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.\footnote{For the characteristics and history of this line of research see, e.g., the Stanford Encyclopedia of Philosophy entry \textit{Formal Learning Theory} by Oliver Schulte.}

In philosophy of science and general epistemology there is no need to assume that theory change is governed by a \textit{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 \textit{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
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:


Chapter 4 is based on:


Chapter 5 is based on:


Chapter 6 is based on:


Chapter 7 is based on:


Chapter 8 is based on: