Belief Revision Based on Information States
Hyperintensionality, Fragmentation, and Consistency

Dissertation von Ayșe Sena Bozdağ

München 2025

Belief Revision Based on Information States
Hyperintensionality, Fragmentation, and Consistency

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
der Ludwig-Maximilians-Universität München

vorgelegt von
Ayşe Sena Bozdağ

aus
Ankara, Türkei

2025

Erstgutachter: Prof. DDr. Hannes Leitgeb

Zweitgutachter: PD Dr. Martin Fischer

Tag der mündlichen Prüfung: 10.02.2025

Contents

List of Figures

Zusammenfassung

Der Hauptbeitrag dieser Dissertation ist die Entwicklung einer hyperintensionalen Semantik für die Überzeugungsrevision und eines entsprechenden Systems der dynamischen doxastischen Logik. Dies erfolgt in der Absicht verschiedene Aspekte der Überidealisierung und des logischen Allwissenheitsproblems zu behandeln, die in der Literatur zur Überzeugungsdynamik immer wieder auftauchen. Die in Kapitel 2 vorgestellten Modelle bilden den Kern dieser Dissertation. Der Ausgangspunkt für die Entwicklung dieser Modelle ist die Verlagerung des Darstellungsfokus doxastischer Modelle von Mengen von Überzeugungen auf Informationssammlungen und die Definition von Überzeugungsänderungen als Artefakte von Informationsänderungen. Auf diese Weise zielt die aktuelle Arbeit darauf ab, eine flexible und weniger idealisierte Darstellung von Überzeugungsrepräsentation und Überzeugungsänderung zu erreichen. Dies geschieht, indem objektlinguistische Überzeugungsänderungsoperatoren in eine Zustandssemantik für Informationsdynamik eingebracht werden. Das Ergebnis ist eine nicht-monotone, hyperintensionale und inkonsistenztolerante Überzeugungsdynamik.

Aus diesen Modellen ergeben sich eine Reihe wichtiger Themen in Bezug auf die Überzeugungsdynamik, wie Hyperintensionalität, Inkonsistenztoleranz und deduktive Geschlossenheit. Im weiteren Verlauf dieser Dissertation werden diese Themen, mit Ausnahme von Kapitel 6, näher untersucht. Insbesondere Kapitel 3 zielt darauf ab, diese Arbeit in der Hyperintensionalitätsliteratur zu verorten, indem es neben einer kurzen Geschichte der Debatte über hyperintensionale Bedeutung einen vergleichenden Überblick über verschiedene Darstellungen hyperintensionaler Überzeugungsänderungen gibt. Aus der Analyse dieser Darstellungen ergeben sich verschiedene Vergleichspunkte, insbesondere zwischen möglichen Welten und zustandssemantischen Darstellungen der hyperintensionalen Überzeugungsrevision, die in Abschnitt 3.3 ausführlich diskutiert werden. Die letztgenannten Themen, nämlich Inkonsistenztoleranz und deduktiver Abschluss, werden zunächst unter dem Begriff der Fragmentierung von Überzeugungszuständen untersucht. Die Debatte über fragmentierte Mengen von Überzeugungen konzentriert sich auf die Anforderungen von deduktiver Geschlossenheit und Konsistenz, die an Mengen von Überzeugungen gestellt werden, und wie diese zu dem Problem der logischen Allwissenheit führen. Der Literaturüberblick zeigt verschiedene Modellierungslösungen für das Problem, Kapitel 4 zielt dann darauf ab, die aktuelle Arbeit in dieser Literatur zu verorten.

Die Frage der Inkonsistenztoleranz taucht in Kapitel 5 im Zusammenhang mit konsistenzsensitiven epistemischen Modalitäten wieder auf. In diesem Teil der Dissertation

werden zwei neue epistemische Modalitäten untersucht. Eine *konsistenzsensitive Evidenzmodalität* wählt die konsistenten Teile einer ansonsten inkonsistenten Informationssammlung aus, indem sie vertrauenswürdige Informationsquellen unterscheidet. Eine konsistenzsensitive sichere Überzeugungsmodalität bestimmt dann eine Reihe von dauerhaften Überzeugungen auf der Grundlage von Evidenz. Die hier vorgestellten Modelle sind zwar stark von der Darstellung in Kapitel 2 inspiriert, weisen aber aufgrund der Wahl der zugrunde liegenden Semantik verschiedene Unterschiede auf.

In Kapitel 6 schließlich wird die Diskussion in eine andere Richtung gelenkt. Hier werden die in Kapitel 2 entwickelten Modelle neu interpretiert und angewandt, um eine besondere Form des begrifflichen Wandels darzustellen. Das Ziel dieser Anwendung ist es, diese Modelle als Brückenkomponente zwischen Theorien der Überzeugungsrevision und des wissenschaftlichen Wandels vorzuschlagen. Insbesondere wird vorgeschlagen, dass ein überzeugungsrevisionsähnliches System verwendet werden kann, um radikale Arten von begrifflichem Wandel abzubilden, die im Rahmen wissenschaftlicher Revolutionen auftreten können. Es ist daher keine Überraschung, dass die Revision von Überzeugungen und die Revision wissenschaftlicher Theorien auffallend ähnlich sind, da beide als Korpus von Informationen und Schlussfolgerungen (oder Hypothesen) verstanden werden können. Die Parallele zwischen den beiden wird in diesem Kapitel auch durch die Übersetzung der Überzeugungsrevisionspostulate in Rationalitätspostulate für begrifflichen Wandel in Abschnitt 6.4 deutlich.

Summary

The main contribution of this dissertation is the development of a hyperintensional semantics for belief revision and a corresponding system of dynamic doxastic logic, with the intention of addressing various aspects of over-idealisation and the logical omniscience problem which repeatedly occurs in the literature on belief dynamics. The models introduced in Chapter 2 make up the core of this dissertation. The starting point in the development of these models is shifting the representational focus of doxastic models from belief sets to collections of information, and defining changes of beliefs as artifacts of changes of information. In this way, the current work is aimed at achieving a flexible and less idealised account of belief representation and belief change. This is done by accommodating object-linguistic belief change operators within a state-semantics for information dynamics, which results in suggesting a non-monotonic, hyperintensional and inconsistency tolerant belief dynamics.

A number of important themes concerning belief dynamics result from these models, such as hyperintensionality, inconsistency-tolerance and deductive closure. In the rest of this dissertation, these themes are investigated in more detail, with the exception of Chapter 6. In particular, Chapter 3 aims for locating this work in the hyperintensionality literature, by presenting a comparative survey of various accounts of hyperintensional belief change, besides a brief history of the debate about hyperintensional meaning. Various comparison points arise from the analysis of these accounts, in particular between possible worlds and state-semantics representations of hyperintensional belief revision, which are discussed in detail in Section 3.3. The latter themes mentioned above, namely inconsistency-tolerance and deductive closure are first investigated under the umbrella of fragmentation of belief states. The debate on fragmented beliefs is focused on the requirements of deductive closure and consistency imposed on belief sets, and how these lead to the problem of logical omniscience. The review of the literature shows various modeling solutions to the problem, Chapter 4 is then aimed at locating the current work in this literature.

The issue of inconsistency-tolerance reoccurs in Chapter 5, in relation to consistency-sensitive epistemic modalities. In this part of the dissertation, two new epistemic modalities are investigated. A *consistency-sensitive evidence modality* picks out the consistent parts of an otherwise inconsistent collection of information, by distinguishing trusted sources of information. A consistency-sensitive safe belief modality then determines a set of persistent beliefs based on evidence. While the models introduced here are heavily inspired by

the presentation in Chapter 2, they exhibit various differences due to the choice of the underlying semantics.

Finally, Chapter 6 steers the discussion towards another direction. Here, the models developed in Chapter 2 are reinterpreted and applied to represent a special form of conceptual change. The aim of this application is to suggest these models as a bridging component between theories of belief revision and of scientific change. In particular, it is proposed that a belief-revision-like system can be used to mirror radical types of conceptual change that might occur as part of scientific revolutions. It is not surprising that belief revision and scientific theory revision are strikingly similar, given that both can be understood as corpus of information and inferences (or hypothesis). The parallel between the two is manifested in this chapter, also by the translation of belief revision postulates into rationality postulates for conceptual change in Section 6.4.

Co-authored and single-authored publications

Chapter 2 is a reprint with minor revisions of the manuscript "A Semantics For Hyperintensional Belief Revision Based On Information Bases", published in Studia Logica, December 2021, [29][1].

Chapter 5 is a reprint in part of the manuscript "Consistency-Sensitive Epistemic Modalities in Information-Based Semantics", published in Studia Logica, May 2024 [126]. The content of this manuscript is joint work with Vít Punčochář, Marta Bílková, and Thomas M. Ferguson.[2]

Chapter 6 is a reprint of the manuscript "Taking Up Thagard's Challenge: A Formal Model of Conceptual Revision", published in Journal of Philosophical Logic, January 2022 [30]. The content of this manuscript is joint work with Matteo De Benedetto.[3]

[1]Reproduced according to http://creativecommons.org/licenses/by/4.0/.

[2]I have contributed to the material included in this dissertation at 30%. Reproduced with permission from Springer Nature.

[3]This paper is product of equal contribution between the authors. Reproduced according to http://creativecommons.org/licenses/by/4.0/.

Acknowledgements

I consider myself one of the few incredibly lucky people to have crossed paths with Hannes Leitgeb and have him as my supervisor. As he is a great philosopher and logician, he is also amazing in supporting a PhD student who often struggles with work itself but also with imposter syndrome. I have learned through the stories of my friends, who were not as lucky as me with their supervisors, that this level of support and encouragement is not a given. Doing a PhD is a mentally and emotionally draining task. Often it feels like a lifetime, even more so in my case. I cannot imagine going on this road without Hannes' support and guidance. I will always be so proud to have worked with him and learned from him.

I am thankful for the amazing research community that is the MCMP. I would like to thank personally to Lavinia Picollo, for being a great example for a woman in academia, Alessandra Marra for being my unofficial mentor since she arrived in Munich, Dominik Klein for actually being my mentor, and Norbert Gratzl for showing up in the most important times. I am afraid I cannot list all the beautiful people I have crossed paths with at the MCMP and beyond, that became parts of this journey either professionally or personally. I will try my best. My office-mates through the years, Adrian Ommundsen, Ana Cholodovskis, Bruno Jacinto, Elio La Rosa, Conrad Friedrich, Matías Osta Vélez, Diego Tajer, Marta Sznajder, Marianna Antonutti Marfori, Martin Fischer, Damian Szmuc, and of course, Ursula Danninger, my co-authors Vít Punčochář, Marta Bílková, and Thomas M. Ferguson. I am grateful for knowing them all.

My luck with supervisors continues as a postdoc, I would like to thank Olivier Roy, for his support, encouragement, and patience during the last phase of my writing. I did not know that one would get this much support and mentoring once not a student anymore. I hope to keep learning from Olivier and work with him in the years to come.

In the wild wild world of being a PhD student, just as important to stay sane are the friends and community. It is (fortunately and unfortunately) a very international community, people move on, you lose some friends, you gain new ones, and if you are lucky enough, the old ones come back too. It would be impossible to survive in this world without my friends whose experiences I share and who share mine. From the first day I arrived in Munich, I shared all with Matteo De Benedetto. I cannot thank him enough for his presence in our office, and in my life outside of it. He transformed Munich into a home for me, something I feel ever stronger next to him.

I am so lucky to share this phase of life with so many beautiful people and grateful for all the conversations that kept us strong and that kept the fun alive. My friendship with Allie Richards has become one of the most valuable things in my life. Allie and I spent so much time together, walking, talking, working, and partying, I wish many more years together as members of the stuck-in-Munich crew. I would like to thank my friend Michal Hladky for all the uplifting talks and all the goofy ones. I could not go through especially the last months of writing without the presence of Jingyi Wu, one of my favourite people on earth. I saw the finish line at our retreat together. I shared my office at the MCMP with many amazing people, some expanded through all aspects of my life, I am grateful that Gianluca Grilletti is one of them.

I am so lucky to have the support of very special people outside of my academic world. My oldest friend Dila Demir has been the person I turn to, to share and to seek comfort with everything going on in my life. I would like to thank Christian Schnurr, although he joined this journey later than the others, he has been one of my strongest sources of support and happiness.

I could not have done all this if I had not been born into the family I have. I would like to thank my parents and my sister for all the love, opportunities, and support they have given me. Sizi seviyorum!

Chapter 1

Introduction

Belief representation and dynamics has a long tradition in the philosophical logic literature. It is possible to put a pin on the development of the AGM belief change theory by Alchourrón, Gärdenfors and Makinson [1] as the starting point of the current debates in this literature. While the AGM-style belief revision has earned the title of *standard theory* of belief change, numerous classical and non-classical ways of changing beliefs have been introduced, referring to the laws of AGM in one way or another, reproducing or challenging them. The challenges directed to the AGM usually stem from concerns of *over-idealisation* imposed by the AGM belief contraction and belief revision postulates. This dissertation is not an exception to this tradition, as the main goal is to put forth a belief revision system which addresses various concerns of over-idealisation.

The debate concerning over-idealisation in modeling belief change, however, has other sources besides the AGM. By far the most common way of representing beliefs and belief dynamics has been using possible worlds semantics. This tradition goes back to Hintikka's conception of knowledge and belief modalities [73]. In this tradition, possible worlds are interpreted as alternative epistemic possibilities for agents, and knowledge and belief are defined as truth in the non-excluded alternatives. It is argued that imposing the structure of possible worlds onto belief sets again leads to over-idealisation in a theory of rational belief change, and can be understood as one of the sources of *the logical omniscience problem*.

The main contribution of this dissertation is the development of a new semantics for hyperintensional belief revision and a corresponding system of dynamic doxastic logic in Chapter 2, with the intention of addressing various aspects of over-idealisation and the logical omniscience problem. The models introduced in this chapter make up the core of this dissertation. The starting point in the development of these models is shifting the representational focus of doxastic models from belief sets to collections of information, and defining changes of beliefs as artifacts of changes of information. In this way, the current work is aimed at achieving a flexible and less idealised account of belief representation and belief change. This is done by accommodating object-linguistic belief change operators within a state-semantics for information dynamics, which results in suggesting a non-monotonic, hyperintensional and inconsistency-tolerant belief dynamics.

A number of important themes concerning belief dynamics result from these models, such as hyperintensionality, inconsistency-tolerance and deductive closure. In the rest of this dissertation, these themes are investigated in more detail, with the exception of Chapter 6. In particular, Chapter 3 aims to locate this work in the hyperintensionality literature, by presenting a comparative survey of various accounts of hyperintensional belief change, besides a brief history of the debate about hyperintensional meaning. Various comparison points arise from the analysis of these accounts, in particular between possible worlds and state-semantics representations of hyperintensional belief revision, which are discussed in detail in Section 3.3. The latter themes mentioned above, namely inconsistency-tolerance and deductive closure are first investigated under the umbrella of *fragmentation of belief states*. The debate on fragmented beliefs is focused on the requirements of deductive closure and consistency imposed on belief sets, and how these lead to the problem of logical omniscience. The review of the literature shows various modeling solutions to the problem, Chapter 4 is then aimed at locating the current work in this literature.

The issue of inconsistency-tolerance reoccurs in Chapter 5, in relation to consistency-sensitive epistemic modalities. In this part of the dissertation, two new epistemic modalities are investigated. A *consistency-sensitive evidence modality* picks out the consistent parts of an otherwise inconsistent collection of information, by distinguishing *trusted sources* of information. A *consistency-sensitive safe belief modality* then determines a set of persistent beliefs based on evidence. While the models introduced here are heavily inspired by the presentation in Chapter 2, they exhibit various differences due to the choice of the underlying semantics.

Finally, Chapter 6 steers the discussion towards another direction. Here, the models developed in Chapter 2 are reinterpreted and applied to represent a special form of *conceptual change*. The aim of this application is to suggest these models as a bridging component between theories of belief revision and of scientific change. In particular, it is proposed that a belief-revision-like system can be used to mirror radical types of conceptual change that might occur as part of scientific revolutions. It is not surprising that belief revision and scientific theory revision are strikingly similar, given that both can be understood as corpuses of information and inferences (or hypothesis). The parallel between the two is manifested in this chapter, also by the translation of belief revision postulates into rationality postulates for conceptual change in Section 6.4.

In the rest of this chapter, the main themes taken up in this dissertation are introduced, and an outline is stated.

