Synthese Library 420
Studies in Epistemology, Logic, Methodology, and Philosophy of Science

Holger Andreas

Dynamic Tractable Reasoning

A Modular Approach to Belief Revision

Springer

Synthese Library

Studies in Epistemology, Logic, Methodology,
and Philosophy of Science

Volume 420

The aim of *Synthese Library* is to provide a forum for the best current work in the methodology and philosophy of science and in epistemology. A wide variety of different approaches have traditionally been represented in the Library, and every effort is made to maintain this variety, not for its own sake, but because we believe that there are many fruitful and illuminating approaches to the philosophy of science and related disciplines.

Special attention is paid to methodological studies which illustrate the interplay of empirical and philosophical viewpoints and to contributions to the formal (logical, set-theoretical, mathematical, information-theoretical, decision-theoretical, etc.) methodology of empirical sciences. Likewise, the applications of logical methods to epistemology as well as philosophically and methodologically relevant studies in logic are strongly encouraged. The emphasis on logic will be tempered by interest in the psychological, historical, and sociological aspects of science.

Besides monographs *Synthese Library* publishes thematically unified anthologies and edited volumes with a well-defined topical focus inside the aim and scope of the book series. The contributions in the volumes are expected to be focused and structurally organized in accordance with the central theme(s), and should be tied together by an extensive editorial introduction or set of introductions if the volume is divided into parts. An extensive bibliography and index are mandatory.

More information about this series at http://www.springer.com/series/6607

Holger Andreas

Dynamic Tractable Reasoning

A Modular Approach to Belief Revision

Springer

Holger Andreas
Department of Economics, Philosophy,
and Political Science
University of British Columbia (Okanagan)
Kelowna, BC, Canada

Synthese Library
ISBN 978-3-030-36232-4 ISBN 978-3-030-36233-1 (eBook)
https://doi.org/10.1007/978-3-030-36233-1

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to Leonard

Modern logic is undergoing a cognitive turn,
side-stepping Frege's 'anti-psychologism'.

Johan van Benthem [112]

Preface

When working on theoretical terms during my PhD thesis, I became intrigued by the Sneed formalism and its set-theoretic predicates. This formalism struck me as a natural continuation of Carnap's logic of science. But I was not convinced by Stegmüller's arguments against a broadly axiomatic and syntactic approach to scientific theories, so I devised an axiomatic variant of the structuralist representation scheme of scientific knowledge.

When looking for applications of my work on structuralism (and opportunities for further funding), I started to explore connections to cognitive science and knowledge representation. I was excited to observe that set-theoretic concepts of the Sneed formalism may help us translate Minsky's ideas about frames into the language of logic and set theory. Thus, it seemed possible to reintegrate frames into the logic-oriented approach to human cognition and artificial intelligence. Hence, I tried to carry on Minsky's research programme while being unconvinced by his arguments against the logic-oriented approach. This endeavour resulted in a novel logic, called *frame logic*.

For Minsky, a major motivation to explore frames was to explain the effectiveness of common-sense thought. How on earth is it possible that human minds process information fast and reliably? According to the present state of the art in computational complexity theory and belief revision, this is not possible at all. Rational belief revision is not tractable. That is, it is not computationally feasible to change one's beliefs – when receiving new epistemic input – in such a manner that certain rationality postulates and logical constraints are respected. So, Minsky seemed to be right, after all, in claiming that logic-oriented approaches to human cognition are on the wrong track. I found this hard to accept, despite all theoretical results pointing towards this conclusion. I therefore explored means to achieve tractability using frame concepts and ideas about truth maintenance. This endeavour resulted in a truth maintenance system for a structuralist theory of belief changes, which in turn leads to a novel belief revision scheme. The belief revision scheme is, in fact, tractable for frame logic.

