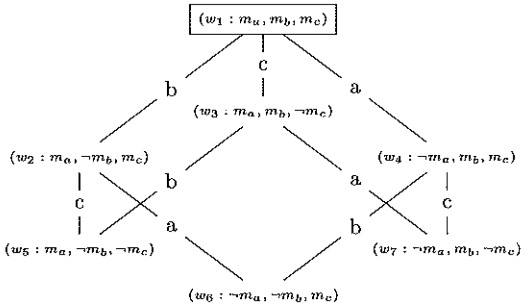
$(w_1 : m_a, m_b, m_c)$
c
b
a
$(w_3 : m_a, m_b, \neg m_c)$
$(w_2 : m_a, \neg m_b, m_c)$
$(w_4 : \neg m_a, m_b, m_c)$
b
a
c
c
a
b
$(w_5 : m_a, \neg m_b, \neg m_c)$
$(w_7 : \neg m_a, m_b, \neg m_c)$
$(w_6 : \neg m_a, \neg m_b, m_c)$

Figure 2.2: Epistemic model after father's announcement

(I) Can you tell for sure whether or not you have mud on your head?

The agents' reasoning can be as follows. In world w_6 agent c knows that he is dirty (there is no uncertainty of agent c between this world and another in which he is clean). Therefore, if the actual world was w_6, agent c would know his state and announce it. The same holds for agents a and b and worlds w_5 and w_7, respectively. But in our story children stay silent. This is in fact equivalent to the announcement that none of the children know whether they are muddy or not. Formally: $\neg(K_a m_a \vee K_a \neg m_a) \wedge \neg(K_b m_b \vee K_b \neg m_b) \wedge \neg(K_c m_c \vee K_c \neg m_c)$. Now all agents eliminate those worlds that do not satisfy this formula: w_5, w_6, w_7. The epistemic model of the next stage is smaller by three worlds (Figure 2.3).

At this stage it is again clear that if one of the w_2, w_3, w_4 was the actual state the respective agent would have announced their knowledge. But in our scenario

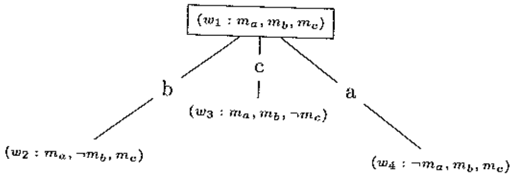
$(w_1 : m_a, m_b, m_c)$
c
b
a
$(w_3 : m_a, m_b, \neg m_c)$
$(w_2 : m_a, \neg m_b, m_c)$
$(w_4 : \neg m_a, m_b, m_c)$

Figure 2.3: Epistemic model in the second stage of epistemic inference

the children still do not respond. Then the father asks again: 'Can you tell for sure whether or not you have mud on your forehead?'. Now the children base their inference on the silence in the previous step, and come to the conclusion that the actual situation cannot be any of w_2, w_3, w_4. So, they all eliminate the three states, which leaves them all with just one possibility (Figure 2.4). All uncertainty disappears and they all know that they are dirty.

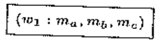

Figure 2.4: Epistemic model in the third stage of epistemic inference

Public Announcement

All announcements made in the above scenario trigger an update of the epistemic model according to Definition 2.2.5. The public character of the announcements makes them influence all agents' uncertainty ranges. Basic epistemic logic, as defined above, can be extended to account for this type of update with a specific 'action' expression of *public announcement*, written as $!\varphi$.

Definition 2.2.7 (Syntax of $\mathcal{L}_{\text{PAL}}$). *The syntax of epistemic language $\mathcal{L}_{\text{PAL}}$ is defined as follows:*

$$\begin{aligned} \varphi &:= \ p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid [A]\varphi \\ A &:= !\varphi \end{aligned}$$

where $p \in$ Prop, *$i \in \mathcal{A}$.*

Definition 2.2.8 (Semantics of $\mathcal{L}_{\text{PAL}}$). *For the epistemic fragment $\mathcal{L}_{\text{EL}}$ the interpretation is given in Definition 2.2.3. The remaining clause of $\mathcal{L}_{\text{PAL}}$ is as follows.*

$$\mathcal{M}, w \models [!\varphi]\psi \quad \text{iff} \quad \text{if } \mathcal{M}, w \models \varphi \text{ then } \mathcal{M} \mid \varphi, w \models \psi$$

An axiomatization PAL of $\mathcal{L}_{\text{PAL}}$ can be composed of the previously given axioms of epistemic logic enriched with the following reduction axioms (Plaza, 1989).