1.1 Belief revision

A major milestone in the literature on belief change is the development of the AGM belief change theory by Alchourrón, Gärdenfors and Makinson [1]. The AGM combines belief change with new information and non-monotonic reasoning, and provides a representation theorem for various postulates that guide rational belief revision and belief contraction. These postulates advanced by the AGM as well as other related works of its authors [63, 64, 65] are commonly taken as the reference point for further discussions and developments in the broader literature on belief change. In terms of the formalisation style, the AGM is a set-theoretic representation of belief change, based on meta-linguistic belief contraction and revision functions. Specifically, the *partial meet contraction* and *partial meet revision* functions of the AGM map a set of formulae (i.e., a belief set) and a formula (i.e., the new information) to a new set of formulae, which can be characterised as a *maximal non-implying set*. The AGM requires that the belief sets of agents are logically closed theories both before and after belief change occurs, and the changes are piece-meal with the new information formulated as a singleton formula.[1]

The AGM can also be represented using possible worlds. This is done famously by Grove [62], to achieve a more *plausible* and *natural* representation theorem than the AGM, for the so-called Gärdenfors postulates. Grove's system models doxastic states of single agents by using a system of spheres based on Lewis's sphere semantics for counterfactuals [101], that orders a set of possible worlds based on a fixed plausibility or preference ordering. In a given model, the set of possible worlds is totally ordered, such that, the worlds contained in the innermost sphere are the ones that the agent considers doxastically most plausible or preferable and the set of sentences determined by the intersection of these worlds constitutes the belief set of the agent. Accordingly, with K fixed as the innermost sphere of a system, contracting the belief set K with respect to a piece of information α can be defined as moving to the closest sphere K' (from K) where $K' \cap \neg\alpha$ is non-empty. Revising the belief set K with the information α can then be defined as removing all $\neg\alpha$-worlds from the system and moving to the closest sphere where α is true.

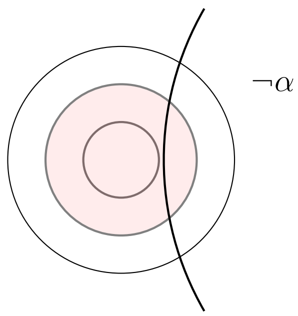

Figure 1.1: The colored area represents the contracted belief set $K \dot{-} \alpha$.

[1]In particular, the AGM partial meet contraction function, denoted by $\dot{-}$ takes as its arguments a belief set K and a formula α, and yields a new belief set K'. K' is the result of intersecting the selected maximal parts of the original belief set K that does not entail α, based on a fixed selection function γ. Hence, $K \dot{-} \alpha$ determined by a selection function γ is defined as $\cap\gamma(K \perp \alpha)$, where $(K \perp \alpha)$ is the set of non-α-implying maximal subsets of K. The AGM partial meet revision (denoted by $*$) on the other hand, is defined based on the contraction function, such that, $K * \alpha$ determined by γ is defined as $Cn((K \dot{-} \neg\alpha) \cup \{\alpha\})$, where Cn is the operation denoting logical closure under classical logic.

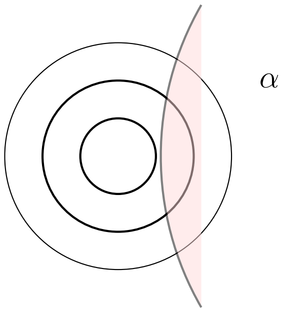

Figure 1.2: The intersection of the α-area with the innermost sphere in the colored area represents the revised belief set $K * \alpha$.

The models offered both in [1] and in [62] then represent a way of changing beliefs which requires, among others, closure and consistency (otherwise trivialisation) of revised belief sets, and equal treatment of logically equivalent sentences in the process of belief revision. Postulates of logical closure, consistency and extensionality mentioned here are respectively:[2]

- (K*1) $K * \alpha = Cn(K * \alpha)$,
- (K*5) if α is a contradiction then $K * \alpha = K_{\perp}$, and
- (K*6) if $\alpha \equiv \beta$ then $K * \alpha \equiv K * \beta$.

There are various attempts in the literature that directly aim to capture the AGM-style belief revision with *object-linguistic* belief change operators[3]. These works uphold some or all AGM revision postulates which are debated in the next sessions, most importantly closure of belief sets, trivialisation of inconsistent belief sets and equivalent treatment of materially equivalent sentences. For instance, the AGM postulate K*6 above is commonly formulated as an axiom of conditional belief or belief change in the following ways [20, p.136]:

- from $\phi \equiv \psi$ infer $B^{\phi}\chi \equiv B^{\psi}\chi$
- from $\phi \equiv \psi$ infer $[*\phi]\chi \equiv [*\psi]\chi$

The above statements say, if ϕ and ψ are classically equivalent sentences, then "if χ is believed conditional on (or upon believing) ϕ, it is also believed conditional on ψ", and "if after revision with ϕ it is believed that χ, then after revision with ψ it is believed that χ", where ϕ, ψ, χ are sentences in a language L. The belief revision models developed in this dissertation challenge these assumptions for belief revision, among others. First and foremost, the models introduced in Chapter 2 are models of *hyperintensional belief change*,

[2] For the full list of the postulates, see [1].

[3] [139, 138, 103, 161, 93, 162, 155, 156, 58] are listed among the works which include belief dynamics upholding the AGM postulates by Berto [20].

hence challenging postulate K*6. That is, they present more fine-grained identification conditions for sentences in the belief context, than what is allowed in the AGM and the Grove models, and in general, any *intensional semantics* including Kripke semantics for modal logic. The models allow, furthermore, belief sets that are not deductively closed, while tolerating inconsistency on different levels of reasoning, hence challenging both K*1 and K*5 above.

An important alternative to the AGM, using the same linguistic level of representation, is the theory of base-generated belief revisions [68, 132]. These theories model belief change using again meta-linguistic contraction and revision operators, however, the objects of belief change are now *structured belief bases* rather than closed belief sets. A belief base can be understood as a possibly incomplete and inconsistent collection of information, consisting of various smaller (possibly ordered) clusters of information. Moreover, they allow belief change with respect to collections of information, unlike the singleton formulation of new information of the AGM. The base-generated revision theories differentiate the information level (belief bases) from the level of inferences (belief sets). Consequently, it is allowed that not all sentences in a belief base are accepted as beliefs by the agents. Evaluated from the point of the AGM postulates, this distinction has important consequences. For instance, the belief sets are not necessarily logically closed. Moreover, although belief revision and contraction are consistency-preserving operations on belief bases, base-generated revision models are inconsistency tolerant, as inconsistencies can be contained in the belief bases. Lastly, beliefs that are merely inferred from other (support) beliefs are not retained after the support beliefs are contracted. Allowing pieces of information to stand and fall together in this way has important consequences for belief dynamics, such as the failure of the AGM recovery postulate.[4] The base-generated revision models relate closer to the frameworks developed here, in virtue of the distinction they suggest between the information level and the belief level, containment of inconsistencies in the information level, and taking structured collections of information as the objects of belief revision.

The AGM-style belief revision is sometimes called *the standard theory* of belief revision. The recent belief revision literature, however, is dominated by object-linguistic formalisations of belief dynamics. Development of Dynamic Epistemic Logic (DEL) and Dynamic Doxastic Logic (DDL) based on Kripke models for modal logics ensured the domination of this tradition.[5] Both DEL and DDL are modal logic approaches to belief and belief

[4]The recovery postulate is a principle of minimal change and it has been the most debated AGM postulate (for more on the debate, see [48]). It results partly from the fact that each piece of information is considered separately as singletons, and they do not stand and fall together unless they logically entail one another. The AGM postulate can be formulated as $K \subseteq Cn((K \dot{-} \alpha) \cup \{\alpha\})$.

[5]Development of Dynamic Epistemic Logic has been a long tradition, inspired by developments in epistemic logic, dynamic modal logic, belief revision, etc. For a historical overview see [162]. Important works in the early literature include [154, 122]. Developments in DEL has inspired the tradition of Dynamic Doxastic Logic with works such as [103, 137] leading the path. In [11, p.2] the authors compare DDL as introduced in [138] and developed further by Segerberg, with the DEL paradigm: “DEL treats dynamic revision as an ‘external’ operation (representing actions as changes of the current model), while in DDL the dynamics is ‘internal’ to the model (i.e., actions are represented as changes of doxastic structure within the same model)”.

revision, built on static sentential knowledge and belief modalities as introduced by Hintikka [73], and they represent changes of belief via dynamic, model-changing operators on the object-language level. They are strictly more expressive than the AGM-style representations, for instance, it is possible to represent multi-agent doxastic models and nested formulas of belief change.[6] Following Hintikka, the static knowledge and belief modalities are formulated as necessity-like modalities. Based on the idea that possible worlds represent epistemic possibilities for agents, knowledge and belief are defined as *truth in all non-excluded possibilities*. Specifically, the belief state of an agent is determined by the doxastic possibilities which the agent considers as candidates for the actual world, i.e., a set of worlds that the agent cannot distinguish from each other and from the actual world.

This well-studied and robust formalisation of belief, however, comes at the cost of extreme idealisation. The primary components of a doxastic model based on Kripke models are possible worlds and binary accessibility relations between worlds, describing indistinguishability from the agents' point of view. While the frame conditions (on the accessibility relations and the set of worlds) of a Kripke model determine some features of belief, such as introspection, factivity, etc., irrespective of these conditions, what is modelled is an idealised sense of belief.[7] In particular, these models uphold the requirements of deductive closure, consistency and intensionality for beliefs and belief change, in virtue of transferring features of possible worlds into belief states:

> It is not a huge surprise that logicians who like possible worlds accounts of belief are generally fond of the logical closure and consistency of belief: essentially, the former is just a semantic way of expressing the latter. ([90, p.40])

These assumptions about belief representation and belief change, however, lead to what is referred to in the literature as *the problem of logical omniscience*. Chapter 4 includes a more detailed discussion of the problem of logical omniscience and how it manifests through the closure conditions imposed on belief sets. Vardi comments on how the possible worlds accounts of belief lead to the problem of logical omniscience:

> At the most fundamental level, possible worlds theory is a theory that takes alternative possibilities as its basic primitive notion. While this theory is controversial in some circles, [w]e are willing to accept it. The only assumption that this theory makes is that there are many conceivable states of affairs. Hintikka [w]ent further to model knowledge and belief as a relation between

[6] For a discussion of advantages of object-level formalisations, see [93]. The syntactic machinery of the AGM change operations offer single-shot belief change, as they take as argument a specified belief set and do not define a revised selection function, and they are not suited to model nested belief change or beliefs about the dynamic dispositions of a doxastic state. Although simple combinations of change can be represented, see the supplementary postulates in [1]. The base-generated revision models are able to represent iterated belief revision, but not nested formulas of belief change.

[7] Stalnaker notes that we can understand this extreme idealisation in two different ways: i)knowledge in the ordinary sense but as it applies to idealised agents, or ii) idealised sense of knowledge that is implicit knowledge as it applies to ordinary agents [143, p.171]

> these conceivable states. According to this approach, at any state an agent has in mind a set of states that are possible relative to that state (the set of possible states is a subset of the set of all conceivable states). It is this set of possible states that captures the agent's knowledge or belief. Unfortunately, this way of capturing epistemic notions is far from being intuitive, and goes a long way beyond the basic assumption underlying the possible worlds theory. Thus by modeling knowledge and belief the way he did, Hintikka made a dubious metaphysical commitment, whose side-effect is the logical omniscience problem. [165, p.295]

Modelling solutions to the problem of logical omniscience vary, as well as the sources of the problem. Fagin et al. present a list of reasons that might be responsible for the gap between the ideal of rational reasoning and the way ordinary agents might reason [45, p.40]:

1. Lack of awareness
2. People are resource bounded
3. People don't always know the relevant rules
4. People don't focus on all issues simultaneously

In this dissertation, various solutions for shortening the gap between ideal and ordinary reasoning are proposed, through the models developed in Chapter 2. First and foremost, the proposed models diverge from the intensional approaches to belief revision and advance an hyperintensional approach instead.

1.2 Hyperintensionality

Modalities representing intentional mental states such as knowing, believing, and imagining are understood to generate *hyperintensional contexts*, which call for an individuation of meanings (of sentences) beyond their intentions (i.e., classical truth conditions), and in a way that is sensitive to other epistemically relevant dimensions of meaning. Intensional semantics on the other hand equate the meanings of sentences with their truth conditions, determined by sets of worlds where the sentences are true. Consequently, these semantics cannot distinguish between necessarily true sentences (true in all possible worlds), nor the contingently equivalent sentences true in all the same worlds based on a fixed domain of possible worlds. They also determine a unique set of worlds, namely the empty set, for falsity. Following Cresswell, hyperintensional contexts are simply defined as contexts where logical equivalences are not respected [36].[8]

[8]Note, however, intensional equivalences and logical equivalences coincide only when the domain of possible worlds include all logically possible worlds.

The debate about the inadequacy of using intensional semantics to model epistemic attitudes goes back to [32, 96, 97, 36, 140]. It is understood that intensions (i.e., sets of possible worlds) are not proper designations for meanings of sentences and intensional identity does not account for identity of meaning in hyperintensional contexts. Various modeling solutions have been proposed, which usually start with identifying the relevant dimensions of meaning. Recent literature on hyperintensional belief revision is focused on supplementing the intensional meaning of sentences with an *aboutness* component, that is *what the sentences are about* or *the issues embedded* in their meaning [171, 99, 19, 175]. Consequently, they require *aboutness preservation* for logical entailment in hyperintensional contexts, as an additional dimension to truth preservation. The survey and comparison of various attempts in this tradition is the subject of Chapter 3.

The hyperintensional belief revision models developed in Chapter 2 diverge from the approaches that depend on the (supplemented) possible worlds semantics by opting instead for a state-semantics for information dynamics. In particular, these models build upon the hyperintensional state-semantics *HYPE* developed in [91]. HYPE-models and propositional HYPE-logic can be used as a background semantics and logic for modeling hyperintensional contexts and modalities in particular in virtue of the partialness and the mereological characterisation of the HYPE-states that are the primary semantic elements of the models.

1.3 Fragmented belief

While Chapter 3 is aimed at locating the belief revision models developed in this work within the hyperintensionality literature, Chapter 4 is aimed at locating them within the literature on *fragmented belief*. Fragmented belief refers to situations where the total belief state of an agent at a time is fragmented into separate belief states. Consequently, the belief set representing the totality of an agent's beliefs is not necessarily closed under logical consequence, independent of whether the fragments are internally closed. Moreover, although the fragments are consistent in themselves, they may not be mutually consistent, allowing the total belief state to contain inconsistent beliefs without being trivial. Fragmented belief states are proposed as a modeling solution to deal with some of the closure requirements that are related to the logical omniscience problem, in particular with requirements of deductive closure, conjunctive closure and consistency.

Deductive closure of belief sets requires, briefly, that agents believe all logical implications of their beliefs, including all logical tautologies, furthermore, that their belief sets are closed under arbitrary conjunctions. It is argued that these are not reasonable rationality standards for ordinary agents, especially in the context of totality of their beliefs. Kyburg discusses, for instance, the limitations of requirements such as the totality of an agent's inductive knowledge to be represented in a *single fat statement*, and that the agent has to believe the conjunction of everything she has a right to believe [88]. Stalnaker and Lewis argue that agents might act on different sets of inferences in different situations and contexts. Some claims against deductive closure follow also from arguments such as the preface paradox and the lottery paradox, where combined with the non-contradiction principle (and consistency-intolerant logics), logical closure leads to trivialisation of belief states. Fragmentation theories then aim to prevent trivialisation by restricting admissible inferences such that the agents cannot reason across (possibly contradicting) fragments.

The exact way fragmentation is modelled in various works varies from introducing centers of reasoning [98, 140], resolution-sensitive models [173], accessibility tables based on elicitation conditions [44], to local reasoning models [45], and models for distinguishing explicit and implicit beliefs [94, 10], use of impossible worlds [23, 77], and of neighborhood models [10]. The models in Chapter 2 relate to this literature by allowing conjunctive closure and closure under believed implications to fail for the totality of beliefs of agents, and by avoiding trivialisation in the face of inconsistent information. While these are not fragmented belief models, Chapter 4 proposes also a direction in modeling fragmentation of beliefs based on them.

1.4 Evidence

Separation of the information level from the belief level, and explicit representation of the *raw information* that is not accepted as belief as part of the doxastic states is not novel to the literature on belief revision. *Evidence models* described by van Benthem and Pacuit [158, 160] propose a similar distinction. The authors use neighborhood semantics to model possibly inconsistent and incomplete states of evidence of agents, and dynamics of evidence, within a matching language for beliefs.[9] On the other hand, the belief base revision (BBR) models introduced in Chapter 2 include only (static and dynamic) modalities of belief. Although the information of the agents plays an important role in determining their belief states, it stays in the background of the belief sentences. In Chapter 5, the structured collections of information are exploited in a different direction, reinterpreting the information states as *sources of information* and defining an evidence modality based on *trusted sources*. The models introduced here deal only with the static aspects of evidence, and they further include a persistent *safe belief* modality based on evidence.