The present book, divided in two parts, develops frame logic and the truth maintenance-inspired belief revision scheme in a stepwise fashion. Part I starts with an exposition of the general problem of dynamic tractable reasoning, and a chapter on frames follows. Then, the reader is introduced to belief revision theory and defeasible reasoning. Part II starts with an axiomatic, Carnapian account of structuralist theory representation. This account is merged with a specific belief revision scheme, resulting in a structuralist theory of belief changes. Thereupon, a truth maintenance system is devised. Finally, frame logic and a novel belief revision scheme will be developed on the basis of the truth maintenance system.

This book brings together different frameworks from different scientific disciplines, so there are a number of people I would like to acknowledge. First of all, I am very grateful to C. Ulises Moulines for his open-mindedness towards my ideas about a syntactic approach to structuralist theory representation when he co-supervised my PhD thesis. My first encounter with belief revision and frames dates back to a postdoctoral stay at Stanford. So, I would like to thank Johan van Benthem and Yoav Shoham for an introduction to dynamic epistemic logic and nonmonotonic reasoning. Mark Musen introduced me to frames. Heinrich Herre from the University of Leipzig kindly supported my research immediately after completion of the PhD, and encouraged me to explore connections between structuralism and knowledge representation.

After my research stay at Stanford, I enjoyed working as an assistant professor, first at the Chair of Andreas Bartels in Bonn and second at the Chair of C. Ulises Moulines at LMU Munich. Moreover, I would like to thank Stephan Hartmann and Hannes Leitgeb for integrating me into Munich Center for Mathematical Philosophy; special thanks are due to Stephan Hartmann for supporting my research after completion of the habilitation. The earlier versions of Chaps. 1 to 7 of this book formed the core part of my habilitation at LMU Munich in 2012. I would like to thank all the committee members for their very valuable comments that helped me improve these chapters: Gerhard Brewka, Hannes Leitgeb, C. Ulises Moulines and Hans Rott.

Not too long after the habilitation, I accepted an offer from the University of British Columbia for a tenure track position. Moving to Canada with the family was quite a challenge. Also, it seemed time to take a break from computational complexity theory and interdisciplinary work. I resumed my work on structuralist belief changes after this break and wrote Chap. 8 of the present book, which expounds frame logic and the truth maintenance-inspired belief revision scheme. I would like to express my sincere gratitude to the department heads, Helen Yanacopulos and Andrew Irvine, for their kind support of my research and to my colleagues, Manuela Ungureanu, Giovanni Grandi, Jim Robinson and Dan Ryder, for a very supportive and cordial work environment. Andrew Irvine and Dan Ryder helped with very valuable comments on the book proposal and the introduction of the book. Moreover, special thanks are due to Josef Zagrodney and Mario Günther for proofreading. Of course, I remain responsible for all the mistakes.

The research stay at Stanford was supported by the DAAD (the German Academic Exchange Service). The Centre for Advanced Studies at LMU Munich, which awarded me a fellowship in the fall term of 2011, provided an excellent research environment for completing the habilitation. Before I came to Canada, I was a Heisenberg fellow of the DFG, the German Research Council.

Finally, I want to thank my wife, Sabine, and my children, Leonard, Laurenz and Mathilde, for their invaluable emotional support and understanding. This book is dedicated to Leonard, the oldest one, which implies a commitment to write two more books to be dedicated to the younger ones.

Kelowna, BC, Canada Holger Andreas

Contents

Part I Foundations

Part II Belief Revision with Frames

List of Figures

Chapter 1
Introduction

In this book, we aim to lay bare the logical foundations of tractable reasoning. We draw on Marvin Minsky's [87] seminal work on frames, which has been highly influential in computer science and, to a lesser extent, in cognitive science. However, only very few attempts have explored ideas about frames in logic. Such is the intended innovation of the present investigation.