1 $\vdash [!\varphi]p \leftrightarrow (\varphi \to p)$, for $p \in$ Prop
2 $\vdash [!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$
3 $\vdash [!\varphi](\psi \vee \xi) \leftrightarrow ([!\varphi]\psi \vee [!\varphi]\xi)$
4 $\vdash [!\varphi]K_i\psi \leftrightarrow (\varphi \to K_i[!\varphi]\psi)$

Theorem 2.2.9 (Plaza 1989). *The axiomatic system* PAL *is complete with respect to the class of epistemic models.*

The change that epistemic models undergo when subjected to public announcement corresponds to the revision with so-called 'hard' information. Such a revision is reasonable if the information originates from a reliable source.

2.2.2 Doxastic Logic

The notion of irrevocable knowledge defined in the previous subsection is very strong. It implicitly indicates that unless complete certainty is reached, the agent does not form any opinion on the state of the world. In order to talk about weaker informational states, like belief, epistemic models have to be modified to account for the order on states given by agents' doxastic attitudes.

Definition 2.2.10 (Baltag & Smets 2006). *An* epistemic-plausibility model *$\mathcal{M}$ is a triple*

$$(W, (\sim_i)_{i \in \mathcal{A}}, (\leq_i)_{i \in \mathcal{A}}, V),$$

where $W \neq \emptyset$ is a set of states, for each $i \in \mathcal{A}$, $\leq_i$ is a total well-founded preorder[6] *on W, and $V :$ Prop $\to \mathcal{P}(W)$ is a valuation.*

A pair $(\mathcal{M}, w)$, where $\mathcal{M} = (W, (\sim_i)_{i \in \mathcal{A}}, (\leq_i)_{i \in \mathcal{A}}, V)$ an epistemic plausibility model and $w \in W$, is called a pointed epistemic plausibility model.

For each $i \in \mathcal{A}$ we will assume that $\leq_i \subseteq \sim_i$.

Now the language of epistemic logic can be extended to account for belief.

Definition 2.2.11 (Syntax of $\mathcal{L}_{\text{DOX}}$). *The syntax of doxastic-epistemic language $\mathcal{L}_{\text{DOX}}$ is defined as follows:*

$$\varphi := \ p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid B_i^{\psi}\varphi$$

where $p \in$ Prop, *$i \in \mathcal{A}$.*

[6] A preorder is a binary relation that is reflexive and transitive. Later we will relax the restriction to well-founded preorders and adjust the relevant definitions.

Definition 2.2.12 (Semantics of $\mathcal{L}_{\mathrm{DOX}}$). *We interpret $\mathcal{L}_{\mathrm{DOX}}$ in the states of doxastic-epistemic models in the following way.*

$$\begin{array}{lll}
\mathcal{M}, w \models p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \text{it is not the case that } \mathcal{M}, w \models \varphi \\
\mathcal{M}, w \models \varphi \vee \psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models K_i\varphi & \text{iff} & \text{for all } v \text{ such that } w \sim_i v \text{ we have } \mathcal{M}, v \models \varphi \\
\mathcal{M}, w \models B_i^{\psi}\varphi & \text{iff} & \text{for all } v \in \mathcal{K}_i[w] \text{ if } v \in \min_{\leq_i}(\mathcal{K}_i[w] \cap \|\psi\|) \text{ then } v \models \varphi
\end{array}$$

We define $\|\varphi\|$ such that $\|\varphi\| = \{w \in W \mid w \models \varphi\}$.

The last clause defines the semantics of the conditional belief operator. An agent is defined to believe φ in state w conditionally on ψ if φ is true in all states that are minimal in the part of the uncertainty range of the agent restricted to those states that make ψ true.

For axiomatizations of $\mathcal{L}_{\mathrm{DOX}}$ the reader is advised to consult (Board, 2004) and (Baltag & Smets, 2008b).

Plausibility Upgrade

Epistemic plausibility models can accommodate public announcements of hard information. Performing update on those structures has an effect analogous to restriction of simple epistemic models. Such a change can of course result in belief change. However, plausibility ordering gives an opportunity to define different, more sophisticated operations on beliefs, operations that do not require state deletion. As we will see in Chapter 4, such revisions are useful if the source of information is not completely trustworthy.

Lexicographic Upgrade The *lexicographic upgrade* of an epistemic plausibility model $\mathcal{M} = (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i)_{i\in\mathcal{A}}, V)$ with a formula φ, rearranges the preorders by putting all states satisfying φ to be more plausible then others. Let us take $\leq_i^{\varphi} = \leq_i\restriction \|\varphi\|$, and $\leq_i^{\bar{\varphi}} = \leq_i\restriction \|\neg\varphi\|$.