1.5 Belief revision meets scientific change

There is a strong relation between the development of belief revision and real sciences. Belief revision theories offer rational ways to select which part of a theory must be given up so that the new data can be integrated in the new theory. A scientific theory, on the other hand, may be considered as a system of beliefs of a scientist who holds it:

> [A] scientific theory is tested against observations. If the observational data are in conflict with the predictions of the theory, then they falsify the theory. Some

[9] In [160], besides a consistency preserving static belief modality, the authors present an extension to the language that stands for *an agent boldly believes* α, that is similar to the extension for fragmented belief proposed in Chapter 4.

> part or other of the all-embracing scientific theory must be given up. [121, p.5]

Berto indeed notes that the AGM originates from insights about theory revision [20, p.134]. Despite this substantial parallel, the two branches have progressed rather separately, and little work has been done to bring them together. In a recent collection, various modeling solutions have been presented to bring the two forms of theory change together, along with discussions of potential causes for the observed distance between them [117]. The application of belief revision theories to model scientific change has usually been in terms of mirroring changes in scientific theories as changes in belief sets [111, 38, 66, 8, 144]. In Chapter 6 it is argued that this belief-centered take on scientific change is exactly the reason why Thagard claims that belief revision theories are not adequate for representing conceptual change [147, 148]. Specifically, he claims, there is a mismatch between the level of belief sets and the level of conceptual change that is pivotal in scientific revolutions. It is, however, possible to lift the methodology of belief revision theories to the conceptual level, as attempted here.

1.6 Outline

In Chapter 2, a novel hyperintensional semantics for belief revision and a corresponding system of dynamic doxastic logic are introduced. The framework developed here constitutes the core of this dissertation. The rest of the chapters are commentaries on the features of this framework and its place in the philosophical logic literature, or rather applications and reinterpretations. Hence I state the extended literature on the relevant issues in the main text. Chapter 3 is an in-depth survey of the literature on hyperintensional semantics and logics as they apply to belief representation and belief revision. It is aimed at a comparative placement of the models introduced in the previous chapter in the literature. In Chapter 4, the requirements of closure and *unity*, that commonly apply to belief sets are reconsidered from the perspective of their relation to the problem of logical omniscience. A new direction towards modeling fragmented belief states is proposed. Chapter 5 explores an alternative formalisation of the main ideas of Chapter 2, and advances a framework for information-based semantics for intuitionistic logic with two novel consistency-sensitive epistemic modalities. Finally, Chapter 6 introduces an application of the framework developed in Chapter 2, as a model for conceptual change in scientific revolutions.

The focus of discussion shifts between static belief representation and belief dynamics throughout this dissertation. While Chapter 2 and 3 focus on dynamics of belief, Chapter 4 is concerned mainly with the (static) closure principles of belief sets. Chapter 5 includes only static evidence and belief modalities, as the dynamic aspects of these are left for future work. Chapter 6 on the other hand, is concerned with conceptual change.

Parts of this dissertation is based on previously published work. Chapter 2 is based on the paper "A Semantics For Hyperintensional Belief Revision Based on Information Bases", published in Studia Logica [29]. The content of Chapter 5 is joint work with Vít Punčochář, Marta Bílková, and Thomas M. Ferguson and it is based on parts of the paper “Consistency-Sensitive Epistemic Modalities in Information-Based Semantics”, published in Studia Logica [126]. The content of Chapter 6 is joint work with Matteo De Benedetto and it is based on the paper “Taking Up Thagard’s Challenge: A Formal Model of Conceptual Revision”, published in Journal of Philosophical Logic [30].

Chapter 2

A Semantics For Hyperintensional Belief Revision Based On Information Bases

In this chapter a novel hyperintensional semantics for belief revision and a corresponding system of dynamic doxastic logic are introduced.[1] The main goal of the proposed framework is to reduce some of the idealisations that are common in the belief revision literature and in dynamic epistemic logic. A state-semantics is used to represent potentially incomplete or inconsistent collections of information, which are the primitive elements of the proposed models. By shifting the representational focus of doxastic models from belief sets to collections of information, and by defining changes of beliefs as artifacts of changes of information, the current work is aimed at achieving a flexible account of belief representation and belief change. The proposal includes dynamic revision operators which represent a non-classical way of changing beliefs: belief revision occurs in non-explosive environments which allow for a non-monotonic and hyperintensional belief dynamics. A logic that is sound with respect to the semantics is also provided.

2.1 Introduction

The doxastic models constructed here are primarily aimed at reducing some of the idealisations that are common in the belief revision literature. These include assumptions about the closure conditions imposed on belief sets such as deductive closure, i.e., agents believe, or at least they are committed to believe, all logical implications of their beliefs, including all logical tautologies. More radically, it is sometimes assumed that the information of an agent, which is the foundation of her beliefs, is complete in the sense that it says something

[1]The content of this chapter is based on the paper "A Semantics For Hyperintensional Belief Revision Based on Information Bases", published in Studia Logica [29].

about every aspect of the world, a feature which possibly passes on to her beliefs.[2] I will also challenge the idea that the primary elements of doxastic representations are belief sets, and that belief changes take place directly on these entities. I show that by shifting the representational focus one step back to the possibly inconsistent and incomplete collections of information (which are not necessarily accepted as beliefs), and by defining changes of beliefs as artifacts of the changes of information, we can achieve a more realistic account of belief representation and of belief change.[3]

The AGM [1] has revolutionised the belief revision literature by introducing a fully formed theory that combines non-monotonic reasoning and belief change. It has significantly influenced the succeeding works on belief revision, especially in terms of the modeling idealisations mentioned above. In particular, the AGM closure postulate requires that belief sets are logically closed before and after belief revision and contraction ($K * \alpha = Cn(K * \alpha)$). Considering such deductively closed belief sets as the primary elements of doxastic representations has been largely criticised as they appear to be too large to be the direct objects of belief change [47, p.41-57]. The theory also suggests that there is only one form of inconsistency, that is trivialisation, and a belief set becomes trivial (i.e., it implies everything in the language) when it is revised with an inconsistent sentence ($K * \alpha = Cn(K * \alpha)$). Last but not least, the AGM change operators treat classically logically equivalent sentences in an equal manner, hence suggesting an intensional theory of belief revision (if $\alpha \equiv \beta$ then $K * \alpha \equiv K * \beta$). We should also note the much debated AGM recovery postulate, which states that merely derived beliefs can be retained even after their bases (supporting beliefs) are withdrawn ($K \subseteq Cn((K \dot{-} \alpha) \cup \{\alpha\})$) .The AGM has also been represented with object-linguistic change operations in several works, contributing to the establishing the Dynamic Epistemic Logic (DEL) and the Dynamic Doxastic Logic (DDL) traditions.[4]

The theories of *base-generated revisions* [132, 68] revoke some of these idealisations established in the AGM paradigm. Base-generated revision models are built on possibly inconsistent sets of beliefs that are not necessarily closed under logical implication, which are called the *belief bases*. Inferential closure of a belief base still serves as the set of sentences an agent is committed to believe.[5] The crucial aspect of base-generated revisions is that changes of belief take place on the belief bases. It is also allowed that agents enjoy inconsistent collections of information without having trivial belief sets, and they can reject making revisions with respect to inconsistent information, hence limiting the AGM success principle [132]. Furthermore, the recovery postulate is no longer satisfied as

[2] As noted, this is rather a radical assumption even among the highly idealised theories of belief revision. Most standard models are still able to represent belief revision via incomplete information.

[3] The idea that changes of belief should take place on entities that are significantly smaller than belief sets is one of the primary motivations behind the theories of base-generated revisions. Rott and Hansson are among the leading figures in the construction of these theories in the philosophical belief revision literature [132, 68].

[4] [138, 164, 137] are considered the first examples of modeling the AGM belief change within the object language and the starting point of the DDL tradition. See also [155, 162]

[5] This claim is accurate only for Rott's theory [132]. Hansson does not assume the closure of belief sets under logical implications [68].

the base-generated revisions stipulate that merely derived beliefs are not kept for their own sakes once the support beliefs are withdrawn. Since pieces of information can stand and fall together, the syntactic structure of information plays a role in determining the belief dynamic. While belief base revision theories also assume an intensional context for belief change, they offer a more fine-grained modeling of belief revision than what the AGM is capable of. Overall, the dynamic operators of base-generated revisions are more general than the AGM change operators (e.g., while the AGM belief revision operations do not allow iterated revisions, it is possible to represent them with the base-generated revision operators.[6] While there is little work towards an object-linguistic representation of base-generated revision operators, the following can be seen a step forward, in virtue of shared motivations and mechanism of belief change.[7]

In the following, I propose a hyperintensional version of base-generated revisions, formalised in the object language following the DEL tradition. Hyperintensionality in the belief revision context means that (classically) logically equivalent content may point out to different change policies of the belief sets, depending on how they are represented in a model. For instance, all (classical) logical tautologies and semantic or mathematical truths are intensionally equivalent (e.g., the following sentences are pairwise intensionally equivalent: "either $x = y$ or $x \neq y$", "all husbands are married", "every integer is the sum of four squares"). This means, within frameworks which are insensitive to hyperintensionality, if an agent comes to believe that all husbands are married (she may have just learned the meaning of the word "husband"), she also comes to believe that every integer is the sum of four squares. This suggests a controversial form of belief dynamics, since learning a piece of vocabulary does not necessarily provide the means to grasp a mathematical truth. A specific form of hyperintensional belief revision has recently been investigated by Berto and Özgün [19, 175]. The authors argue that hyperintensionality occurs in belief revision due to subject matter sensitivity. They obtain the hyperintensionality results without disowning classical logic, hence the works are very significant in integrating the *framing effects*, or hyperintensionality, into the classical approaches of logic of belief revision. On the other hand, the present work assumes a weaker logic and a non-standard semantics.

The representation of doxastic states primarily based on possibly incomplete and inconsistent collections of information is motivated by the assumption that prior to a belief set, an agent possesses possibly incomplete and inconsistent information about the world. These collections of information are formalised in the models via states in a state-space. The use of states to represent incomplete sets of data can be traced back to the *Situation Semantics* developed in [15], while more recent examples include [50, 91]. In the following, the states are characterised by a valuation function which maps them to the literals in the language, and by a fusion function which structures the state-space as a join semi-lattice. The fusion function represents also the dynamic dispositions of an agent by specifying the possible ways of expanding her information. To the point of (static) characterisation of

[6] See [47] for a comparison of the two approaches.

[7] Emiliano Lorini's works on epistemic logic with belief bases has been brought to my attention as one of the projects in this area, as they exploit the belief bases for an epistemic logic of implicit and explicit belief [106].

the states and the state-space, the models are based on the *HYPE-semantics* [91].

Over and above exploiting the HYPE-semantics, the doxastic models introduced here include (uniform) preference orderings between sets of states. A preference ordering represents an agent's epistemic preferences over various collections of information. It determines which parts of the available information are accepted as beliefs by the agent. The current formalisation of epistemic preferences diverges from the most common examples in the literature due to formal concerns which will be pointed out in the next section. Outside of this framework, they are usually defined between sets of worlds (see Grove's seminal work [62] in which he generalises the epistemic preference relations, and the primary works of the DEL tradition such as [155]), or they are defined between possible worlds (e.g, in [159], where the authors model preference change).

A model-shifting dynamic operation (formalised with an object-linguistic change operator) is proposed to model belief revision with new information. This operation is carried out on the possible collections of information accessible by the agent, rather than on the respective belief sets. The changes of belief follow from the changes on these information collections. The change operation preserves the structure of the state-space along with the valuations, however, it alters the preference ordering and the information of the agent. Keep in mind that this is rather a simple revision operation, as most of the complexity of reasoning is sustained by the static belief modality. Supported by the non-classical features of the static models, the revision operation suggests a non-classical way of changing beliefs. In particular, belief revision occurs in a non-explosive environment which also allows for a non-monotonic and hyperintensional belief dynamics. While the belief sets that are formed within this framework are not necessarily closed under logical implication, the proposed system of belief revision is still subject to some idealising assumptions. Most importantly, it is assumed that the belief set of an agent is always consistent, and this is manifested by the consistency axiom for belief. The proposed system satisfies the axiom schemas of cumulative transitivity (cut) and cautious monotony, which are usually desired as common properties of various non-monotonic logics [56]. In the following only revising belief sets is modelled, and contracting beliefs are left for future work.

I start the next section setting the preliminaries for the static portion of the *belief-base-revision models*, and expand these with a revision operation in Section 2.3. In section 2.4 a sound axiomatisation of the logic of belief change is presented based on the these models. Section 2.5 is about some (negative) principles of belief base revision, such as non-monotonicity, non-explosiveness and hyperintensionality.

2.2 Preliminaries

I start with the static belief-base models with the intention of expanding them to revision models later on. A static belief-base model represents the doxastic state of a single agent at a time. A (static) doxastic state is the space of all possible belief states of an agent. It is represented by a state-space, structured with a parthood ordering between the states. The models build upon HYPE-models in [91], expanding them with a preorder for epistemic preference ordering on sets of states.

The following set of formulas specifies the language L_B for the static portion of the models. The modal operator B is the static belief operator:

- $\phi ::= AT \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid B\phi$.
- $\top$ (always true) and $\bot$ $(\neg\top)$ (always false).
- $L_{prop} \subseteq L_B$ is the modality-free portion of L_B.
- $l \subseteq L_B$ is the set of literals for L_B such that $l = \{p, \bar{p}, q, \bar{q}, ...\}$. ($\bar{p}$ is used to denote $\neg p$ and $\bar{\bar{p}}$ denotes p again).

Definition 2.2.1. *A belief-base model* is a tuple $M = \langle S, V, \circ, \bot, \leqslant \rangle$ such that

- S is a non-empty and finite state-space, and the states are denoted by "s_i" with or without the subscript.
- V is a mapping from S to the power set $\mathcal{P}(l)$ of literals. $V(s)$ is the local content of a state s.
- $\circ$ is a binary partial function from $S \times S$ to S that satisfies the following conditions:
  - If $s \circ s'$ is defined, it is required that $V(s \circ s') \supseteq V(s) \cup V(s')$
  - $s \circ s$ is always defined and is equal to s (idempotence)
  - if $s' \circ s$ is defined, then $s \circ s'$ is also defined, and $s' \circ s = s \circ s'$ (commutativity)
  - if $(s \circ s') \circ s''$ is defined, then $s \circ ((s \circ s') \circ s'')$ is defined, and $s \circ ((s \circ s') \circ s'') = (s \circ s') \circ s''$
- $\bot$ is a binary symmetric relation on S that satisfies the following conditions:
  - For all $v \in l$, if $v \in V(s)$ and $\neg v \in V(s')$ then $s \bot s'$
  - if $s \circ s''$ is defined and $s' \circ s'''$ is defined, if $s \bot s'$ then $s \circ s'' \bot s' \circ s'''$
- For all $s \in S$, there is a unique $s^* \in S$ *(the star image of s)* such that
  - $V(s^*) = \{\bar{v} | v \notin V(s)\}$
  - $s^{**} = s$
  - $s^* \not\bot s$
  - if $s' \not\bot s$ then $s' \circ s^*$ is defined, and $s' \circ s^* = s^*$
- $\leqslant$ is a total (transitive, reflexive and connected) preorder on $\mathcal{P}(S)$. For all $A, B \subseteq S$, if $A \leqslant B$ we say that A is at least as preferred as B.

In the above definition $v, \bar{v}$ are used as metavariables for the literals of the language. The literals that a state is mapped to via V represent the atomic pieces of information. The models allow mapping of a state to a set of contradictory literals, such as $\{p, \bar{p}\}$. Such states are called *glutty*. This assumption allows us to represent real world scenarios where agents have contradictory information about their world. It is also usually the case that the agents have incomplete information about the world. Representation of such scenarios are possible by allowing the existence of *gappy* states: a state s is gappy iff for some $p \in l$, neither p nor its negation is in $V(s)$.

Via the fusion function, the states may overlap with each other, be part of other states, or be the product of two or more states fused together. The fusion function determines a partial order on the states which structures the state-space in a join semi-lattice.

Definition 2.2.2. Given a belief-base model M on a state-space S and the state s, s' in S, $s' \sqsubseteq s$ iff $s \circ s' = s$.