In cognitive science, the idea that the human mind works in a modular fashion continues to be very popular. This idea is also referred to as the *massive modularity hypothesis* [100]. The present logical investigation of modularity is directed toward two unresolved problems that have arisen in cognitive science and have yet to receive a proper solution. First, the problem of tractable reasoning, and second the problem of transmodular reasoning with a modular architecture. The first problem concerns the cognitive feasibility of global inferential reasoning: we know very little about the cognitive and logical means that allow us to draw inferences that involve larger amounts of language and memory. Global inferential reasoning appears to be intractable.

Ideas about a modular architecture of human cognition promise to explain how human minds cope with the challenges of computational complexity [41, 100]. However, cognitive scientists have not delivered a more detailed demonstration that modularity actually helps achieve tractability. This gives rise to the second problem: human reasoning is often transmodular and is therefore not confined to a small set of beliefs in a specific domain. A decomposition of reasoning into modular operations thus is needed to maintain the modularity hypothesis for central cognition.

Particularly challenging from a computational perspective is the operation of changing one's beliefs in light of new epistemic input. This is so for two reasons. First, a new piece of information may affect various parts of our overall body of beliefs. Second, new information, if trustworthy, may force us to retract some of our present beliefs. Our reasoning is sometimes dynamic in the sense that we retract certain premises and conclusions that have been accepted so far. This observation gave rise to new areas in formal epistemology and logic: belief revision

H. Andreas, *Dynamic Tractable Reasoning*, Synthese Library 420,
https://doi.org/10.1007/978-3-030-36233-1_1

and nonmonotonic reasoning. Research in these areas is dedicated to the problem of dynamic reasoning, and we now have a variety of approaches to this type of reasoning. But the problem of dynamic tractable reasoning has remained largely unsolved.

The apparent computational infeasibility of inferential reasoning – be it dynamic or static – is a major objection to the *computational theory of mind* and the *language-of-thought* hypothesis propounded by Fodor [55, 59]. But the problem is not tied to these two theoretical hypotheses: as soon as we acknowledge that human minds engage in drawing inferences, we face the problem of computational and cognitive feasibility. This problem does not result from a particular cognitive paradigm.

Why are belief changes computationally challenging, if not even computationally infeasible? Why do these computational issues matter for our understanding of human cognition? Why should a modular account of human reasoning – in terms of frames and frame concepts – help achieve tractability? In what follows, we shall outline an answer to these questions.

1.1 Apparent Intractability of Belief Changes

In science and everyday life, our beliefs are changing continuously. Belief changes are often initiated by a new piece of information. More specifically, we can distinguish between two types of impact that a new piece of information may have on our present beliefs: first, the new epistemic input may allow us to infer further consequences from our present beliefs. Second, we may be forced to give up some of our present beliefs because the new epistemic input is inconsistent with what we presently believe.

This simplified variant of a story by Gärdenfors [61, p. 1] exemplifies how new epistemic input may force us to give up some of our present beliefs. Oscar used to believe that he had given his wife a ring made of gold at their wedding. Later, he realises that his wedding ring has been stained by sulfuric acid. However, he remembers from high school chemistry that sulfuric acid does not stain gold. As he could not deny that the ring is stained and as he also believed that their wedding rings are made of the same material, his beliefs implied a contradiction. Hence, the new epistemic input – i.e., the observation that a certain liquid (believed to be sulfuric acid) stains his wedding ring – forces Oscar to give up some of his present beliefs. From a logical point of view, there are at least three options to regain consistency: (i) he could retract the belief that sulfuric acid stained the ring, (ii) he could call into question that sulfuric acid does not stain gold, and finally, (iii) he could give up the belief that their wedding rings are made of gold. As he paid a somewhat lower price for their wedding rings than might normally be expected, he found himself forced to accept the third option.

The logical study of belief changes has given rise to a new discipline in philosophical logic called *belief revision theory*. For the most part, this discipline has been founded by Alchourrón, Gärdenfors and Makinson [1]. It is for this reason also referred to as *AGM*, or *AGM-style*, *belief revision theory*. The original AGM theory consists in an account of the laws and the semantics of belief changes. A large number of further belief revision schemes have been devised in the wake of the AGM account. Each of these comes with a specific analysis of belief changes.