Definition 2.2.13. *The* lexicographic upgrade *of an epistemic plausibility model $\mathcal{M} = (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i)_{i\in\mathcal{A}}, V)$ with a formula φ is defined as follows:*

$$\mathcal{M} \Uparrow \varphi := (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i')_{i\in\mathcal{A}}, V),$$

where for each $i \in \mathcal{A}$ and for all $v, w \in \mathcal{K}_i[w]$:

$$v \leq_i' w \text{ iff } (v \leq_i^{\varphi} w \text{ or } v \leq_i^{\bar{\varphi}} w \text{ or } (v \models \varphi \text{ and } w \models \neg\varphi)).$$

The language of announcements that trigger lexicographic upgrade is given in the following way.

Definition 2.2.14 (Syntax of $\mathcal{L}_{\Uparrow}$). *The syntax of the doxastic-epistemic language $\mathcal{L}_{\Uparrow}$ is defined as follows:*

$$\begin{array}{l}
\varphi := p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid B_i^{\psi}\varphi \mid [A]\varphi \\
A := \Uparrow\varphi
\end{array}$$

where $p \in$ Prop, $i \in \mathcal{A}$.

Definition 2.2.15 (Semantics of $\mathcal{L}_{\Uparrow}$). *For the doxastic-epistemic fragment $\mathcal{L}_{\mathrm{DOX}}$ the interpretation is given in Definition 2.2.12. The remaining clause of $\mathcal{L}_{\Uparrow}$ is as follows.*

$$\mathcal{M}, w \models [\Uparrow\varphi]\psi \quad \text{iff} \quad \mathcal{M} \Uparrow \varphi, w \models \psi$$

Conservative Upgrade The *conservative upgrade* (also known as *minimal upgrade* or *elite change*, see Van Benthem, 2007) of an epistemic plausibility model $\mathcal{M} = (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i)_{i\in\mathcal{A}}, V)$ with a formula φ, rearranges the preorders by making only the most plausible states satisfying φ more plausible than all others, leaving the rest of the preorder the same. Let $\leq_i^{\mathrm{rest}\varphi} = \leq_i \restriction \{t \in S \mid t \notin \min_{\leq_i} \|\varphi\|\}$.

Definition 2.2.16. *The* conservative upgrade *of an epistemic plausibility model $\mathcal{M} = (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i)_{i\in\mathcal{A}}, V)$ with a formula φ is defined as follows:*

$$\mathcal{M}{\uparrow}\varphi := (W, (\sim_i)_{i\in\mathcal{A}}, (\leq_i')_{i\in\mathcal{A}}, V),$$

where for each $i \in \mathcal{A}$ and for all $v, w \in \mathcal{K}_i[w]$:

$$v \leq_i' w \text{ iff } (v \leq_i^{\mathrm{rest}\varphi} w \text{ or } v \in \min_{\leq_i}\|\varphi\|).$$

Definition 2.2.17 (Syntax of $\mathcal{L}_{\uparrow}$). *The syntax of the doxastic-epistemic language $\mathcal{L}_{\uparrow}$ is defined as follows:*

$$\begin{array}{l}
\varphi := p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid B_i^{\psi}\varphi \mid [A]\varphi \\
A := \uparrow\varphi
\end{array}$$

where $p \in$ Prop, $i \in \mathcal{A}$.

Definition 2.2.18 (Semantics of $\mathcal{L}_{\uparrow}$). *For the doxastic-epistemic fragment $\mathcal{L}_{\mathrm{DOX}}$ the interpretation is given in Definition 2.2.12. The remaining clause of $\mathcal{L}_{\uparrow}$ is as follows.*

$$\mathcal{M}, w \models [\uparrow\varphi]\psi \quad \text{iff} \quad \mathcal{M}{\uparrow}\varphi, w \models \psi$$

Complete axiomatization for the logics of the two types of upgrades can be given by a complete axiomatic system for conditional belief complemented with reduction axioms. Van Benthem (2007) gives a detailed discussion on the subject, together with explicitly formulated axioms.

In Chapter 4 we will cover these upgrade methods again in a systematic way. We will compare their reliability in the context of single-agent belief-revision. In this, we will follow other attempts to analyze some classical belief-revision problems within the framework of dynamic epistemic and doxastic logic.

Chapter 3
Learning and Epistemic Change

In the present chapter we show how the paradigms of learning theory and dynamic epistemic logic can be linked. We will discuss the interface between learning theory and dynamic epistemic logic in the context of iterated information change and belief revision.

3.1 Identification as an Epistemic Process

In Chapter 2 we gave the prerequisites of formal learning theory with its central notion of identification. Assuming the reader's familiarity with those standard tools, we will now discuss the epistemology behind finite and limiting identification.