If $s' \sqsubseteq s$, it is said that the state s' is part of the state s. The *parthood ordering* $\sqsubseteq$ is reflexive, transitive and antisymmetric.[8] In this framework, the parts of the states are as important as their local contents in their characterisation. That is, states with the same local content are not necessarily identical, they can be distinct states in virtue of their parts.[9]

The $\perp$ relation is an incompatibility relation between the states. The incompatibility of two states may become manifest through contradictory literals $(p, \bar{p})$ in their local content. The star operation is known from the relevance logic [43]. It is however, not a primitive element of the models. Its existence depends on the assumption that the models are rich enough to include s^* whenever they include s. The star image gives the largest compatible state for each state in a model. Its existence means that the ideal agents are capable of expanding their information to a maximally consistent collection, within the limits of the language.[10]

The preorder $\leqslant$ represents the epistemic preference ordering between sets of states. An epistemic preference ordering represents the agent's dispositions for making rational selections among collections of information. Defining the preference ordering on sets of states, rather than on states, simplifies the models significantly. Reflexivity and transitivity are common characteristics of orderings which are used for making rational choices. It is furthermore stipulated that this is a connected order. When applied to real agents, a connected preference ordering means the agents always prefer some collections of information over the others. As a model assumption, this allows one to avoid cases where the agents have access to some information yet fail to form beliefs because of their (lack of) preferences. Lastly, the preference ordering in a belief-base model is not relativized to the states. The shifts in epistemic preferences of the agents are represented only via the model-shifting dynamic operations.

[8] See [91] for proof.
[9] See example 2.3.3.
[10] For a detailed discussion of the formal aspects of the star image see [91].

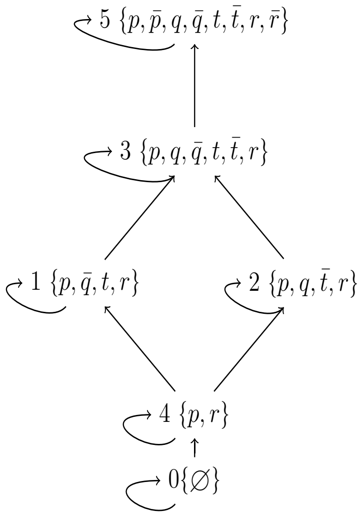

Figure 2.1: A belief-base model. The nodes represent the states in the state-space S, and the arrows represent the (transitive) parthood ordering on S.

Example 2.2.1. The first example is a static belief-base model (pictured in Figure 2.1) which displays the basic principles stated for the construction of such a model. Let $l = \{p, \bar{p}, q, \bar{q}, t, \bar{t}, r, \bar{r}\}$ be the set of literals of interest for this example. Let $S = \{0, 1, 2, 3, 4, 5\}$ be the state-space of model M with $V(0) = \varnothing$, $V(1) = \{p, \bar{q}, t, r\}$, $V(2) = \{p, q, \bar{t}, r\}$, $V(3) = \{p, q, \bar{q}, t, \bar{t}, r\}$, $V(4) = \{p, r\}$, $V(5) = \{p, \bar{p}, q, \bar{q}, t, \bar{t}, r, \bar{r}\}$. Let the fusion function of M be as the reflexive and transitive closure of the following: $\{0 \circ 4 = 4, 4 \circ 1 = 1, 4 \circ 2 = 2, 1 \circ 3 = 3, 2 \circ 3 = 3, 3 \circ 5 = 5\}$. Let the incompatibility relation in the model be determined via the literals. Thus, we have, $1 \perp 2, 1 \perp 3, 1 \perp 5, 2 \perp 1, 2 \perp 3, 2 \perp 5, 3 \perp 1, 3 \perp 2, 3 \perp 3, 3 \perp 5, 4 \perp 5, 5 \perp 1, 5 \perp 2, 5 \perp 3, 5 \perp 4, 5 \perp 5$. So, the following holds for the star images of the states: $0^* = 5, 1^* = 1, 2^* = 2, 3^* = 4, 4^* = 3, 5^* = 0$. Let the preference ordering $\leqslant_M$ be such that for all $A, B \subseteq S$ it holds that $A \leqslant_M B$, i.e., all sets of state in M are preferred equally.

Recall that states in a belief-base model represent collections of information, and a belief-base model represents the doxastic state of an agent. That is, the states stand for the collections of information about the world, the agent possibly possesses. An important notion throughout this paper is the *information base* of an agent. An information base consists of a set of states in the state-space, structured by the epistemic preference ordering and the parthood ordering, and which has an upper bound with respect to the latter. Consider the set of states $H = \{0, 1, 4\}$ from the above example. The state 1 is the upper bound of this set according to the parthood ordering of the model since $1 \circ (0 \circ 4) = 1$. The set of states $H = \{0, 1, 4\}$ is also ordered by the epistemic preference ordering $\leqslant_M$. Let $\mathcal{H}$ denote the corresponding information base. If $\mathcal{H}$ is the current information base of the agent, then her total information is given in the state 1 since by the model assumptions of the fusion function, the local content (i.e., the propositional or non-belief content) of the state 1 includes the local contents of its parts.[11] In the proposed framework, the upper bound of an information base is also where the beliefs of the agent is located, given that information base. To keep things simple, I will often say that an information base $\mathcal{H}$ is determined by a state s, if s is the upper bound of the set of states in the information base according to the parthood ordering. If I intend to address only the (structured) information of the agent at time t, I will talk about the information base at time t. If, on the other hand, I mean to refer to the beliefs of the agent together with their information, I will talk about their *belief state* at time t. In this sense, I will also say that the belief state of the agent at time t is determined by the state s that is the upper bound of the set of states in the relevant information base. Note that, as a possible location of the total information and beliefs of an agent, each state s in a belief-base model determines a possible information base, hence a possible belief state of the agent.

As mentioned at the beginning of this section, the doxastic state of the agent includes possible belief states, which may fall outside of her belief state at time t. A belief-base model may then include states that are not parts of the information base or the belief state of the agent at time t, since an information base does not necessarily exhaust the state-space. It might be that only some of the states in the model are available to the agent through information growth, while the others are not. In particular, the states which are located above the information base of the agent according to the parthood ordering partly determine the dynamics of her belief state at time t: they indicate which collections of information are possibly available to the agent via information growth. As for the other states included in a model which are not parts of the agent's current information base, and which are not available to the agent via information growth (in virtue of the partialness of the fusion function), these states allow one to make hypothetical cases about what the agent would accept as beliefs, and how she would change her beliefs accordingly, if the available information base is such and such. In the most basic case, this indicates which pieces of information the agent cannot learn, by virtue the states which are not connected to her current information base via the parthood ordering.[12]

One of the aims of this framework is to model the assignment of consistent belief sets to possibly inconsistent information bases. Hence, the agents do not necessarily believe everything in their information base. A consistency aiming, cautious process for determining a belief set starts with identifying the *consistent parts of an information base.* Informally, (consistent) parts of an information base amounts to the (consistent) chunks of an agents total information.[13] Formally, a consistent part of an information base is a pairwise con-

[11]Note that the belief content of a state may not include the belief contents of its parts. This is a symptom of the non-monotonicity of the logic of belief revision determined by the proposed framework. In particular, only the propositional content is preserved through information growth, i.e., up the parthood ordering.

[12]These hypothetical cases are limited by how the information is represented in the models in terms of the states that include the information and the parthood ordering between these states.

[13]We talk about grouping the information of an agent only in order to form consistent chunks. One might think that reasoning also involves the parting of the information with respect to subject matter. However plausible this assumption might be, in this paper we focus on a very simple model of reasoning, working with relatively small collections of information. However, a framework which involves grouping

sistent set of states (i.e., for all states s, s' in said set, it holds that $s \not\perp s'$) *within* the information base. A maximality principle is in play in identifying these parts in order to keep the amount of information loss at a minimum in the transition from information bases to the belief sets. *Maximally consistent parts of an information base* are its consistent parts that cannot be expanded within the information base by the addition of more states without breaking the pairwise consistency. When there are multiple maximally consistent parts of an information base, the preference ordering marks off the *best* maximally consistent parts of an information base.

The pieces of information which are given in *all* of the best maximally consistent parts of an information base will constitute the belief set for that base. This definition indicates that the belief operator is a box-like modal operator, hence posing another layer of maximality.[14] The following definition formally specifies the consistency, maximality and preference requirements mentioned above. The support-condition for modal formulas resorts heavily to this definition. The clauses in Definition 2.2.4 are based on the HYPE-logic [91], except for the modal clause.

Definition 2.2.3. Given a belief-base model M on a state-space S,

- *A state $s \in S$ is consistent* (in M) iff $s \not\perp s$. Otherwise it is inconsistent.
- *A set of states $A \subseteq S$ is consistent* (in M) iff for all $s, s' \in A$, $s \not\perp s'$. Otherwise it is inconsistent.
- A set of states $A \subseteq S$ is *maximally consistent with respect to a state $s \in S$* (in M) iff A is consistent, for all $s' \in A$ it holds that $s' \sqsubseteq s$, and for all $s'' \in S$ if $s'' \sqsubseteq s$ and $s'' \notin A$ it holds that $A \cup \{s''\}$ is inconsistent.
- The *best sets of states* in a set $I \subseteq \mathcal{P}(S)$ (in M) are given by the following:

  $min_{\leqslant_M}(I) = \{A \in I | \forall\ B \in I,\ A \leqslant_M B\}$.
- The *best of a state s* (in M) is given by the following:

  $Best_M(s) = min_{\leqslant_M}(\{A \subseteq S \mid A \text{ is maximally consistent w.r.t. } s\})$.

Definition 2.2.4. Given a belief-base model M on a state-space S, for all $s \in S$, the support-conditions for formulas of L_B are as follows:

of information based on topic might open up the discussion to another form of hyperintensional belief revision.

[14]In this framework, the use of maximally consistent parts of an information base transforms the common diamond-like modality of belief into a box-like modality, while producing similar semantical and logical results. For instance, [26] propose a diamond-like knowledge operator in a framework developed with similar motivations of reducing the idealisation of reasoning in epistemic settings. Their framework is also based on structures such as states - called (partial) information states and a (parthood-like) ordering on them. However, they impose a *mutual* consistency requirement while identifying the consistent parts of an information state, instead of maximal consistency. The box-like belief operator is also a reminiscent of the inference operator of Rott's base-generated revision system [132].

$M, s \vDash v$ iff $v \in V(s)$
$M, s \vDash \neg v$ iff $\neg v \in V(s)$
$M, s \vDash \neg\phi$ iff for all s', if $s' \vDash \phi$ then $s \bot s'$[15]
$M, s \vDash \phi \wedge \psi$ iff $s \vDash \phi$ and $s \vDash \psi$
$M, s \vDash \phi \vee \psi$ iff $s \vDash \phi$ or $s \vDash \psi$
$M, s \vDash \phi \to \psi$ iff for all s', if $s \circ s' = s'$ and $s' \vDash \phi$ then $s' \vDash \psi$
$M, s \vDash B\phi$ iff for all $A \in Best_M(s)$, there is $s' \in A$, $s' \vDash \phi$
$M, s \vDash \top$

The support-condition of the biconditional is as usual: $s \vDash \phi \leftrightarrow \psi$ iff $s \vDash \phi \to \psi$ and $s \vDash \psi \to \phi$. $s \vDash \phi$ says that *the state s supports ϕ*. When there is need for specifying the models, we write $s \vDash_M \phi$ and say that *the state s supports ϕ in the model M*.

Based on the definitions above, an agent's belief set consists of the sentences that are supported by all of the best maximally consistent sets of states (by some state in these sets), which are parts of her information base. Formally, the proposed belief modality is a reminiscence of the (non-monotonic) partial-meet operations used to define the AGM contraction and revision, as well as the base-generated revisions and contractions by Hansson and Rott (see [1] for the partial-meet contraction and revision operations, and see [68] and [132] for their application on possibly non-closed sets; for a more general discussion of partial-meet consequence relations and non-monotonicity see [109]).

An important feature of the proposed belief modality is that its objects are the collections of information, represented by the states, rather than the pieces of information (whereas in the above mentioned applications of similar inference operations, the objects are singleton sentences). This, for instance, makes the following scenario possible. Suppose that in a belief-base model, a piece of information ϕ is only supported in a ψ-state, while ϕ is not logically entailed by ψ. That is, the information that ϕ is available to the agent only with the additional information that ψ. Suppose all ψ-states contradict with the current belief state of the agent. Hence, it might be the case that ϕ is not accepted as belief only because the collection of information of which it is a part of (i.e., the ψ-theory) is refuted. The intuition here is that the circumstances surrounding a piece of information matters. Acquiring pieces of information in isolation from other pieces of information mostly occurs in idealised states. Usually, the agents are confronted with possibly incomplete and inconsistent theories about the world, and it is not always reasonable to believe only a part of a refuted theory on the basis that that particular part is not directly refuted. Some pieces of information stand and fall together. For instance, consider reading a certain newspaper. Suppose you are heavily set on your belief that any piece of information given in this paper is highly doubtful, and generally incorrect. Thus, when encountered with a piece of information ϕ, which looks reasonable, due to the non-logical circumstances around this piece of information, such as other information that comes along with it, you do not accept it as a belief.[16]

---

[15] Lemma 8 in [91] shows that $s \vDash \bar{v}$ iff for all $s' \in S$, if $s' \vDash v$ then $s \perp s'$. The lemma is also satisfied in my framework.

[16] To see the formal possibility of such scenarios, consider a belief-base model M on the state-space

A unique belief set for each state $s \in S$ can be defined as follows:

Definition 2.2.5. Given a belief-base model M on a state-space S, and $s \in S$, K_s is the set of beliefs supported by s:
$K_s = \{\phi \in L_B | s \vDash B\phi\}$

Finally, logical consequence and truth in belief-base models are defined as usual:

- $\phi_1, \ldots, \phi_n \vDash \psi$ iff for all models $M = \langle S, V, \circ, \perp, \leqslant \rangle$, for all states $s \in S$, if $s \vDash \phi_1, \ldots, \phi_n$ then $s \vDash \psi$.

- $\vDash \psi$ iff for all models $M = \langle S, V, \circ, \perp, \leqslant \rangle$, for all states $s \in S$, $s \vDash \psi$.

Example 2.2.2. This example shows how a belief set might be determined in a belief-base model. Suppose the agent, whose total information is given in the state 3 in M in Example 2.2.1, is investigating the responsible person for the robbery of a very valuable book from a personal library. Hence, her information base is determined by the state 3 in M, and equally by the set of states $\{0, 1, 2, 3, 4\}$ in S and $\leqslant_M$. She then has the information that the butler has a key to the library (p) and that there are only two keys to the library (r) (e.g., $4 \vDash p \wedge r$). She also has the information that the maid has a key to the library (t), and that if the maid has a key then gardener does not have a key to the library (e.g., $1 \vDash t \to \neg q$), but also that the gardener has a key to the library (q), and if the gardener has a key then the maid does not have a key to the library (e.g., $2 \vDash q \to \neg t$). Therefore, her information about who possess a key to the library is contradictory.

Start with identifying the consistent parts of the agent's information base. In M, there are two maximally consistent sets of states w.r.t. the state 3: $\{0, 1, 4\}$ and $\{0, 2, 4\}$. Based on the preference ordering of M, it holds that $\{0, 1, 4\} \leqslant_M \{0, 2, 4\}$ and $\{0, 2, 4\} \leqslant_M \{0, 1, 4\}$. So, both sets are among the *best of the state* 3: $Best_M(3) = \{\{0, 1, 4\}, \{0, 2, 4\}\}$. By the support-condition for the belief formulas, then, $3 \vDash B(p \wedge r) \wedge B((q \wedge \neg t) \vee (t \wedge \neg q))$. Therefore, the agent believes that there are only two keys to the library and the butler has a key to the library. She also believes that either the maid or the gardener has a key, but not both.

The following are some observations concerning the belief-base models and the belief sets determined via these models.

Lemma 2.2.1. [The implication lemma] Given a belief-base model M on a state-space S, and $\phi, \psi \in L_B$, for all $s \in S$, if $s \vDash \phi \to \psi$ then it holds that if $s \vDash \phi$ then $s \vDash \psi$.

$S = \{1, 2, 3, 4\}$, on a language whose literals are $l = \{p, \bar{p}, q, \bar{q}, s, \bar{s}\}$. Let $V(1) = \{p, q, \bar{q}\}$, $V(2) = \{p, s\}$, $V(3) = \{p, q, \bar{q}, s\}$, $V(4) = \{p, s, \bar{s}\}$. Let $(1 \circ 2) \circ 3 = 3$, $2 \circ 4 = 4$, $1 \circ 1 = 1, 2 \circ 2 = 2, 3 \circ 3 = 3, 4 \circ 4 = 4$. Finally, let $1 \perp 1, 3 \perp 3, 4 \perp 4, 1 \perp 3, 2 \perp 4, 3 \perp 4$. Thus, it holds that $1^* = 4$ and $2^* = 3$. Given that the current information of the agent is given in the state 1, the agent does not believe that p (their belief set is empty) although p is among the information of the agent. However, if the current information of the agent is given in the state 3, they believe that p. This is because, in the former case, the information that p is available only as part of an inconsistent theory. Whereas, in the latter case, it is also available as part of a consistent and unrefuted theory.