Why should it be computationally and cognitively impossible to perform belief changes in a rational way? In brief, the answer is that the amount of calculation demanded for a single belief change is likely to exceed the cognitive power of our mind. Theoretical arguments showing this come from the theory of computational complexity and related complexity results for standard, logical approaches to belief revision [89]. These results strongly suggest that belief revision is an *intractable* problem. In more technical terms: if $\mathbf{P} \neq \mathbf{NP}$ (which is quite firmly believed by most computer scientists), then the problem of rationally revising one's beliefs is intractable. We shall explain the computational complexity of belief changes in greater detail in the course of this investigation.

Intractability of a problem means, in less technical terms, that our best computers take too much time to solve non-trivial instances of the problem, despite all the progress in the development of hardware that we have witnessed in past decades. To give an example, no efficient method has been found for determining the satisfiability of a propositional formula. For, the number of relevant valuations grows exponentially with the size of such a formula so that, in the worst case, there is an exponential growth of the number of computation steps needed to test satisfiability.

Cherniak [36, p. 93] has given a telling illustration of how devastating this type of exponential growth is for the feasibility of determining whether or not a propositional formula (or a set of such formulas) is satisfiable: suppose our super computer is able to check whether a given propositional valuation verifies the members of a set of propositions in the time that light takes to traverse the diameter of a proton. Assume, furthermore, that our propositional belief system contains just 138 logically independent propositions. If so, the estimated time from the Big Bang to the present would not suffice to go through all valuations of the system of propositions. Hence, even this super computer would not provide us with the means to test consistency of a moderately complex propositional belief system in such a manner that the test is conclusive for any possible input. We might of course be lucky and find a propositional valuation that verifies all items of the belief system after going through only a few valuations, but it is much more probable that we will be unlucky.

Exponential growth of the computation steps needed to solve a problem, in the worst case, is considered a mark of intractability. Tractability of a problem, by contrast, means that it can be solved with a reasonable number of computation steps. To be more precise, a problem is tractable if and only if the number of computation steps needed to solve an arbitrary instance of it is bounded by a polynomial function whose argument is a parameter that characterises the syntactic size of the instance. The computation of arithmetic functions, such as addition, multiplication, etc., for natural and rational numbers is tractable because corresponding algorithms exhibit no exponential growth of the number of computation steps.

The problem of revising a belief system is computationally even harder than the problem of testing satisfiability of a propositional formula. Belief revision must therefore be considered an intractable problem, at the present state of the art. As

Nebel [89, p. 121] puts it, "The general revision problem for propositional logic appears to be hopelessly infeasible from a computational point of view because they are located on the second level of the polynomial hierarchy." It is fair to say that there is no reasonably expressive and tractable belief revision scheme to be found in the computer science literature to date.[1]

Why is the problem of rational belief revision even harder than testing satisfiability of a propositional formula? The runtime of algorithms that determine belief changes exactly is not only exponential, but super-exponential. That is, these algorithms require exponentially many calls of a subroutine that consists of exponentially many computation steps. Few attempts have been made toward approximate solutions to the general problem of belief revision.[2] And the potentialities of a modular approach to tractable reasoning and belief revision have not yet been fully exploited.

1.2 Computational Theory of Mind

The distinction between tractable and intractable problems has a certain bearing on the understanding of human cognition, as observed by Cherniak [36], Fodor [58], Johnson-Laird [74], Stenning and van Lambalgen [107], Woods [120] and others. Human minds are far more creative than computers and for this reason are certainly distinct from the latter. But they are not capable of executing algorithms and computations faster and more reliably than computers. As for belief revision, there is a large class of problems that appear to require formal reasoning and calculation more than creativity.