What are the epistemic components of identification in the limit? The entanglement of the notions of knowledge, certainty and belief in limiting learning is widely used in explanations of the paradigm. We quote Gold (1967) in his seminal paper *Language identification in the limit*:

> In the case of identifiability in the limit the learner does not necessarily *know*[1] when his guess is correct. He must go on processing the information forever because there is always the possibility that information will appear which will force him to change his guess.

With time the epistemic metaphor in identification in the limit became even more explicit, involving notions of certainty, justification, possible worlds, etc.:

> [...] Thus the Scientist is never *justified* in feeling *certain* that her last conjecture will be her last.
>
> On the other hand, [identifiability in the limit] does warrant a different kind of *confidence*, namely that systematic application of guessing rule will eventually lead to an accurate, last conjecture [...]. If we *know*

[1] The emphasis is mine.

> that the *actual world* is drawn from [a class identifiable in the limit], then we can be *certain* that our inquiry will ultimately succeed [...]. (Jain et al., 1999, pp. 11–12)

Later, even notions of introspection of knowledge, belief and reliability were introduced:

> This does not entail that [the learner] *knows he knows* the answer, since [...] [the learner] may lack any reason to *believe* that his hypotheses have begun to converge. Nonetheless, to the extent that the *reliability perspective on knowledge* can be sustained, our paradigms concern scientific discovery in the sense of *acquiring knowledge*. (Martin & Osherson, 1998, p. 13)

Finally, the epistemic dominance of limiting identification over certainty has been once summed up in the following way:

> True, there are good reasons for preferring the computable way of deriving *knowledge*. We *know* the results of computations, and only *think we know* the results of trial and error procedures [viz. limiting computation]. There are many reasons for *preferring knowing to thinking* (as Popper, 1966, observed). But that does not change the fact that sometimes *thinking* may be more appropriate. (Kugel, 1986, p. 155)

Our aim is to expose the epistemology that runs the limiting learning process from behind the scenes. Let us start by overviewing the components of identification and discussing their correspondence with the approach of epistemic logic as described in Chapter 2.

Class of hypotheses The procedure of learning starts with a class of hypotheses, a class of possible states of the world. It can be interpreted as the background knowledge of Scientist, his uncertainty range (see, e.g., Martin & Osherson, 1997). Scientist expects that one of the possibilities is true, and in the framework it is guaranteed that he is right—Nature indeed chooses one from the class fixed in the beginning. Among the consequences of such a treatment of background knowledge is that the actual world is always one of the options Scientist considers possible. Another implication is that learning is not simply verifying or falsifying a single hypothesis, although those two processes can be viewed as important components of identification (Gierasimczuk, 2009b). The fact of picking *one from a class* is an important factor in learnability analysis. It allows considering learnability as a property of classes of hypotheses determined by some external properties.

Different nature of data and conclusions The key word "learning" is often used in the context of belief revision and dynamic epistemic logic. There it takes the form of one-step "learning that φ", followed by a modification of the informational state of the agent—usually by various ways of simply accepting φ as it is. In other words, the agent "learned that φ" means that the agent "got to know that φ". In the setting of formal learning theory it requires more effort than that to be declared to have learned something. First of all, the incoming information is by default spread over more than one step. The inductive, step-by-step nature of this inference is essential; the incoming pieces of data are of a different nature than the actual 'thing' being learned. Typically, at each finite step the environment gives only partial information about a potentially infinite set. The relationship between data and hypothesis is like the one between sentences and grammars, natural numbers as such and Turing machines. Namely, if we know the hypothesis, we can infer what kind of possible data are going to appear, but in principle we will not be able to make a conclusive inference from data to hypotheses. Therefore, in learning theory we say that an agent "learned that a hypothesis holds" if he converged to this hypothesis on data that are consistent with the actual world.

Positive, true, and readable data There are three important assumptions that the incoming data can satisfy:

1. Truthfulness (soundness). Scientist receives only true data, no false information is included. This assumption leads to, e.g., the priority of incoming data over the current conjecture and background preferences of Scientist.

2. Positiveness. Scientist receives only elements of positive presentation (*text*) of the object being learned. Alternatively, together with positive also all negative information could be included (*informant*), e.g., for set learning the graph of the characteristic function of the set could be enumerated.

3. Readability. Scientist has a complete clarity about what information he receives. A further step would be to analyze the situation of uncertainty about the incoming information.