Proof. The lemma follows from the idempotence of $\circ$ and the support-condition for $\rightarrow$. □

Lemma 2.2.2. [Persistency for non-modal formulas] Given a belief-base model M on a state-space S, for all $s \in S$ and for all $\phi \in L_{prop}$, if $s \models \phi$ and $s \circ s' = s'$, then $s' \models \phi$.

Proof. See lemma 9 in [91, p.31]. □

Observation 2.2.1. [Non-persistency of modal formulas] The modal formulas of the language L_B are not necessarily persistent through parthood ordering in a belief-base model M.

Proof. See the model in Example 2.2.1. It holds that $1 \models Bt$ since $Best_M(1) = \{\{0, 4, 1\}\}$ and $1 \models t$. It also holds that $1 \circ 3 = 3$, however, $3 \not\models Bt$. □

Lemma 2.2.3. Given a belief-base model M on a state-space S, for all $s \in S$ it holds that $Best_M(s)$ is non-empty.

Proof. Let M be an arbitrary belief-base model on a state-space S, and let s be an arbitrary state in this model. Suppose for some $s' \in S$ with $s' \not\perp s'$ it holds that $s' \sqsubseteq s$. That is, s has some consistent parts. So, it follows that there is an $A \subseteq S$ such that A is maximally consistent w.r.t. s and $s' \in A$. By the assumptions for the preference ordering, either $A \in Best_M(s)$ or there is an $A' \subseteq S$ such that A' is maximally consistent w.r.t. s and $A' \leqslant_M A$ (and $A \not\leqslant_M A'$). In the latter case however, it holds that $A' \in Best_M(s)$. Therefore, if s has some consistent part, it cannot be the case that $Best_M(s) = \varnothing$. Now suppose for all $s' \sqsubseteq s$ it is the case that $s' \perp s'$. That is, s has no consistent parts. In this case, it follows from the Definition 2.2.3 that $Best_M(s) = \{\varnothing\}$. Hence, $Best_M(s)$ has the empty set of states as its unique member. Since M and s are arbitrary, this holds for all belief-base models. □

Lemma 2.2.4. Given a belief-base model M on a state-space S, for all $s \in S$ and for all $\phi \in L_B$, if $s \not\perp s$, then $s \models \phi$ iff $s \models B\phi$.[17]

Proof. Let M be an arbitrary belief-base model on a state-space S, and let $s \in S$ an arbitrary state such that $s \not\perp s$. The proof for the direction from left-to-right is simple.

[17] Another observation, which might be worrying to some, follows from this lemma in combination with the support-condition for disjunction. The present lemma predicts that an agent whose total information is consistent, believes a disjunction iff she believes one of the disjuncts, provided that both disjuncts does not involve a conditional sub-formula. One might however, think that an agent can be in a consistent doxastic state and believe, for instance, that the butler has a key to the library or the maid has a key to the library, without being decidedly opinionated about either of the disjuncts. Hence, the current system puts the criteria for an agent to believe a disjunction without necessarily believing one of the disjuncts as her having contradictory information (not necessarily directly about either of the disjuncts). While the justification of this state is not very clear, one should keep in mind that completely consistent information bases is the most idealised scenario in this framework. Moreover, an agent can still believe a disjunction (when both disjuncts involve some conditional sub-formula) without believing one of the disjuncts in case she has inconsistent ways of expanding her information base, i.e., if there are inconsistent collections of information (states) with which she can combine (fuse) her current information base.

Suppose $s \models \phi$. By assumptions for the incompatibility relation, since $s \not\perp s$, it follows for all s', s'' in S that if $s \circ (s' \circ s'') = s$ then $s' \not\perp s''$. Hence, there is a unique set $A \in Best_M(s)$ such that for all $s' \sqsubseteq s$ it holds that $s' \in A$. Therefore, since $s \sqsubseteq s$, it follows that $s \in A$, and since $s \models \phi$, it holds that $s \models B\phi$ as desired.

For the direction from right-to-left, we prove by cases. For the first case suppose $\phi \in l$ when l is the set of literals in L_B. So, suppose $s \models B\phi$. Hence, for all $A \in Best_M(s)$ there is an $s' \in A$ such that $s' \models \phi$. Pick an arbitrary $A \in Best_M(s)$ and an arbitrary $s' \in A$ with $s' \models \phi$. Since $s' \circ s = s$, by the persistency lemma it follows that $s \models \phi$ as desired.

For the second case, suppose ϕ is in the form $\neg\psi$ for some $\psi \in L_B$. So, suppose $s \models B\neg\psi$. Thus, for all $A \in Best_M(s)$ there is an $s' \in A$ such that $s' \models \neg\psi$. Pick an arbitrary $A \in Best_M(s)$ and an arbitrary $s' \in A$ with $s' \models \neg\psi$. Consider an arbitrary $s'' \in S$ such that $s'' \models \psi$. By the clause for negation, it follows that $s' \perp s''$. Since $s \circ s' = s$, by the assumptions for the incompatibility relation, it holds that $s \perp s''$. Since s'' is arbitrary, again by the clause for negation, it follows that $s \models \neg\psi$.

For the third case, suppose ϕ is in the form $\psi \to \chi$ for some $\psi, \chi \in L_B$. So, suppose $s \models B(\psi \to \chi)$. So, for all $A \in Best_M(s)$ there is an $s' \in A$ such that $s' \models \psi \to \chi$. Pick an arbitrary $A \in Best_M(s)$ and an arbitrary $s' \in A$ with $s' \models \psi \to \chi$. For reductio, suppose $s \not\models \psi \to \chi$. Hence, there is an $s'' \in S$ such that $s \circ s'' = s''$, $s'' \models \psi$ but $s'' \not\models \chi$. By the assumptions for the fusion function, it also holds that $s' \circ s'' = s''$. So, it is the case that $s' \not\models \psi \to \chi$. However, this contradicts with the assumption that $s' \models \psi \to \chi$. Therefore, it follows that $s \models \psi \to \chi$.

For the forth case, suppose ϕ is in the form $B\psi$ for some $\psi \in L_B$. So, suppose $s \models BB\psi$. Thus, for all $A \in Best_M(s)$ there is an $s' \in A$ such that $s' \models B\psi$. Pick an arbitrary $A \in Best_M(s)$ and an arbitrary $s' \in A$ with $s' \models B\psi$. It follows that, for all $A' \in Best_M(s')$ there is an $s'' \in A'$ such that $s'' \models \psi$. By the assumptions for the fusion function, it also holds that $s'' \circ s = s$. Since $s \not\perp s$, there is a unique set $A \in Best_M(s)$ and $s'' \in A$ (see the proof for the left-to-right direction). Therefore, $s \models B\psi$ as desired.

The cases for formulas in the form of conjunctions and in the form of disjunctions follow by induction using the above cases. Since M and s are arbitrary, this holds for all belief-base models.

□

Lemma 2.2.5. Given a belief-base model M on a state-space S, for all $s \in S$ and for all $\phi \in L_B$, if $s \models B\phi$ then $s \models \phi$.

Proof. Let M be an arbitrary belief-base model on a state-space S, let $s \in S$ an arbitrary state. The proof is then similar to the proof of right-to-left direction of Lemma 2.2.4. We only modify the case for the formulas in the form $B\psi$ for some $\psi \in L_B$. We show that if $s \models BB\psi$ then $s \models B\psi$. So, suppose $s \models BB\psi$. Thus, for all $A \in Best_M(s)$ there is an $s' \in A$ such that $s' \models B\psi$. Pick an arbitrary $A \in Best_M(s)$ and an arbitrary $s' \in A$ with $s' \models B\psi$. Thus, for all $A' \in Best_M(s')$ there is an $s'' \in A'$ such that $s'' \models \psi$. Consider an arbitrary $A' \in Best_M(s')$ and an arbitrary $s'' \in A'$ with $s'' \models \psi$. We know by the assumption for the fusion function that $s'' \circ s = s$. For reductio, suppose for some $B \in Best_M(s)$, $s'' \notin B$. So, by the maximality of $Best(s)$, it follows that there is a $u \in B$

such that $u \perp s''$. By the assumptions for the incompatibility relation, it follows that also $s' \perp u$ (since $s'' \circ s' = s'$). However, since $A \in Best_M(s)$ and $s' \in A$ are arbitrary, it follows that B is inconsistent. Since this contradicts with our model assumptions, it should be the case that $s'' \in B$ for all $B \in Best_M(s)$. Therefore, $s \vDash B\psi$. Since M and s are arbitrary, this holds for all belief-base models. □

Lemma 2.2.6. Given a belief-base model M on a state-space S, for all states $s \in S$, K_s is always consistent, i.e., for all $\phi \in L_B$, it cannot be the case that $s \vDash B\phi \wedge B\neg\phi$ or that $s \vDash B(\phi \wedge \neg\phi)$.

Proof. Let M be an arbitrary belief-base model on a state-space S, let $s \in S$ an arbitrary state. First we show that for all $\phi \in L_B$, $s \nvDash B\phi \wedge B\neg\phi$. We proceed by induction on the complexity of the formulas of L_B. The base case is when $\phi \in l$ with l is the set of literals for L_B. For reductio, assume $s \vDash Bp \wedge B\neg p$. So, it is the case that for all $A \in Best_M(s)$, there is a state u in A such that $u \vDash p$, and also a state u' in A such that $u' \vDash \neg p$. Since $u \perp u'$, it follows that A is inconsistent. Hence, a contradiction follows from the requirement of the models that for all $B \in Best(s)$ it holds that B is consistent. Therefore, it cannot be the case that $s \vDash Bp \wedge B\neg p$.

Next we prove for the formulas in the form $\neg\psi$ for some $\psi \in L_B$. So, for reductio assume $s \vDash B\neg\psi \wedge B\neg\neg\psi$. So, it is the case that for all $A \in Best_M(s)$, there is a state u in A such that $u \vDash \neg\psi$, and also a state u' in A such that $u' \vDash \neg\neg\psi$. Pick an arbitrary $A \in Best_M(s)$ and arbitrary $u, u' \in A$ with $u \vDash \neg\psi$ and $u' \vDash \neg\neg\psi$. By the support-condition for negation (since $u' \vDash \neg\neg\psi$) it follows that for all $s' \in S$, if $s' \vDash \neg\psi$ then $u' \perp s'$. So it follows that $u' \perp u$ and that A is inconsistent. Hence, a contradiction follows from the requirement of the models that for all $B \in Best(s)$ it holds that B is consistent. Therefore, it cannot be the case that $s \vDash B\neg\psi \wedge B\neg\neg\psi$. The rest of the cases can be proved easily by the above cases. I leave them for the reader.

To show that for all $\phi \in L_B$, $s \nvDash B(\phi \wedge \neg\phi)$, we again use proof by induction. I only state the base case and leave the rest of the cases out for space issues. The base case is when $\phi \in l$ with l is the set of literals for L_B. For reductio, assume $s \vDash B(p \wedge \neg p)$. So, it is the case that for all $A \in Best_M(s)$, there is a state u in A such that $u \vDash p \wedge \neg p$. So, it follows that $u \perp u$ and A is inconsistent. Hence, a contradiction follows from the requirement of the models that for all $B \in Best(s)$ it holds that B is consistent. Therefore, it cannot be the case that $s \vDash B(p \wedge \neg p)$. □

Observation 2.2.2. There is a belief-base model M on a state-space S, and a state $s \in S$ such that $s \vDash B\phi \wedge \neg B\phi$.

Proof. Let M be a belief-base model constructed on the state-space $S = \{1, 2, 3\}$, and on a language L_B with the literals $l = \{p, \bar{p}\}$. Let $V(1) = \{p, \bar{p}, q\}, V(2) = \{p, q\}, V(3) = \{q\}$. Let $1 \circ 1 = 1, 2 \circ 2 = 2, 3 \circ 3 = 3, 2 \circ 3 = 2, 1 \circ 2 = 1, 1 \circ 3 = 1, (2 \circ 3) \circ 1 = 1$; and let the incompatibility relation as the following: $1 \perp 1, 1 \perp 2, 2 \perp 1$. So, it follows that $1 = 3^*, 2 = 2^*, 3 = 1^*$. Finally, let $\leqslant_M$ is such that for all $A, B \subseteq S$, $A \leqslant_M B$.

We show that $1 \vDash Bp \wedge \neg Bp$. There is a unique maximally consistent set of states w.r.t. 1, that is the set $\{2, 3\}$. By the connectivity of the preference ordering it holds that

$Best_M(1) = \{\{2,3\}\}$. Since $2 \vDash p$, it holds that $1 \vDash Bp$. It also holds that $2 \vDash Bp$: since $Best_M(2) = \{\{2,3\}\}$ and $2 \vDash p$. However, since $Best_M(3) = \{\{3\}\}$ and $3 \nvDash p$, it follows that $3 \nvDash Bp$. Since $1 \perp 1$ and $1 \perp 2$, and since 1 and 2 are all and only states which support Bp, by the support-condition for negation, it holds that $1 \vDash \neg Bp$. Therefore, by the support-condition for conjunction, $1 \vDash Bp \wedge \neg Bp$. □

These lemmas and observations are used to simplify some proofs later on. They are more important, however, as indicators of some of the consequences of the framework. Lemma 2.2.1 shows that the proposed conditional ($\rightarrow$) appeals to the intuition of conditional reasoning.[18] Lemma 2.2.2 and Observation 2.2.1 reflect the non-monotonic nature of beliefs, even though the non-belief content is persistent through information growth. Lemma 2.2.3 shows how the trivial, or inconsistent belief sets are blocked by the models since an empty $Best_M(s)$ for a state s in a model M would lead to a trivial belief set which is equal to the language L_B. Lemma 2.2.4 says that when the total information of an agent is consistent (in itself), the agent believes every part of her information base. The equality of the information and the beliefs is the ideal belief state in the proposed frameworks. Lemma 2.2.5 says, on the other hand, regardless of the consistency of information, the agent believes only what is part of her information. Lemma 2.2.6 indicates the consistency of beliefs as a strong property of the proposed framework. Observation 2.2.2 comments on the previous one stating that "believe that ϕ" and "not believe that ϕ" are not contradictory. What the latter means may present a lengthy discussion, I only want to highlight that "not believe that ϕ" is not same as "believing that not ϕ".

2.3 Belief revision

Recall the agent, the investigator from the examples in the previous section. Suppose initially she believes that the butler has a key to the library (p), that there are only two keys that could open the library (r), and also that the maid has a key to the library (t). Hence, we assume that her information base is determined by the state 1 in the model M in Example 2.2.1. Suppose she thereafter learns that the gardener may have stolen the maid's key ($(q \wedge \neg t) \vee (t \wedge \neg q)$). This section introduces how an agent should revise her beliefs with new information, within a dynamic framework that is be constructed based on the belief-base models above. In this context, revision means adding new beliefs to a belief set while preserving its internal consistency.[19] As preservation of the consistency of the beliefs is already achieved by the static aspects of the models, the dynamic part covers

[18]The conditional ($\rightarrow$) is, however, stronger than the common intuition when the other direction is considered, the two expressions "if $s \vDash \phi$ then $s \vDash \psi$" and "$s \vDash \phi \rightarrow \psi$" are not equivalent.

[19]Other common forms of belief change are belief contraction and belief expansion. Belief expansion is the operation of adding new beliefs without a concern for restoring the consistency of the new belief set. The current framework does not allow the construction of inconsistent belief sets (see Lemma 2.2.6). Belief contraction is the operation of eliminating some of the beliefs from a belief set. I leave the formalisation of belief contraction via belief-base models for future work.

the expansion of the information base with new information and revision of the preference ordering accordingly.

The belief base revision (BBR) models are constructed by expanding the static belief-base models with a revision relation. This is a relation from a state in a belief-base model and a formula in the language, to a new state in a new belief-base model. In particular, the state which determines the initial information base of an agent is expanded to another state to incorporate the new information. In this way, a shift occurs to a new information base. At the same time, the epistemic preference ordering of the agent changes to ensure that the new information is accepted as a new belief, hence the shift to a new model. In this transition, the structure of the initial belief-base model is preserved except for the preference ordering. Since the state-space of the initial model is among what is preserved, the existence of the expanded information base is a precondition for belief base revision.