In cognitive science, the *computational theory of mind* (CTM) is a research programme that aims to exploit presumed commonalities between computers and the biological machinery underlying our cognition. It follows quite directly from CTM that any problem that is intractable for computers is intractable for human minds as well. As we shall see later on, rather weak formulations of CTM suffice to show this implication. In particular, CTM implies that we are not capable of obeying the norms of standard approaches to rational belief revision. We do not even have a clear idea of how we could approximately conform to the norms of rational belief change.

Modularity, on the other hand, appears to be a means to escape the combinatorial explosion of belief formation. This is so, however, only if we can decompose global reasoning into modular operations. For, belief formation is often a global affair in the sense that new pieces of information potentially impact a large variety of beliefs across a number of different domains. In molecular biology, for example, a new finding about the homology of two protein sequences may imply hypotheses

---

[1] See again Nebel [89], who has given the most comprehensive survey of the complexity results of belief revision schemes. Those belief revision schemes that yield, together with a tractable logic, a tractable determination of belief changes are highly counter-intuitive. We shall be more explicit about this in Sect. 1.8 and Chap. 7.

[2] See [115] and [2]. We shall discuss these proposals in Chap. 7.

about their functional similarity, which in turn may be relevant to our theories about specific biochemical pathways. Even in daily life it happens that new information has an impact across domains. Last winter, I met a passionate skier who sold his property at the local ski hill because of a few mediocre winters in a row and because of scientific evidence for global warming.

We are unable, in general, to tell in advance for which beliefs a new epistemic input is relevant. The modularity hypothesis, *by contrast*, asserts that cognitive modules work in a domain-specific way and are encapsulated. In order to maintain the modularity of central cognition, one therefore must resort to an account of interacting modular units of reasoning [35]. This idea, however, has not yet been formulated in a precise manner. We lack, in particular, something like a logical analysis of modular reasoning. Cognitive scientists have not yet been able to prove that a modular account of human cognition explains how global inferential reasoning is cognitively feasible.[3] At the same time, considerations of tractability have been used to advance the massive modularity hypothesis in cognitive science. The presumed intractability of an a-modular account of human cognition serves as one of three core arguments in favour of this hypothesis [41].

In a similar vein, Fodor [58] has observed an impasse in cognitive science: the undeniable existence of global cognition, i.e., cognition that involves larger amounts of language and knowledge, is fundamentally at odds with the computational theory of mind, given the limited success of logic-oriented artificial intelligence and the difficulties of computing global inferences reliably and fast. Fodor observes that we simply have no idea how global cognition is feasible for minds with finitely bounded resources of computation and memory (see [58, Chaps. 2 and 3] and [59, Chap. 4.4]).

Who is the culprit for this impasse? On the face of it, there appear to be only two candidates: CTM and the view that formal logic has some role to play in an account of human cognition. So, shall we discard CTM or some variant of it? Fodor [58] introduces CTM in the more specific sense that higher cognitive processes are classical computations, i.e., computations that consist of operations upon syntactic items. This understanding of CTM forms the core of the classical computationalist paradigm in cognitive science [20]. Broadly logic-oriented research on human cognition has been driven by this paradigm. Assuming a syntactic nature of computation appears to justify applying the computational complexity theory to human cognition in the first place. So it seems natural to discard CTM – more precisely, the syntactic formulation of CTM – in order to solve our problem. Computational complexity theory, however, pertains to human cognition quite independently of this assumption since the scope of complexity theory is governed by the *Church-Turing thesis*. This thesis asserts that any physically realisable computation device – whether it is based on silicon, neurones, DNA or some other technology – can be simulated by a Turing machine [14, p. 26]. Issues of computational complexity, therefore, pertain to human cognition quite independently of a commitment to a syntactic variant of CTM.

---

[3] The famous work on bounded rationality by Gigerenzer [64, 65] is not concerned with modularity.