4. Completeness. The data that are consistent with the actual world are all eventually enumerated.

In formal learning theory it is usually assumed that the incoming information is readable and complete. The source of data is also taken to be truthful. Occasional errors are rarely taken into account, and in more applied disciplines are interpreted as noise (see, e.g., Grabowski, 1987). In contrast, the general epistemic framework allows erroneous information in form both of mistakes and intentional lies. In this respect the original learning theory conforms more to the assumptions of the philosophy of scientific inquiry (Nature never lies) than to, e.g., conversational situations (see, e.g., Grice, 1975). Another classic requirement put on data is that it is positive, i.e., data enumerates only elements of the language. This assumption is often challenged by involving negative information, data indicating which elements are *not* in the set. This setting boosts the power of learning immensely (see Gold, 1967). It should be noted here that data including both positive and negative samples gives remedy to errors. There is enough expressive power so that any information inconsistent with the actual world can be accounted for truthfully later on in the process.

Inductive, step-by-step process As briefly mentioned in the previous points, the process of restricting the hypothesis space to only those hypotheses that are consistent with the incoming data resembles update or public announcement (Baltag et al., 1998). Can learning in the limit of hypothesis h be viewed as the result of announcing the conjunction of data that lead to stabilization on h? First let us observe that the point of convergence to a correct hypothesis is unknown and in general uncomputable, which makes it also uncomputable to discover which finite sequence resulted in the success of the learning process. Even more importantly, finite sequences of data cannot be seen as a single announcement of a given hypothesis, because which hypothesis is in fact announced by the data heavily depends on the initial hypothesis space. For instance, let us consider two classes: $\mathcal{C}_1 = \{\{1\}, \{2\}\}$ and $\mathcal{C}_2 = \{\{1\}, \{1, 2\}\}$, and let h_1 be a hypothesis corresponding to the set $\{1\}$. In this case the single event of updating with 1 is equivalent to announcing h_1 in case $\mathcal{C}_1$ had been the initial set of hypotheses, but it does not announce h_1 when Scientist has to pick from $\mathcal{C}_2$, since the other hypothesis is still possible.

Infinite procedures The learning theory framework is defined for potentially infinite universes, but even for finite worlds the sequences of data are infinite. The reason for this is that we want to account for situations when Scientist does not know the finiteness or size of the entity he investigates. If the initial class of hypotheses is not drastically restrictive, Scientist can never know whether all the elements have already been enumerated. This leads to infinite procedures and conditions defined in the limit. Our epistemic setting should reflect these properties. It should allow talking about epistemic states as invariant from some point onwards, without specifying when this happens. Such an approach to learning is not unheard of in epistemic logic and belief-revision. There is an ongoing philosophical debate about iterated belief revision, iterated epistemic update, stability of knowledge, etc. (see, e.g., Stalnaker, 2009). As we will show they directly correspond to our limiting processes.

Non-introspective knowledge The success of limiting learning can be defined as reaching an epistemic state that can be called 'knowledge'. What kind of knowledge is it? On the surface it seems to be pretty close to the classical justified true belief (see, e.g., Chisholm, 1982), the definition ascribed to Plato. Indeed, eventually Scientist puts forward a hypothesis that is true, he believes that it is true, and moreover he has some reasons to choose it and those reasons can be viewed as a (often very limited) justification. However, from the perspective of the agent this 'knowledge', preceded by a sequence of belief changes, is strictly operational, the work is always in progress. There seems to be some issue with introspection here—Scientist is not able to point out the successful guess, he does not know whether he will not be forced to change his guess again in the light of future data (for the discussion of the introspection of knowledge in inductive inference see Hendricks, 2003). On the other hand it is more than just a true belief—it is immune to change under new true information.

Single agent As mentioned before, in learning theory the data are assumed to be complete and true. In our view, this is the reason why learners are pretty lonely in this paradigm. Although in principle, science as well as learning seem to be at least a two-player game that includes a teacher and a learner (a sender of the information and a receiver), for many algorithmic reasons the role of the former has been minimized. As a result we are concerned here only with the role of Scientist. Nature can be viewed as an objective, uninvolved source of data. In a sense this constitutes an assumption of fairness. Nature does not intend to help or disturb the process. As a result, learning theory is predominantly a one-agent business. A hint of multi-agency can be associated with team-learning, a framework suggested by Blum & Blum (1975), explicitly introduced by Smith (1982) and since then extensively studied (for an overview see Jain & Sharma, 1996). However, multi-agency understood in this way can be summed up as learners working on their own contributing to some common, bigger goal. The topic of communication and (non-)cooperativeness of the learners is marginal here. Dynamic epistemic and dynamic doxastic logics study these notions of multi-agency explicitly, and this is in fact their main focus (for the benefits of a multi-agent approach to epistemic issues see Van Benthem, 2006).