While various forms of revising beliefs can be defined which differ in terms of the severity and the range of effect, particularly on the preference ordering (see [133] for various ways of changing beliefs within the DDL framework), only one option is presented here. The revisions are represented in the object language with a pair of dynamic modal operators. Crucially in this setting, there may not be a unique revised model as the result of a revision. Therefore, a box-like revision operator and dual a diamond-like revision operator are introduced. The language L_B is then extended to include formulas of the form $[\psi]\phi$ and $\langle\psi\rangle\phi$, where ψ is a brackets-free formula, constituting the new language L_D. The box-like dynamic operator means that the right-hand-side (sub)formula is supported by *all* of the revised models after the revision with the left-hand-side (sub)formula, and the diamond-like dynamic operator means that the right-hand-side formula is supported by *some* of the revised models.[20]

Definition 2.3.1. The following set of formulas specify the language L^D:

- $\phi ::= AT \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \to \phi \mid B\psi \mid [\psi]\phi \mid \langle\psi\rangle\phi$, where ψ is a bracket-free formula
- $\top$ (always true) and $\bot$ $(\neg\top)$ (always false)

Note that we do not allow bracketed formulas to appear in the brackets or after the modality B. On the other hand, the right-hand-side subformula of the revision formulas can contain brackets, hence allowing us to represent in the object language iterated revisions of the form $[\psi][\chi]\phi$, where, again, ψ and χ are brackets-free.

We can now define the belief base revision models with a dynamic revision relation between models that share a state-space, a valuation function, a fusion function, and an

[20]The literature on indeterministic belief change focuses on approaches of belief revision that allows revisions to result in multiple new models. The approach is motivated by the idea that there may be more than one admissible way of changing ones beliefs, none of whom necessarily a better option than the others. Indeterministic belief change is also referred to as *relational belief change*. For various motivations leading to the investigation of relational belief change operations and indeterministic belief change see [41], [102] and [104], and the discussion by [116]. Hansson states that indeterministic belief change confirms most of the results and expectations from deterministic models [68].

incompatibility relation, with possibly different preference orderings. Let us call such class of models MOD. A revision relation R is then simply a relation between the members of MOD. We first define revision models with general revision relations, and briefly specify the relation R to define the belief base-revision models.

Definition 2.3.2. *A pre-model* is a tuple $M^P = \langle MOD, R \rangle$ such that

- MOD is a class of belief base models with identical $S, V, \circ, \perp$ and possibly different $\leqslant$ and
- R is a relation from a triple $\langle \phi, M, s \rangle$ to pairs $\langle M', s' \rangle$ such that
  - $\phi \in L_B$, $M = \langle S, V, \circ, \perp, \leqslant \rangle \in MOD$,
  - $M' = \langle S, V, \circ, \perp, \leqslant' \rangle \in MOD$,
  - $s, s' \in S$.

The initial model M and the revised model M' differ only with respect to the preference ordering, while the rest of the structure of the initial model is preserved during the revisions. We will use the notion $sR^{\phi}s'$ to denote $\langle \phi, M, s \rangle R \langle M', s' \rangle$ when the context allows.

Definition 2.3.3. The following extends the support conditions on definition 2.2.4 to include the new formulas in L_D:

- $M, s \models [\phi]\psi$ iff $\forall \langle M', s' \rangle : \langle \phi, M, s \rangle R \langle M', s' \rangle$ it holds that $M's' \models \psi$
- $M, s \models \langle \phi \rangle \psi$ iff $\exists \langle M', s' \rangle : \langle \phi, M, s \rangle R \langle M', s' \rangle$ and $M', s' \models \psi$.

Again, the right-hand-side subformula of a revision formula may include brackets to denote iterated revisions. Accordingly, given that ψ, χ are brackets-free, the support condition for an iterated revision formula is constructed as follows: $M, s \models [\psi][\chi]\phi$ iff $\forall \langle M', s' \rangle : \langle \psi, M, s \rangle R \langle M', s' \rangle$ it holds that $M', s' \models [\chi]\phi$ iff $\forall \langle M'', s'' \rangle : \langle \chi, M' , s' \rangle R \langle M'' , s'' \rangle$ it holds that $M'', s'' \models \phi$. The support condition for iterated diamond-formulas are similarly constructed. The iterations are then executed starting with the outermost bracketed formula. It is also allowed that formulas with the modality B can be embedded in the revision formulas, in the form $[B\psi]\phi$, where ψ is brackets-free. Revision with a belief formula in the current framework is a particular case since it mixes two different preference relations on the right-hand-side and on the left-hand-side of the revision formula.

Some new terminology will be used for the specification of the revision relation. For all $\phi \in L_D$, a *basic ϕ-state* in a belief base model M is a state which satisfies ϕ (*a ϕ-state*) and which do not have any parts other than itself which are also ϕ-states. These are the smallest ϕ-states in a model, and taken as the unique sources of the new information. This restriction is proposed in line with the well-known *minimal change principle*: while revising (and contracting) a belief set, the changes that occur in the new belief set shall be minimal. That is, one should only add (or delete) the beliefs which are necessary for the intended change to be successful.

Definition 2.3.4. Given a belief base model M on a state-space S with $s \in S$, for all $\phi \in L_D$, s is a *basic ϕ-state* (in M) iff $s \models_M \phi$ and for all $s' \sqsubseteq s$, if $s' \neq s$ it holds that $s' \not\models_M \phi$.

It is important to note that, for $\phi \in L_D$, there might be multiple basic-ϕ-states. It is also possible that a model M does not contain any basic-ϕ-states. The former observation leads to indeterministic belief change, in the sense that there might be more than one way to revise a state in a model with a formula ϕ, hence based on a triple $\langle \phi, M, s \rangle$, R might determine multiple pairs $\langle M', s' \rangle$. The latter observation means that revision with a formula in a model might not always be possible.

We can now define a maximal revision relation R which meets the conditions specified below.

Definition 2.3.5. *A belief base revision model* is a tuple $M^D = \langle MOD, R \rangle$ such that

- MOD is a class of belief base models with identical $S, V, \circ, \perp$ and possibly different $\leqslant$ and
- R is a relation from a triple $\langle \phi, M, s \rangle$ to pairs $\langle M', s' \rangle$ such that
  - $\phi \in L_B$, $M = \langle S, V, \circ, \perp, \leqslant \rangle \in MOD$,
  - $M' = \langle S, V, \circ, \perp, \leqslant' \rangle \in MOD$,
  - $s, s' \in S$,
  - $\exists t \in S$ such that t is a basic ϕ-state in M and it holds that $(s \circ t) \circ s' = s'$ such that there is no $s'' \in S$ with $s'' \neq s'$, $s'' \circ s' = s'$, and $(s \circ t) \circ s'' = s''$, otherwise no pair $\langle M', s' \rangle$ stands in the relation R to $\langle \phi, M, s \rangle$,
  - for all $A, B \subseteq S$, if there is a $u \in A$ with $u \models_M \phi$ and for all $u' \in B$, $u' \not\models_M \phi$, then $A \leqslant_{M'} B$, and if for all $u \in A$, $u \not\models_M \phi$ and there is a $u' \in B$ with $u' \models_M \phi$, then $A \nleqslant_{M'} B$; otherwise $A \leqslant_{M'} B$ iff $A \leqslant_M B$.[21]

The existence requirement in the above definition states that there is a basic ϕ-state (t) whose fusion with the initial state (s) is defined, and s' is the lowest in the parthood ordering which includes the fusion $s \circ t$. Hence, s' is the smallest state that we can pick as the revised state.

Observation 2.3.1. R is a partial relation.

[21] The current definition of the revision relation mixes syntactic aspects of a model with the semantics. This is a choice I made, in order to have a general relation rather than a family of relations, indexed to the formulas of the language, e.g., $\mathcal{R} = \{R'_\phi : \phi \in L_B\}$. Application of the latter formulation could also be considered, however, it is likely to generate differences in the logic of belief revision, in particular when nested or iterated revisions are in question.

Proof. Consider $M^D = \langle MOD, R \rangle$ and let $M = \langle S, V, \circ, \perp, \leqslant \rangle \in MOD$ with $S = \{s, s'\}$, let $\circ$ be empty, and let for all $A, B \subseteq S$ it holds that $A \leqslant_M B$. We limit the language L_D with $l \subseteq L_D = \{p, \bar{p}\}$. Suppose $V(s) = \{p\}$ and $V(s') = \{\bar{p}\}$. Therefore, $s \perp s'$ holds in M^D, and $s^* = s$, $s'^* = s'$. Suppose we want to revise s with the sentence $\neg p$. Since there is no basic $\neg p$-state in S whose fusion with s is defined, a revised state cannot be determined in S. Therefore, there are no pairs R-related to the triple $\langle \neg p, M, s \rangle$. □

Example 2.3.1. For the first revision example, consider again the scenario in Example 2.2.2. In order to indicate how the preference order is affected by iterated revisions, multiple revision processes are included in this example. Suppose, at the beginning, the investigator is ensured by the owners (only) that there are only two keys to the library, and the butler has one of those. Assume at this point that the agent's information base is determined by the state 4 in the model M. Assume also a different preference ordering for the model M from the one given in Example 2.2.2. Since "there are only two keys to the library" and "the butler has a key to the library" constitute the only information the agent has so far, let her preferences be such that all sets of states which include a $(p \wedge r)$-state are preferred over the ones which does not include such a state. For simplicity, the rest of the preference ordering is fixed as plain such that all sets of states are preferred equally. Suppose at the first phase of the investigation the agent is informed that the second key is held by the maid (t).

Start with identifying the basic t-states: the state 1 is the unique basic t-state in S. Since $1 \circ 4 = 1$ holds in M, the information base of the agent after the expansion is determined by the state 1. Her epistemic preference order is then adjusted so the sets of states which include a t-state are preferred over the ones which do not include any t-states, and the sets of states which do not include any t-states are no longer preferred over the ones which include some t-states. The rest of her preferences remain as in the beginning. Call this new model with the revised preference ordering M'. M' then satisfies the following: $\{1\} \leqslant_{M'} \{4\} \leqslant_{M'} \{2\} \leqslant_{M'} \{4\} \leqslant_{M'} \{0\}$ while $\{4\} \nleqslant_{M'} \{1\}$ and $\{0\} \nleqslant_{M'} \{4\}$.

The revised belief set is then determined by the state 1 in M'. There is a unique maximally consistent set of states w.r.t. 1: $Best_{M'}(1) = \{\{0, 4, 1\}\}$. It follows that $1 \vDash_{M'} B(p \wedge t) \wedge B(r \wedge \neg q)$. Therefore, after the revision, the agent believes that the butler and the maid has the only two keys to the library, while the gardener does not have a key. As M' is the unique model for this revision, it follows that $4 \vDash_M [t](B(p \wedge t) \wedge B(r \wedge \neg q))$.

Suppose at the second phase of the investigation, the agent is told of the owners suspicions about whether or not the gardener stole the maid's key. She then wants to revise her belief set with the information that either the maid has the second key or the gardener has it. Given that her current doxastic model is represented at M', there are two ways she can use this information to change her beliefs. That is because both 1 and 2 are basic $(q \vee t)$-states in M'. Call these revised models obtained by expanding her information base with the state 1 and with the state 2, M^1 and M^2 respectively.

At M^1, her expanded information base is again determined by the state 1 since $1 \circ 1 = 1$. According to the revised preference ordering, the sets of states which include a $(q \vee t)$-state are preferred over the ones which do not include any $(q \vee t)$-states, and the sets of states which do not include any $(q \vee t)$-states are no longer preferred over the ones which include some $(q \vee t)$-states. The rest of the preferences remain as in M'. Note that with the revision, since all t-states are also $(q \vee t)$-states, they remain minimal in the preference ordering on the subsets of S. However, since all q-states are also $(q \vee t)$-states, they also move among the most preferred in M^1. There is, however, a unique maximally consistent set of states w.r.t. 1, thus, $Best_{M^1}(1) = \{\{0,4,1\}\}$. So, it is the case that $1 \vDash_{M^1} (B(p \wedge r) \wedge B(q \vee t)) \wedge B(t \wedge \neg q)$.

At M^2, her expanded information base is determined by the state 3 since $1 \circ 2 = 3$. According to the revised preference ordering, the sets of states which include a $(q \vee t)$-state are preferred over the ones which do not include any $(q \vee t)$-states, and the sets of states which do not include any $(q \vee t)$-states are no longer preferred over the ones which include some $(q \vee t)$-states. The rest of the preferences remain as in M'. (Hence the preference ordering of M^2 is identical to that of M^1.) There are two maximally consistent sets of states w.r.t. the state 3, these are the sets $\{0,1,4\}$ and $\{0,2,4\}$. Based on the revised preference ordering $\leqslant_{M^2}$, we have that $\{0,1,4\} \leqslant_{M^2} \{0,2,4\} \leqslant_{M^2} \{0,1,4\}$ since both 1 and 2 are $(q \vee t)$-states. Hence, $Best_{M^2}(3) = \{\{0,1,4\},\{0,2,4\}\}$. So, it follows that $3 \vDash_{M^2} B(p \wedge r) \wedge B(q \vee t)$. Therefore, after revising her beliefs, the agent still believes that there are only two keys to the library and the butler has a key to the library, however, she no longer believes that the maid has a key to the library. She also does not believe that the gardener has a key to the library, while she believes that either one of them has the second key.

We express this indeterministic way of changing beliefs with the help of the diamond-like revision operators in the language: $1 \vDash_{M'} [q \vee t](B(p \wedge r) \wedge B(q \vee t)) \wedge \langle q \vee t \rangle B(t \wedge \neg q)$. That is, after revising her beliefs with the disjunction, the agent believes the disjunction $(q \vee t)$, while believing neither q nor t, and there is a way of changing her beliefs in which she also comes to believe that t and also $\neg q$ as a result (as in the revised model M^1).

Example 2.3.2. This example shows how indeterministic belief change is interpreted in this framework. First, let the literals of the language L_D be limited to $l = \{p, \bar{p}, q, \bar{q}\}$. Consider a belief base revision model $M^D = \langle S, V, \circ, \perp, \leqslant, R\rangle$ based on $S = \{1,2,3,4\}$ with $V(1) = \varnothing, V(2) = \{p, \bar{q}\}, V(3) = \{p, q\}, V(4) = \{p, \bar{p}, q, \bar{q}\}$. Let the following be defined on S: $1 \circ 1 = 1, 1 \circ 2 = 2, 1 \circ 3 = 3, 1 \circ 4 = 4, 2 \circ 2 = 2, 2 \circ 3 = 4, 2 \circ 4 = 4, 3 \circ 3 = 3, 3 \circ 4 = 4, 4 \circ 4 = 4$. Let the incompatibility relation to be given by the literals such that $2 \perp 3, 2 \perp 4, 3 \perp 4$ and $4 \perp 4$. So, the following holds: $1 = 4^*, 2 = 2^*, 3 = 3^*, 4 = 1^*$. Let $M = \langle S, V, \circ, \perp, \leqslant\rangle$. Finally, let the preference ordering $\leqslant_M$ be such that for all $A, B \subseteq S$ it holds that $A \leqslant_M B$.

Suppose the information base of the agent is determined by the state 1 in M. Since $Best_M(1) = \{\{1\}\}$, at this point she only has beliefs in the form of $B(\phi \rightarrow \psi)$. Suppose she learns that p. We show how she should revise her beliefs accordingly. Given the model M, there are two basic p-states: 2 and 3. Hence, there are two ways she can revise her belief set. Either she expands her information base with the state 2, or with the state 3. So, it follows that there are two revised models based on the model M, call them M^2 and M^3 respectively. At each model, the preference ordering of the agent shifts from the preference ordering of the model M in the way that all sets of states which include a p-state are strictly preferred over all sets of states which do not include a p-state.

At M^2, her (new) information base is determined by the state 2. There is a unique maximally consistent set of states w.r.t. 2, so, $Best_{M^2}(2) = \{\{1, 2\}\}$. It follows that $2 \vDash_{M^2} (Bp \wedge B\neg q) \wedge B(q \vee \neg q)$. At M^3, her (new) information base is determined by the state 3. There is a unique maximally consistent set of states w.r.t. 3, $Best_{M^3}(3) = \{\{1, 3\}\}$. It follows that $3 \vDash_{M^3} (Bp \wedge Bq) \wedge B(q \vee \neg q)$. After the revision, the agent believes that p and that $(q \vee \neg q)$. However, it is indetermined whether she believes that q or that $\neg q$: $1 \vDash_M [p](Bp \wedge B(q \vee \neg q)) \wedge (\langle p \rangle Bq \wedge \langle p \rangle B\neg q)$.

Before moving to the last example of belief base revision, note that belief revision is not always successful in the current framework. That is, it is not necessarily the case that an information piece ϕ is accepted as a belief after the revision with ϕ. In fact, neither $\vDash [\phi]B\phi$ nor $\vDash \langle \phi \rangle B\phi$ are valid in this framework. After the revision of a belief set with ϕ, it is accepted as a belief iff there is some consistent part of the revised information base in which ϕ is supported by a state. Moreover, if ϕ is a contradictory sentence in the form $\psi \wedge \neg\psi$ for some $\psi \in L_B$, the revision is bound to be unsuccessful.

Example 2.3.3. The last example indicates that representation of information is more fine grained in the belief base revision models than it is in the traditional belief change models. The models allow the existence of multiple states mapped to the same set of propositional letters since we do not identify the states with their local content. That means, these states may differ in terms of their dynamic aspects although their local contents are the same. Consider the following fractions of belief base revision models; the nodes represent the states in the state-space with their local content given in parenthesis, and the arrows represent the parthood ordering, the parthood ordering should be read as transitively closed.

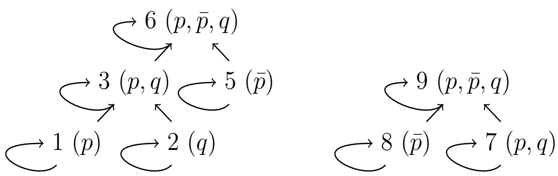

Figure 2.2: Distinct states with identical local contents

The local contents of the states 3 and 7 are equal, although their parts differ. If the belief set supported by the state 3 is revised with the information that $(\neg p)$, the revised belief set would still include q, whereas the same revision on the belief set supported by the state 7 would see both p and q being eliminated from the new belief set.

2.4 Logic

The logic of belief base revision models restricted to the sublanguage L_{prop} is the logic of HYPE.[22]

Theorem 2.4.1. The axioms and rules listed in Figure 2.3 is sound for the system of belief base revision (for all formulas $\phi \in L_D$, if $\vdash \phi$ then $\vDash \phi$).

Proof. To prove soundness, I demonstrate the detailed proofs for selected axiom schemas. Let $M^D = \langle MOD, R\rangle$ be an arbitrary belief base revision model, let $M = \langle S, V, \circ, \bot, \leqslant\rangle$ and let s be an arbitrary state in S.

(11) For reductio, assume $s \nvDash \phi \to \neg\neg\phi$. So, there is an $s' \in S$ with $s' \circ s = s'$ and $s' \vDash \phi$, but $s' \nvDash \neg\neg\phi$. Hence, there is an $s'' \in S$ with $s'' \vDash \neg\phi$ and it holds that $s' \not\perp s''$. However, since $s' \vDash \phi$, it should be the case that $s' \perp s''$. Hence we have a contradiction. Therefore, $s' \vDash \neg\neg\phi$, and it follows that $s \vDash \phi \to \neg\neg\phi$. Since s and M^D are arbitrary, $\vDash \phi \to \neg\neg\phi$ is true in all belief base revision models.

(16) Let arbitrary $s' \in S$ with $s' \circ s = s'$. Suppose $s' \vDash B\phi$. So, for all $A \in Best_M(s')$, there is an $s'' \in A$ with $s'' \vDash \phi$. Consider an arbitrary $A \in Best_M(s')$ and an arbitrary $s'' \in A$ with $s'' \vDash \phi$. Since A is consistent, it follows that $s'' \not\perp s''$. By the Lemma 2.2.4, it holds that $s'' \vDash B\phi$. Since A and s'' are arbitrary, it holds that for all $A' \in Best_M(s')$ there is a $u \in A'$ with $u \vDash B\phi$. Therefore, $s' \vDash BB\phi$. Since s' is arbitrary, it follows that $s \vDash B\phi \to BB\phi$. Since s and M^D are arbitrary, $\vDash B\phi \to BB\phi$ is true in all belief base revision models.

(17) Let arbitrary $s' \in S$ with $s' \circ s = s'$. Suppose $s' \vDash B\phi \vee B\psi$. So, either $s' \vDash B\phi$ or $s' \vDash B\psi$. Suppose the former. Hence, for all $A \in Best_M(s')$, there is an $s'' \in A$ with $s'' \vDash \phi$. Pick an arbitrary $A \in Best_M(s')$ and an arbitrary $s'' \in A$ with $s'' \vDash \phi$. By the support-condition for disjunction, it holds that $s'' \vDash \phi \vee \psi$. Since A and s'' are arbitrary, it follows that for all $A' \in Best_M(s')$ there is a $u \in A'$ with $u \vDash \phi \vee \psi$. Therefore, $s' \vDash B(\phi \vee \psi)$. Similarly, if $s' \vDash B\psi$, it follows that $s' \vDash B(\phi \vee \psi)$. Since s' is arbitrary, it follows that $s \vDash B\phi \vee B\psi \to B(\phi \vee \psi)$. Since s and M^D are arbitrary, $\vDash B\phi \vee B\psi \to B(\phi \vee \psi)$ is true in all belief base revision models.

(18) Let arbitrary $s' \in S$ with $s' \circ s = s'$. Suppose $s' \vDash B(\phi \wedge \psi)$. So, for all $A \in Best_M(s')$, there is an $s'' \in A$ with $s'' \vDash \phi \wedge \psi$. Pick an arbitrary $A \in Best_M(s')$ and an arbitrary $s'' \in A$ with $s'' \vDash \phi \wedge \psi$. It follows that $s'' \vDash \phi$ and also $s'' \vDash \psi$. Since A and s'' are arbitrary, $s' \vDash B\phi \wedge B\psi$. Since s' is arbitrary, it follows that $s \vDash B(\phi \wedge \psi) \to B\phi \wedge B\psi$. Since s and M^D are arbitrary, $\vDash B(\phi \wedge \psi) \to B\phi \wedge B\psi$ is true in all belief base revision models.

[22] A sound and complete axiom system for the HYPE logic is given in [91].

Axioms

1	$\vdash \top$
2	$\vdash \phi \to \phi$
3	$\vdash (\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$
4	$\vdash \phi \wedge \psi \to \phi$
5	$\vdash \phi \wedge \psi \to \psi$
6	$\vdash \phi \to \phi \vee \psi$
7	$\vdash \psi \to \phi \vee \psi$
8	$\vdash (\phi \to \chi) \to ((\psi \to \chi) \to (\phi \vee \psi \to \chi))$
9	$\vdash \phi \wedge (\psi \vee \chi) \leftrightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)$
10	$\vdash \phi \vee (\psi \wedge \chi) \leftrightarrow (\phi \vee \psi) \wedge (\phi \vee \chi)$
11	$\vdash \phi \to \neg\neg\phi$
12	$\vdash \neg\phi \vee \neg\psi \to \neg(\phi \wedge \psi)$
13	$\vdash \neg\phi \wedge \neg\psi \leftrightarrow \neg(\phi \vee \psi)$
14	$\vdash (\phi \to \psi) \to ((\psi \to \chi) \to (\phi \to \chi))$
15	$\vdash (\phi \to (\phi \to \psi)) \to (\phi \to \psi)$
16	$\vdash B\phi \to BB\phi$ (Positive Introspection)
17	$\vdash B\phi \vee B\psi \to B(\phi \vee \psi)$ (Disjunctive closure)
18	$\vdash B(\phi \wedge \psi) \to B\phi \wedge B\psi$ ($\wedge$ distribution)
19	$\vdash [\phi]B\chi \to [\phi]B(\psi \vee \chi)$ (Disjunction)
20	$\vdash [\phi]B(\psi \wedge \chi) \to [\phi]B\psi$ (Simplification)
21	$\vdash \langle\phi\rangle B\chi \to \langle\phi\rangle B(\psi \vee \chi)$ (Disjunction$\langle\rangle$)
22	$\vdash \langle\phi\rangle B(\psi \wedge \chi) \to \langle\phi\rangle B\psi$ (Simplification$\langle\rangle$)
23	$\vdash B\phi \wedge B\neg\phi \to \bot$ (Consistency 1)
24	$\vdash B(\phi \wedge \neg\phi) \to \bot$ (Consistency 2)

Rules

(MP)	$\phi, \phi \to \psi \vdash \psi$
(Cont)	$\dfrac{\vdash \phi \to \psi}{\vdash \neg\psi \to \neg\phi}$
(CM)	$[\phi]B\psi, [\phi]B\chi \vdash [\phi \wedge \psi]B\chi$
(Cut)	$[\phi]B\psi, [\phi \wedge \psi]B\chi \vdash [\phi]B\chi$
(F)	$\dfrac{\vdash B\phi}{\vdash \phi}$

Figure 2.3: A Hilbert style axiomatization of logic of belief base revision

(CM) Let $M, M' \in MOD$ be belief-base models on a state-space S and let s, s' be states in S with $\langle M, s, \phi\rangle R\langle M', s'\rangle$. Suppose $s \vDash_M [\phi]B\psi$. Since all $(\phi \wedge \psi)$-states are ϕ-states, and since $s \vDash_M [\phi]B\psi$, it holds that the set of pairs of models and states obtained from revising s in M with $(\phi \wedge \psi)$ constitutes a subset of the set of pairs of models and states obtained from revising s in M with ϕ (that is, if that all sets of states in M' which include some ϕ-states are preferred to the sets of states in M' which do not include any ϕ-states entails that it is already the case that all sets of states in $Best_{M'}(s')$ also include some ψ-states then, provided that there are any, the maximally consistent sets of states under s' with some $(\phi \wedge \psi)$-states are already among the $Best_{M'}(s')$). Therefore, if $s \vDash_M [\phi]B\chi$, it holds that $s \vDash_M [\phi \wedge \psi]B\chi$. Since s and M^D are arbitrary, the CM rule holds in all belief base revision models.

(Cut) Let $M, M' \in MOD$ be belief-base models on a state-space S and let s, s' be states in S with $\langle M, s, \phi\rangle R\langle M', s'\rangle$. Suppose $s \vDash_M [\phi]B\psi$. So, it holds that $s \vDash_M [\phi](\phi \wedge \psi)$. Moreover, if all sets of states in M' which include some ϕ-states are preferred to the sets of states in M' which do not include any ϕ-states, then it is the case that all sets of states in $Best(s')$ also include some ψ-states. Hence, these are exactly the maximally consistent sets of states under s' with some $(\phi \wedge \psi)$-state, provided that there are any. Therefore, the set of pairs of models and states obtained from revising s in M with ϕ constitutes a subset of the set of pairs of models and states obtained from revising s in M with $(\phi \wedge \psi)$. Since $s \vDash_M [\phi \wedge \psi]B\chi$ holds, all members of $Best_{M'}(s')$ include also some χ-states. Therefore, $s \vDash_M [\phi]B\chi$. Since s and M^D are arbitrary, the cut rule holds in all belief base revision models.

□

The MP rule and the axiom schema 2 follows from the idempotence assumption of $\circ$ and the support-condition for $\rightarrow$. The validities of axiom schemas 3, 8, 12 - 15 can easily be shown via reductio ad absurdum. The validity of the axiom schema 19 follows from 17 and that of 20 follows from 18. For the proofs of 23 and 24 see Lemma 2.2.6. The last rule in the list, labelled (F) for faithful, can be proved using Lemma 2.2.5. The rest can be proved using only the support-conditions. In the above theorem, when possible, the claims which include the belief modality and the dynamic operators are stated as axiom schemas rather than as rules. (For instance, the rule for the positive introspection would be $B\phi \vdash BB\phi$.) By the implication lemma, the proofs for the doxastic axiom schemas entail the proofs for the respective rules.

Lemma 2.4.1. The following deduction theorem is logically valid in BBR models iff the language of the models are restricted to the sublanguage L_{prop} (i.e., when $\phi_1, ...\phi_n, \psi, \chi \in L_{prop}$):
$\phi_1, ...\phi_n, \psi \vdash \chi$ iff $\phi_1, ...\phi_n \vdash \psi \rightarrow \chi$.

Proof. Use (MP) and the axioms schemas (2) and (3) from Theorem 2.4.1, together with the schema $\vdash \phi \rightarrow (\psi \rightarrow \phi)$. The latter is valid in the current framework only when the models are restricted to the language L_{prop}. □

2.5 More properties of belief and belief revision

This section states some principles concerning the belief sets and belief revision, which fail in the proposed framework although they are valid axiom schemas or rules in some of the more common and well-known theories in the literature.

Theorem 2.5.1. The following list of axiom schemas and rules are not valid in BBR models.

1. $B\phi$, $B(\phi \to \psi) \models B\psi$ (Modal modus ponens)
2. $B(\phi \to \psi) \models B\phi \to B\psi$ (K-rule)
3. $\phi \to \psi \models B\phi \to B\psi$ (Monotonicity of belief)
4. $B\phi \wedge B\psi \models B(\phi \wedge \psi)$ (Conjunctive closure)
5. $\neg B\phi \models B\neg B\phi$ (Negative introspection)
6. $B(\phi \vee \psi) \models B\phi \vee B\psi$ ($\vee$ distribution)
7. $\dfrac{\vdash \phi}{\vdash B\phi}$ (Necessitation)

Proof. As a counterexample to the first four principles, consider a belief base revision model $M^D = \langle S, V, \circ, \perp, \leqslant, R\rangle$ on the state-space $S = \{1, 2, 3, 4, 5, 6\}$, and let literals of the language L_D for the model M^D be $l = \{p, \bar{p}, q, \bar{q}, r, \bar{r}, t, \bar{t}, s, \bar{s}\}$. Let $V(1) = \{r, t\}, V(2) = \{p, t\}, V(3) = \{p, \bar{p}, q, \bar{q}, r, t, s, \bar{s}\}, V(4) = \{p, q, \bar{q}, r, \bar{r}, t, s, \bar{s}\}, V(5) = \{p, q, r, t, \bar{t}\}, V(6) = \{p, q, r, s, \bar{s}\}$. Let $((1 \circ 2) \circ 6) \circ 3 = 3, ((1 \circ 2) \circ 6) \circ 4 = 4, (1 \circ 2) \circ 5 = 5, 1 \circ 1 = 1, 2 \circ 2 = 2, 3 \circ 3 = 3, 4 \circ 4 = 4, 5 \circ 5 = 5, 6 \circ 6 = 6$, and let the parthood relation be transitively closed on these fusions. Let the incompatibility relation on S be given via the literals. It follows that $1 = 3^*, 2 = 4^*, 3 = 1^*, 4 = 2^*, 5 = 6^*, 6 = 5^*$. Let $M = \langle S, V, \circ, \perp, \leqslant\rangle$ and finally, for all $A, B \in S$, $A \leqslant_M B$.

1. Substitute ϕ in the schema with p, and ψ in the schema with q. In the model M, it holds that $5 \models Bp$ and $5 \models B(p \to q)$, but $5 \not\models Bq$.
2. Use the same substitution of the formulas. It holds in M, that $5 \models B(p \to q)$, however, $5 \not\models Bp \to Bq$.
3. Similarly, $5 \models p \to q$, however, $5 \not\models Bp \to Bq$.
4. Substitute ϕ in the schema with p, and ψ in the schema with r. In the model M, it holds that $5 \models Bp \wedge Br$, but $5 \not\models B(p \wedge r)$.

As a counterexample to the remaining principles, consider a belief-base model M on the state-space $S = \{1, 2, 3, 4\}$, and let literals of the language L_D for the model M' be $l = \{p, \bar{p}\}$. Let $V(1) = \{p\}, V(2) = \{\bar{p}\}, V(3) = \{p, \bar{p}\}, V(4) = \varnothing$. Let $4 \circ 1 = 1, 4 \circ 2 = 2, 4 \circ 3 = 3, 1 \circ 2 = 3, 1 \circ 3 = 3, 2 \circ 3 = 3, 4 \circ 4 = 4, 1 \circ 1 = 1, 2 \circ 2 = 2, 3 \circ 3 = 3$. Let the incompatibility relation of M be given via the literals, hence $1 \perp 2, 1 \perp 3, 2 \perp 3, 3 \perp 3$. It follows that $1 = 1^*, 2 = 2^*, 3 = 4^*, 4 = 3^*$. Let the preference ordering of M' be such that for all $A, B \in S$, $A \leqslant_{M'} B$.

(5) Substitute ϕ in the schema with p. In the model M', it holds that $3 \vDash \neg Bp$, however, $3 \nvDash B\neg Bp$

(6) Substitute ϕ in the schema with p, and the 'ψ' in the schema with '$\neg p$'. In the model M', $3 \vDash B(p \vee \neg p)$, but $3 \nvDash Bp \vee B\neg p$.

The necessitation rule fails when for some belief-base model M, for some $s \in S$ it holds that $Best_M(s) = \{\varnothing\}$. That is, when s has no consistent parts. Because in this case, for no $\psi \in L_B$ it holds that $s \vDash B\psi$. □

The above principles are stated as rules rather than axioms since by the contraposition of the implication lemma, their failures entail the failures of the respective axiom schemas.