Shall we discard, then, logical approaches to human reasoning and cognition? The *connectionist* paradigm in cognitive science is motivated by the significant success of artificial neural networks in pattern recognition and the relatively minor success of *good old-fashioned artificial intelligence*, which is logic-oriented. Proponents of the connectionist paradigm have suggested that formal logic has hardly any explanatory value for human intelligence [39]. It is undeniable, however, that scientific and quotidian reasoning have a genuine propositional structure since reasoning and argumentation essentially consists in making inferential transitions from antecedently acquired or accepted propositions. It has been moreover shown that propositional reasoning can be implemented by means of neural networks [79]. Stenning and van Lambalgen [107], Johnson-Laird [74] and others have successfully supplemented logical approaches to human cognition with empirical research in psychology.

One might also try to dissolve issues of tractability by emphasising the normative role of logic. Logical systems, one could argue, explicate norms of reasoning without actually describing human reasoning. But even a purely normative view of logic would not solve the problem under consideration. In order for a standard to have a normative role, it must be possible to, at least approximately, meet it. A norm that we cannot obey – neither exactly nor approximately – cannot be considered a norm in the first place. *Ought* implies *Can*.[4] Following van Benthem [112], we view the role of logical systems in an analysis of human reasoning as partly normative and partly descriptive.

A closer look at our problem reveals more culprits to consider. There might be something wrong with the distinction between tractable and intractable problems in the theory of computational complexity. In particular, the decision to focus on worst-case scenarios in the original distinction is open to question.[5] This line, however, is not pursued here.

Yet another option is to work on the cognitive adequacy of logical systems themselves. This is the strategy pursued here. We tackle the computational issues of belief revision using frames and frame concepts, and thus resume a research programme originated by Minsky [87]. Even though Minsky himself was rather hostile toward logic-oriented approaches to human cognition, there is a logical and set-theoretic core recognisable in his account of frames. We explicate and further develop this core using set-theoretic predicates in the tradition of Sneed [104] and Balzer et al. [18].

A note on the infamous frame problem is in order here. In an investigation of frames, one would expect to find a thorough discussion of this problem. Fodor [57, 58] makes much out of the frame problem, but is charged with not knowing 'the frame problem from a bunch of bananas' by Hayes [72]. In fact, the account that

---

[4] Thanks to Hans Rott for this point.

[5] Focusing on worst-case scenarios means that the computational complexity of a problem is determined by the maximal number of computation steps that are needed to solve any possible instance of the problem.

Fodor [57, 58] gives of the frame problem is a reinterpretation of the original frame problem as described by McCarthy and Hayes [82]. While there is some substantial connection between the original frame problem and Fodor's reinterpretation, there is no need to discuss any variant of the frame problem in order to explain the computational challenges of dynamic, inferential reasoning. Likewise, there is no need to discuss abductive reasoning, as Fodor [58] does, for this purpose. Analysis of the computational complexity of belief revision gives us a more concise and less controversial exposition of the problem that dynamic, inferential reasoning is intractable in the setting of classical propositional logic. Minsky himself, in his seminal work on frames [87], makes no explicit reference to the frame problem.

1.3 Frame Logic

Why may frame concepts help reduce the computational complexity of belief changes? Such concepts have a richer structure than ordinary concepts. A telling example used by Minsky [87, p. 47] is that of a child's birthday party. Unlike an ordinary concept, this concept does not seem to apply well to a certain individual or tuples of individuals. We would not say that the concept in question applies to the birthday child, the union of birthday child and guests, or to the place where the party is given. What then are objects to which the concept of a birthday party is applied? Minsky says that it describes a situation that involves a number of different things: guests, games, presents, a birthday cake, a party meal, decor, etc. These things, normally, satisfy certain conditions: the guests are friends of the host, the games must be fun, the gifts must please the birthday child, etc.

From a logical point of view, a frame concept is a concept that applies to sequences of sets of objects as opposed to mere tuples of objects (which are not sets). Furthermore, frame concepts impose semantic constraints upon the (first-order) predicates of a small fragment of our language. For example, the guests of the birthday party are, normally, friends of the host. Frame concepts can therefore be used to interpret a piece of language in a small domain. This amounts to subdividing our global language into small sublanguages. These sublanguages, in turn, have different and yet interrelated interpretations in small subdomains. It is thus the notion of a frame concept by means of which we try to semantically explicate the notion of a cognitive module.

Modularising semantics in this way allows us to distinguish easily between intramodular and transmodular reasoning. The former type of reasoning is confined to a single module – which concerns the interpretation of a piece of language in a small domain – whereas the latter communicates information from one module to another. The distinction between intra- and transmodular reasoning gives rise to a proper logic of frames that emerges from our investigation. This logic is guided by the following two principles:

(1) Classical first-order logic remains valid within the application of a frame concept.
(2) Only atomic sentences and their negation can be inferred from one application of a frame concept to another.

Restricting the scope of applying classical logic is key to reducing the computational complexity of global inferential reasoning. Inspired by the methods of object-oriented programming, we furthermore allow compositions of frame concepts. Compositions, however, must be bounded in size so as to retain the tractability of frame logic. Note that object-oriented programming is a distinctive style of modular programming. It is commonly viewed as the greatest success story of Minsky's methodology of frames in [87].

For frame logic to be devised and investigated, some formal work has to be done. First, we show how the notion of a frame concept can be formalised using set-theoretic predicates. This will allow us to give an axiomatic account of reasoning with frame concepts. Then, we merge the set-theoretic account of frames with some AGM-style belief revision scheme.[6] The result is a belief revision theory with frame concepts. For this theory, we finally devise a truth maintenance system (TMS), i.e., an algorithm that determines how presently accepted truth values change upon new epistemic input. The TMS determines belief changes in a tractable manner.

The TMS is, furthermore, shown to serve as a powerful approximation of first-order reasoning, but it is not sound and complete with respect to first-order logic, even when confined to finite domains. Soundness and completeness, however, can be achieved for frame logic on condition of two further constraints: first, the conclusion is quantifier-free, while premises may well contain quantifiers. Second, the domain of any modular unit of reasoning is finite.

Frame logic is developed on the basis of a natural deduction system of classical first-order logic. We shall also speak of natural deduction frame logic to refer to this logic. A variant of natural deduction frame logic will be devised in the propositional resolution calculus. Resolution frame logic is shown to be sound and complete, without any further qualifications. Of course, we explain the basic concepts of the resolution calculus so as to make this investigation as self-contained as possible.

One word on unit resolution and unit propagation is in order in the context of frame logic. Unit resolution is a specific inference rule in the setting of the resolution calculus. At least one of the two premises of this inference rule must be a literal, i.e., an atom or the negation of an atom. At the semantic level, inferences licensed by unit resolution are described as *unit propagation*. Obviously, transmodular reasoning in frame logic amounts to the propagation of determinate literals to other applications of frame concepts. However, to the best of my knowledge, ideas about unit propagation have not yet been taken to the development of a proper logic of modular reasoning.

---

[6] The belief revision scheme of preferred subtheories by Brewka [24] proved well suited for this purpose.

1.4 TM Belief Revision

Natural deduction and resolution frame logic approximate classical first-order and classical propositional reasoning, respectively. Likewise, the TMS yields only an approximate determination of belief changes, approximate with respect to standard approaches in the AGM tradition. We turn, however, the approximate nature of the TMS into a virtue by devising a belief revision scheme that mirrors the working of the TMS. This belief revision scheme is tractable for natural deduction and resolution frame logic. Since the belief revision scheme is inspired by ideas about truth maintenance, we speak of *TM belief changes* in order to refer it. The TMS, thus, serves as a ladder by means of which we reach the two frame logics and a novel belief revision scheme.

Recall that some modification of an AGM-style approach to belief revision is necessary since a tractable logic alone does not suffice to resolve the computational issues of belief revision [89]. If we understand the notion of rational belief change in terms of AGM-style approaches, then $\mathbf{NP} \neq \mathbf{P}$ implies that there is simply no exact solution to the problem of tractable, rational belief revision. (We shall make this claim more precise in Sect. 1.8 below.) Approximations and computational simplifications of AGM-style approaches to belief revision are therefore of theoretical interest, at least from a cognitive point of view.

The semantics and the inference rules of frame logic, together with the TM belief revision scheme, are aimed at analysing our means of coping with the computational and cognitive challenges of belief revision. While we do not claim that the present account is literally true in all respects, we think that this account makes significant progress toward a logical analysis of quotidian human reasoning that is cognitively plausible. Frame logic and the TM belief revision scheme are intended to take the *cognitive turn in logic* one step further. Johan van Benthem [112, p. 67] came to speak of such a turn when reviewing specific trends in logic, such as belief revision theory and the related dynamic epistemic logics. Issues of computational complexity are explicitly mentioned as well [112, p. 74].

The first TMS was devised by Doyle [47]. For the expert reader it may be instructive to note, at this point, the differences between the present attempt at truth maintenance and Doyle's original TMS. First, the present system is more liberal concerning the logical form of what Doyle calls *justifications*. Any instance of an axiom can be a justification. Second, justifications may well become retracted. Third, there is an epistemic ranking of justifications.

1.5 Modularity in Cognitive Science

Now that we have outlined the key results of our investigation, let us relate these, in somewhat greater detail, to the modularity hypothesis in cognitive science. To a great extent, this hypothesis originated from Fodor's *The Modularity of Mind* [56].

There, Fodor develops a twofold thesis, which has been described as *minimal* or *peripheral modularity*: input and output systems of the mind, such as sensory and motor-control systems, work in a modular fashion. Central cognitive systems, by contrast, are non-modular. It is central cognitive systems that realise our capacities of explicit reasoning, belief formation and decision making. The notion of a module itself is characterised by the following properties and features in Fodor [56]: (i) domain specifity, (ii) information encapsulation, (iii) mandatoriness, (iv) fast output, (v) shallow, i.e., non-conceptual output, (vi) neural localisation and (vii) innateness.

From the perspective of evolutionary psychology, more radically modular proposals have been made, also comprising central cognition. As indicated above, this work aims to show that the human mind works in a massively modular fashion (see in particular Cosmides and Tooby [41]). According to this thesis, both peripheral and central cognition work with a modular architecture. Tractability is one of three arguments advanced in favour of the massive modularity hypothesis in Cosmides and Tooby [41]. It is the only argument we are concerned with here.

Even though this is only a very rough sketch of the modularity map in cognitive science and evolutionary psychology, it is precise enough to locate our results on this map. In devising a proper logic of modular reasoning, we aim to contribute to an understanding of modularity at the level of central cognition. The modular units of reasoning characterised by frame logic share at least two important properties with the Fodorian notion of a module: they are domain-specific and work with information encapsulation. This is good news for proponents of the massive modularity hypothesis since encapsulation and domain specifity are considered most central to this hypothesis [100, p. 63].

Let us further compare our logical explanation of a module with Fodor's notion. Elementary modules (which are not composed of other modules) have fast output insofar as they are associated with very simple patterns of inference. Furthermore, our specific notion of a module is perfectly consistent with having neural localisation. It is more than plausible to assume this property. As we are concerned with central cognition, our modules have non-shallow, conceptual output and input. Their working may or may not be mandatory. While Fodorian modules are not necessarily interactive, our logical modules are.

The present account of frames may well be viewed as a logical variant of the massive modularity hypothesis. This variant comes with a precise hypothesis about the working of information encapsulation:

(1) First-order and propositional reasoning within a module are encapsulated from extra-modular disjunctive information.
(2) Information in the form of disjunctions and implications is encapsulated in the sense that it is located within a module and that it cannot directly be accessed by other modules.
(3) Only literals can be communicated between modular units of reasoning.
(4) The elementary modular units of reasoning are given by applications of generalisations.