3.2 Learning via Updates and Upgrades

With the above discussion in mind we can now turn to the question of how learning-theoretic notions can be approached from the perspective of the epistemic framework.[2]

Let us fix $\mathcal{C} = \{S_1, S_2, \ldots\}$ to be a class of sets. It can be interpreted as the initial epistemic model, representing the background knowledge of Scientist

[2] Our considerations are of semantic nature and therefore differ from the computable framework of learning theory. E.g., one of the consequences is that we assume the property of consistency of learning, which in formal learning theory is optional.

together with his uncertainty about which world is the actual one. Let us take the initial epistemic model to be formally defined as

$$\mathcal{M} = (\mathcal{C}, \sim),$$

where $\mathcal{C}$ is the set of worlds and $\sim \subseteq \mathcal{C} \times \mathcal{C}$ is an uncertainty relation for Scientist. For now we do not require any particular preference of the scientist over $\mathcal{C}$—all possibilities are equally plausible. Hence, we can for now assume that $\sim$ is a universal equivalence relation over $\mathcal{C}$. The initial epistemic state of the Scientist is depicted in Figure 3.1. This model corresponds to the starting point of the scientific discovery process. In the beginning Scientist considers all of them possible. Scientist is *given* the class of hypotheses $\mathcal{C}$, i.e., he knows what the alternatives are.

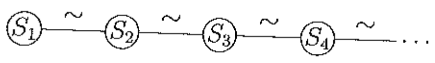

Figure 3.1: Initial epistemic model

Next, Nature decides on some state of the world by choosing one possibility from $\mathcal{C}$. Let us assume that, as a result, S_4 is the chosen world. Then, she decides on some particular environment ε, consistent with S_4. We picture this enumeration in Figure 3.2 below.

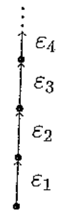

Figure 3.2: Environment ε consistent with S_4

The sequence ε is successively given to Scientist. Let us focus now on the first step of the procedure. A piece of data ε_1 is given to the scientist. In Figure 3.3 Scientist's confrontation with ε_1 is depicted. Scientist can react to this new information by adjusting his epistemic state in different ways.

3.2.1 Learning via Update

Epistemic Update

One way for Scientist to incorporate a new piece of data is to *update*[3] his status with ε_1. This is done by eliminating all the sets that do not include ε_1. We can represent the process formally by the update of $\mathcal{M}$ with ε_1, $(\mathcal{M} \mid \varepsilon_1)$, resulting in a new epistemic model $\mathcal{M}' = (\mathcal{C}', \sim')$, where: $\mathcal{C}' = \{S_n \in \mathcal{C} \mid \varepsilon_1 \in S_n\}$ and $\sim' = \sim\restriction \mathcal{C}'$.

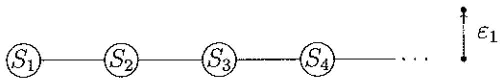

Figure 3.3: Confrontation with data

Scientist tests $\mathcal{C}$ with ε_1. If a set includes the information, it remains as a possibility, if it does not, it is eliminated (see Figure 3.4). Let us assume that ε_1 is not consistent with S_1 and S_3.

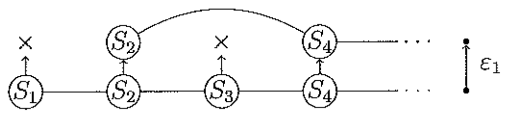

Figure 3.4: Epistemic update

This epistemic update can be iterated infinitely many times along ε resulting in an infinite sequence of models whose result according to the lines of DEL can be called the ε-Generated Epistemic Model (see, e.g., Van Benthem, Gerbrandy, Hoshi, & Pacuit, 2009).

Definition 3.2.1 (Generated epistemic model). *The generated epistemic model* $\mathcal{M}^{\varepsilon}$, *with* $\varepsilon = \varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$, *is the result of update* $(((\mathcal{M} \mid \varepsilon_1) \mid \varepsilon_2) \mid \varepsilon_3) \mid \ldots$

To stay true to our original learning-theoretic motivation we want to investigate how the epistemic model changes when ε is given in a stepwise fashion. In particular, we would like to focus on its convergence properties. Our modeling involves only the equivalence relation, which mirrors not only the agent's uncertainty, but also indifference with respect to what is the actual world. This approach is especially and, we could argue, exclusively suited for interpreting the rise of irrevocable knowledge. That is, the agent is said to know something if this

[3] The event of update is a simple single-agent version of public announcement (Baltag et al., 1998).

something is true in all worlds in his uncertainty range defined by the equivalence relation. Therefore, we will be particularly interested in the convergence to the state of such knowledge, i.e., in our case in convergence to the situation in which only one, true set is left. Then we will say that the scientist learned with certainty what is the actual world. The possibility of reaching certainty in an epistemic model by the use of updates resembles the setting of finite identifiability. To recall the latter let us give a short example.

Example 3.2.2. *Let us take $\mathcal{C} = \{S_1, S_2, S_3\}$, such that $S_n = \{1, ..., n\}$, for $n \in \{1, 2, 3\}$. Nature makes her choice regarding the identity of the world. Let us assume that, as a result, S_3 is the actual world. Then, Nature chooses an enumeration $\varepsilon = 1, 2, 1, 3, 2, \ldots$. After the first piece of data, 1, the uncertainty range of the scientist includes the whole $\mathcal{C}$. After the second, 2, the scientist eliminates S_1 since it does not contain the event 2 and now he hesitates between S_2 and S_3. The third piece, 1, does not change anything; however, the next one, 3, eliminates S_2. Uncertainty is eliminated. He knows that S_3 is the actual world. Therefore, we can say that he learned it conclusively, with certainty.*

In Chapter 5 we will show that finite identifiability can be modeled within the dynamic epistemic logic framework, with the use of: possible worlds for sets; propositions for the incoming information; and update for the progress in eliminating uncertainty over the hypothesis space.

Plausibility Update

The epistemic, update-based approach as set out above is very restrictive with respect to the outcome of learning. At best, we have been able to account only for finite identification, and not for learning in the limit. In order to move to identification in the limit we need to be able to talk about *sequences of conjectures* of Scientist. Until now this was impossible because the only 'conjecture' that we were able to define was a final irrevocable conclusion. So we want to enrich the framework to account for a *current conjecture*--a hypothesis that is considered appropriate in a given step of the procedure.

Let us consider the following example of a learning scenario, in which the uncertainty is never eliminated.

Example 3.2.3. *In Example 3.2.2 Scientist was very lucky. Let us assume for a moment that nature had chosen $S_2 = \{1, 2\}$, and had fixed the enumeration $\varepsilon = 1, 2, 1, 2, 2, 2, 2, \ldots$ In this case Scientist's uncertainty can never be eliminated.*

This example indicates that the central element of the identification in the limit model is the unavoidable presence of uncertainty. The limiting framework allows, however, introducing some kind of *operational* knowledge (for an account of procedural knowledge see Hoshi, 2009), that is expressed by the stability of the conjectures of the learning function.

To model an algorithmic nature of the learning process that includes the actual guess and other not-yet-eliminated possibilities, we can enrich the epistemic model with some plausibility relation. The relation $\leq$ represents some preference over the set of hypotheses. E.g., if Scientist is an Occamist, the preference would be defined according to the simplicity of hypotheses. In the initial epistemic state the uncertainty of the scientist again ranges over the whole of $\mathcal{C}$. This time however the class is ordered and Scientist's current belief is the most preferred hypothesis.[4] Therefore, we consider the initial epistemic plausibility state of Scientist to be:

$$\mathcal{M} = (\mathcal{C}, \sim, \leq).$$

The procedure of erasing worlds that are inconsistent with successively incoming data is the same as in the previous section. This time however let us introduce the current-guess state which is interpreted as the actual conjecture of the Scientist. It is always the one that is most preferred—the smallest one according to $\leq$. In doxastic logic a set of most preferred hypotheses is almost invariably interpreted as the ones that the agent *believes* in. Let us go back to Example 3.2.3, where Nature chose world S_2. After seeing 2 and eliminating S_1, Scientist's attention focuses on S_2; then S_2 is his current belief. It is the most preferred hypothesis, and as such can be repeated as long as it is consistent with ε. In this particular case, since Nature chose a world consistent with S_2, it will never be contradicted, so Scientist will always be uncertain between S_2 and S_3. However, his preference directs him to believe in the correct hypothesis, without his being aware of the correctness. The belief in a hypothesis may become safe—whatever true information is given, it will not force the scientist to change his mind. And this state of safety while maintaining uncertainty is intuitively the one that occurs in identification in the limit. According to the picture sketched here, we will show (in Chapter 4) that learning in the limit can be modeled within the dynamic doxastic logic framework, using: possible worlds for sets; propositions for incoming information; update for the progress in eliminating uncertainty over the hypothesis space; a plausibility relation for the underlying hypothesis space; in each step of the procedure, the most preferred hypothesis as the actual positive guess of the learning function.

3.2.2 Learning via Plausibility Upgrades

Extending this approach we will also investigate different ways of reacting to the incoming information: except for update we will also consider ways of upgrading the preference relation as a reaction to new data. Upgrades are useful when update is too strong—in the situations in which the source of information is not entirely

[4] For now we do not pose any restriction on the plausibility ordering. The conditions of well-foundedness or connectedness of the plausibility ordering are often assumed of such doxastic situations. As we will see later, in our setting, the well-foundedness of the initial plausibility preorder might not always be possible.

reliable. We want to focus on two types of upgrades: lexicographic and minimal (see Chapter 2). Upgrades can be performed on the epistemic plausibility models step-by-step as in the case of iterated update. Interpreting the minimal hypotheses as the ones that the agents believes in at any finite point of the procedure, again allows considering sequences of conjectures.

3.3 Learning as a Temporal Process

In the above-described paradigm each hypothesis from the given class is associated with the corresponding set of environments. The latter can be seen as possible "streams of events" or "histories" that may occur if the relevant hypothesis is true. A history can in its turn be represented as a branch in the tree of all possible courses of events. Accordingly, hypotheses can be viewed as sets of histories or trees. The intuitive way to deal with hypotheses in a temporal framework is to introduce a temporal model of all the possible streams of information determined by the hypothesis.

Let us consider a set $\mathcal{C} = \{\{1\}, \{1,2\}, \{1,2,3\}\}$ and the corresponding set of hypotheses $I_\mathcal{C} = \{h_1, h_2, h_3\}$. We know that h_1 corresponds to the set $\{1\}$, so it is consistent with only one environment $\varepsilon = 1, 1, 1, 1, 1, \ldots$ Therefore, it can be identified with only one possible sequence of events, history H, which is represented by the frame presented in Figure 3.5.

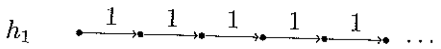

Figure 3.5: History for hypothesis h_1

It is of course different for the hypothesis h_2 which corresponds to the set $\{1,2\}$. Here, possible histories are all ω-sequences over the set $\{1,2\}$, that include at least one occurrence of 1 and 2. Therefore the hypothesis is represented as a binary tree.

Let us put this idea formally. If S is a set, then S^* is the set of all finite sequences over S (all finite strings of elements of S). Let us take a class $\mathcal{C}$ and $S_n \in \mathcal{C}$. The set S_n determines an epistemic temporal logic frame

$$\mathcal{F} = (S_n, H_n, \sim),$$

where $H_n = S_n^*$ is a protocol (says which sequences of events are allowed), that is closed under non-empty prefixes; and $\sim$ is a binary relation on H_n.

Such an epistemic temporal frame indicates which sequences of data can be expected when the corresponding hypothesis is true. This way of thinking allows viewing the class of hypotheses $\mathcal{C}$ as a set of protocols, a forest of temporal frames (see Figure 3.6).

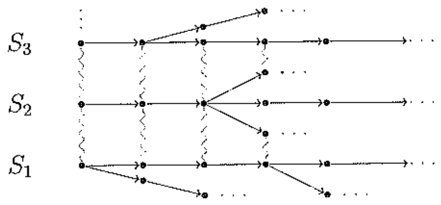

Figure 3.6: Epistemic temporal forest $\mathcal{F}$

To sum up, we interpret hypotheses to be sets of histories, i.e., sets of sequences enumerating events. Therefore, we can reinterpret the possible realities as *sets of functions*. This approach leads to a generalized, uniform view of learnability of various structures. Function learning and set learning become analyzable on a common ground.

To account for identification in the limit, following the argumentation of previous sections it seems to be necessary to enrich the temporal models with plausibility ordering that will account for the beliefs at each level of the temporal forest. The latter can be generated from the initial class as in the previous case. Then the temporal epistemic plausibility frame is given as follows:

$$\mathcal{F}_{\leq} = (S_n, H_n, \sim, \leq).$$

Our aim in all the above described semantic interpretations is to give an epistemic (temporal) characterization of learnability.

3.4 Summary

In this chapter we gave an introduction to our modeling of the process of inductive inference in dynamic epistemic logic and dynamic doxastic logic. For now we avoided formalism in order to first provide motivation and basics of the transition from one framework to another. In particular, we indicated that update is appropriate to analyze the notion of finite identifiability as convergence to knowledge. Learning in the limit, on the other hand, has to be supported by an underlying ordering of the hypothesis space. This indicates that it should be formalized in doxastic logic, where the preference or plausibility relation is a standard element of any model, and identification in the limit is viewed as reaching safe belief. We also proposed to view one component of the learning paradigm, hypotheses and hypothesis spaces, as temporal models. This allows investigating properties of the epistemic revision that requires certain *sequences* of events, conforming to some temporal protocols. We postulate that identifiability can be expressed in temporal logic interpreted over the corresponding epistemic temporal forests.