Theorem 2.5.2. The following list of axiom schemas and rules are not valid in belief base revision models.

1. $\vDash [\phi]B(\psi \vee \neg\psi)$ (Excluded middle)
2. $[\phi]B\psi$, $[\phi]B\chi \vDash [\phi]B(\psi \wedge \chi)$ (Adjunction)
3. $[\phi]B(\psi \vee \chi) \vDash [\phi]B\psi \vee [\phi]B\chi$ (Disjunction 2)
4. $[\phi]B\psi$, $[\phi]B(\psi \to \chi) \vDash [\phi]B\chi$ (Closure under belief implication)
5. $[\phi]B\chi \vDash [\phi \wedge \psi]B\chi$ (Monotony)
6. $\neg[\phi]B\neg\psi, [\phi]B\chi \vDash [\phi \wedge \psi]B\chi$ (Rational monotony)[23]

Proof. The invalidity of the first axiom schema follows from the failure of general excluded middle ($\vDash \phi \vee \neg\phi$). The next three invalidities follow respectively from the failures of conjunctive closure, $\vee$ distribution and modal modus ponens in Theorem 2.5.1.

To show the invalidity of the remaining principles, construct the following model. Let $M^D = \langle MOD, R\rangle$ be a belief base revision model. We limit the language L_D with $l \subseteq L_D = \{p, \bar{p}, q, \bar{q}, r, \bar{r}\}$. Let $M = \langle S, V, \circ, \perp, \leqslant\rangle \in MOD$, with $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ such that $V(1) = \{p, r\}$, $V(2) = \{q, \bar{r}\}$, $V(3) = \{p, \bar{p}, q, \bar{q}, r, \bar{r}\}$, $V(4) = \{p, q, \bar{q}, r\}$, $V(5) = \{p, \bar{p}, q, \bar{r}\}$, $V(6) = \varnothing$, $V(7) = \{p, q\}$, $V(8) = \{p, q, r, \bar{r}\}$. Let the following fusions be defined on

[23]The first premise of this argument could be read as the satisfaction of the negated formula, or as the non-satisfaction of $[\phi]B\neg\psi$.

S: $6 \circ 1 = 1, 6 \circ 2 = 2, 6 \circ 7 = 7, (1 \circ 7) \circ 4 = 4, (2 \circ 7) \circ 5 = 5, ((1 \circ 2) \circ 7) \circ 8 = 8, ((4 \circ 5) \circ 8) \circ 3 = 3$, and let the parthood relation be transitively closed on these fusions. Let the incompatibility relation on S be given via the literals. Thus, the following holds: $1 = 4^*, 2 = 5^*, 3 = 6^*, 4 = 1^*, 5 = 2^*, 6 = 3^*, 7 = 8^*, 8 = 7^*$. Finally, suppose $A, B \in S$, $A \leqslant_M B$.

(5) Substitute ϕ in the schema with r, ψ in the schema with q, and χ in the schema with $(r \wedge p)$. We show that $6 \vDash_M [r]B(r \wedge p)$, but $6 \nvDash_M [r \wedge q]B(r \wedge p)$. Suppose we revise the state 6 in M with r. Let M' be the revised model. The revised information base is determined by the state 1 ($6R^r1$). It holds that $Best_{M'}(1) = \{\{1, 6\}\}$, hence $1 \vDash_{M'} B(r \wedge p)$. Therefore, $6 \vDash_M [r]B(r \wedge p)$. Now suppose we revise the state 6 in M with $(r \wedge q)$. There are two ways to do this since both 4 and 8 in M are basic $(r \wedge q)$-states. It suffices to show that in one of the revised models, $B(r \wedge p)$ is not supported by the revised information base. Let M^8 be the revised model through state 8 ($6R^{r \wedge q}8$). It holds that $Best_{M^8}(8) = \{\{1, 6, 7\}, \{2, 6, 7\}\}$, hence $8 \nvDash_{M^8} B(r \wedge p)$. Therefore, $6 \nvDash_M [r \wedge q]B(r \wedge p)$.

(6) It suffices that $6 \nvDash_M [r]B\neg q$ in addition to what we have shown above. In fact, $6 \vDash_M \neg[r]B\neg q$ since $[r]B\neg q$ is not supported anywhere in the model M'. Hence we comply with the first premise in both forms of its reading.

□

Theorem 2.5.3. The following metarules are not valid in belief base revision models. $\Rightarrow$ stands for classical logical entailment.

1. From $\phi \Rightarrow \psi$ infer $[\phi]B\psi$ (Intensionality)

2. From $[\phi]B\psi$, $\psi \Rightarrow \chi$ infer $[\phi]B\chi$ (Right weakening)

3. From $[\phi]B\chi$, $\phi \Leftrightarrow \psi$ infer $[\psi]B\chi$ (Left logical equivalence)

Proof. We use the model constructed above.

(1) Substitute ϕ in the schema with r, and ψ in the schema with $(q \vee \neg q)$. It holds that $r \Rightarrow (q \vee \neg q)$, however, $1 \nvDash_M [r]B(q \vee \neg q)$.

(2) Substitute ϕ in the formula with r, ψ with p, and χ with $(q \vee \neg q)$. It holds that $1 \vDash_M [r]Bp$, and that $p \Rightarrow (q \vee \neg q)$, however, $1 \nvDash_M [r]B(q \vee \neg q)$.

(3) Substitute ϕ in the schema with $(p \vee \neg p)$, ψ in the schema with $(q \vee \neg q)$, and χ in the schema with p. Thus, it holds that $6 \vDash_M [p \vee \neg p]Bp$, and $(p \vee \neg p) \Leftrightarrow (q \vee \neg q)$, however, $6 \nvDash_M [q \vee \neg q]Bp$.

□

To conclude this section some remarks are in order, particularly about the specific form of hyperintensionality manifested here and the (lack of) reduction axioms. Theorem 2.5.3 addresses the influence of classical logic on belief base revision. The invalidities indicate the hyperintensionality of the revision system. That is, intensionally equivalent formulas cannot be substituted within the revision formulas. That we formulated these schemas in the metalanguage referring to two different logics marks an important difference between this framework and the framework for hyperintensional belief revision presented in [19]. Berto formulates the intensionality, right weakening and left logical equivalence rules within his object language. Therefore, the failures of the rules indicate some limitations concerning the influence of the already underlying (classical propositional) logic on belief revision.

Finally, introducing reduction axioms of dynamic formulas to static ones proves challenging in the proposed framework. The formulation of the conditional ($\rightarrow$) successfully hints at some properties of belief revision as it is a forward looking modality. Belief-base revision, in most cases, causes a shift from one state to another. For instance, in order to revise a state s with a piece of information ϕ, we move to the states which expand s with ϕ. These are (some of) the states that are determined by a conditional on s whose antecedent is ϕ. In this respect, the conditional still underdetermines the states relevant for the revision since when revising s, belief base revision relation marks the states which expand s with basic ϕ-states only. The language, however, is not rich enough to allow precoding revisions completely. That is because, the information supported by a state depends in part on the preference ordering of the model. It might be the case that, for some ψ in the language, ψ is not supported by a ϕ-state until after the preference ordering of the model is revised. Although this is the case in most belief revision systems which include changing the preferences, in some of these systems, the revised models can be given as sub-models of the original one. Hence, it is possible to give reduction axioms by referring to a relativized version of the original model. The system presented in [155] is an example to this sort of preference change.

There is an exception to this hardship by the persistency lemma, which ensures that the propositional content of a state does not change via model-shifts. So, a reduction axiom only concerning the formulas of L_{prop} can be presented: given a belief base revision model $M^D = \langle MOD, R\rangle$ with $M = \langle S, V, \circ, \bot, \leqslant\rangle \in MOD$, for all states $s \in S$, for all $\phi \in L_B$ and for all $\psi \in L_{prop}$, $s \vDash_M [\phi]\psi \leftrightarrow (\phi \rightarrow \psi)$. For the base case of the proof, suppose $s \vDash_M [\phi]p$. Hence, for all $s' \in S$, if $sR^{\phi}s'$, it holds that $s' \vDash_M p$. Let arbitrary $s'' \in S$ such that $s' \circ s'' = s''$. By the persistency lemma, also $s'' \vDash_M p$ holds. By the assumptions for the fusion function, it holds that $s'' \circ s = s''$. Since s'' is arbitrary, by the support-condition for $\rightarrow$, it holds that $s \vDash_M \phi \rightarrow p$. For the other direction, suppose $s \vDash_M \phi \rightarrow p$. That is, whenever s in M in expanded with a (basic) ϕ-state, satisfaction of p follows in the expanded state. Hence, for all $s' \in S$, if $sR^{\phi}s'$ then $s' \vDash p$. Therefore, $s \vDash_M [\phi]p$. The validity of the claim for non-atomic propositional formulas of the language can be shown easily by induction on the complexity.

2.6 Concluding remarks

I have presented a new hyperintensional semantics for belief revision, which also allows non-monotonic and non-explosive belief revision. The non-classical features of the revision framework principally follow from the underlying non-classical semantics. In particular, the dynamics of potentially incomplete and inconsistent collections of information is formalised using a form of state semantics. Adoption of states as the principle elements of the models separates my framework from the DEL paradigm. At the same time, the introduction of the revision operators in the object language marks the effective difference between the proposed semantics and the base-generated belief revision theories in the literature. Syntactically, the underlying (propositional) logic of belief base revision is significantly weaker than classical logic. This yields a more adaptable logic of belief representation and belief dynamics.

The belief base revision models reflects a non-classical way of changing beliefs. Its features relate in particular to issues of non-monotonicity, indeterminacy of information, hyperintensional sensitivity, and fragmentation of information. Non-monotonicity of the system is apparent by the items (5) and (6) of Theorem 2.5.1. It is the belief modality that I employ here that causes the non-monotonicity of belief. Indeterminacy of information is specifically related to disjunctive beliefs. The item (6) in Theorem 2.5.1 and its dynamic counterpart, the item (3) in Theorem 2.5.2 indicate that an agent can believe a disjunction without necessarily believing one of the disjuncts. However, believing a disjunction without believing one of the disjuncts is possible only if the agent has inconsistent information, not necessarily about the disjunction in question.

Hyperintensionality of the logic of belief base revision is presented via the item (1) in Theorem 2.5.2 and Theorem 2.5.3. This is a consequence of the partial content of the states. Hyperintensional sensitivity has usually been introduced as subject-matter sensitivity. In this respect, the content of a state can be understood to determine a subject-matter. Hence, although two sentences α and β are classically logically equivalent, while the subject-matter of α is included the subject-matter fixed by a situation s, the subject-matter of β may not be included. Thus, an agent, whose beliefs are determined by the state s may not believe the latter on the basis of the former.

Some models of hyperintensional belief revision reject also the principle of disjunctive closure on the grounds of subject-matter inclusion requirement for belief entailment [19]. It follows from disjunctive closure, that if an agent believes a sentence α, she also believes the disjunction $\alpha \vee \beta$ (for any β). The subject-matter inclusion requirement is such that given that a sentence α logically entails a sentence β, an agent believes that β based on the belief α only if the subject-matter of α includes the subject-matter of β. Briefly, logical entailment of sentences with foreign subject-matters do not carry over to the beliefs. However, disjunctive closure is a valid principle of belief base revision framework that I presented in this work. This is because, although the models are sensitive to the hyperintensional contexts, the requirement for entailment in these contexts is weaker than subject-matter inclusion. In fact, it seems that for logical entailment to carry over to the beliefs of an agent, shared subject-matter between the two sentences suffices. That is, it holds that the agent believes that $\alpha \vee \beta$ based on the belief α since the former is partly about α.

Lastly, one of the most important features of the models is the fragmentation of information, due to the partial fusion function and the partial parthood ordering of the models. The consequences of this structure are presented in the paper as the failures of the principles of conjunctive closure and of the principles of closure under implication, in the items (1)-(4) in Theorem 2.5.1 and in the items (2) and (4) in Theorem 2.5.2. These consequences are quite similar to that of fragmented belief approaches, where it is allowed that the doxastic system of an agent involves different centers of rationality. Thus, the agent's belief state is fragmented such that the agent may believe that α in one fragment and believe that β in another, and not be able to put the two beliefs together. These systems allow also contradictory beliefs located in different fragments. The models presented here falls short of admitting the full consequences of fragmented belief, by virtue of the employment of a partial-meet-consequence-like belief modality and a total epistemic preference ordering. The construction of belief contraction models and a complete axiom system for belief base revision models are also left for future work.

Chapter 3

A Comparative Assessment Of Approaches To Hyperintensional Belief Revision

Belief like other intentional mental attitude modalities such as knowledge, imagining, etc., is a sentential modality. These modalities take as their objects sentences of a given language. Thus, in sentential semantics, sentences are the primary truth-bearers. Semantical analysis and the interpretation of logical structures depend to a large degree on what we take as the meaning or the content conveyed by sentences. Hyperintensional contexts generated by hyperintensional modalities such as the belief modality, suggest that the relevant aspects of the meaning of sentences, in these contexts, go beyond their truth-conditions.

In the following, we investigate hyperintensionality in the context of belief revision. The problems of assuming an intensional semantical analysis of belief revision are manifested in various forms, contributing to the problem of logical omniscience. An intensional analysis, for instance, lacks the resources to individuate logical and necessary truths from one another, particularly from the sentence determined by the set of all possible worlds. This means, that belief in all necessary truths, no matter their content, follow from belief in one necessary truth, for instance from a tautology of classical logic "A or not A". It also lacks the resources to differentiate necessary falsehoods from one another. Developments in the literature on belief revision have suggested that agents might not believe all tautologies of logic, nor all contingent necessities, and they might fail to believe all classical implications of their beliefs regardless of their relevance to each other in terms of content, without being irrational. Moreover, they might, non-trivially, believe (necessarily) false sentences and can distinguish these sentences from other falsehoods. The search for a more fine-grained analysis of the meaning of sentences has then led to the development of hyperintensional semantics and logics for belief revision.

Traditional analysis of logical consequence is based on truth-preservation. In the context of belief revision, however, truth-preservation does not suffice to determine warranted implications between sentences since further dimensions make up the relevant aspects of their meaning. It is then suggested, in these contexts, the consequence relation is strengthen in a way that it accounts also for the non-truth-functional aspects of meaning. This is sometimes done by supplementing the consequence relation with topic-preservation [19], aboutness preservation [171], and issue or partition-preservation [99]. Alternatively, altering the primary elements of representation, what determines a *proposition* expressed by a sentence can be redefined, and the sentences can be individuated in a more fine-grained manner than intensional semantics. A support-based state-semantics, such as the one proposed in Chapter 2 for instance, is able to account for hyperintensional sensitivity within the context of belief revision based on support-preservation.

The following is a limited survey of various modeling solutions proposed in the literature with the aim of capturing hyperintensional sensitivity in relation to belief revision. A comparative assessment is then provided based on issues of topic-transitivity, topic-inclusion, and the notion of semantical opposition, between the support-based models introduced in Chapter 2 and the topic-preserving hyperintensional belief revision models such as proposed in [19].

3.1 Introduction

Lewis defines the meaning of a sentence as what determines the conditions that make the sentence true or false [97].[1] His definition refers to the *intension* of a sentence, which can be defined as a function from the relevant factors that make up the meaning of a sentence to truth values. The meaning of a sentence then depends on what these relevant factors are taken to be, and the means of representation. In the tradition going back to Carnap, and further established by Kripke, alternative states of affairs in the form of possible worlds are used to represent these factors [32, 85]. The way a sentence structures a logical space of possible worlds based on its assigned truth value (true or false) at each world determines the intension of that sentence. The sentences that are true in the same set of possible worlds are then treated identically and can be mutually substituted within *intensional contexts*. Alethic modalities such as necessity and possibility, for instance, are interpreted as *intensional modalities*. In an intensional context, all necessary truths and logical tautologies express the same unique intension, i.e. the same *proposition*, that is the set of all possible worlds. Similarly, all necessary falsehoods express the same unique intension that is the empty set of possible worlds. If, on the other hand, the domain of possible worlds include all logically possible worlds, intensionally equivalent sentences correspond to *logically equivalent sentences* with respect to classical logic.

[1] Defined in this way, the meaning of a sentence is distinct from its *extensional meaning*. The extension of a sentence is its actual or assigned truth value, given a certain states of affairs and an interpretation. Hence, in a sense the extension of a sentence depends on its meaning [97, p.23]. The extension of a sentence can also be defined as a function from a sentence to a truth value. Unary and binary logical connectives $\wedge$, $\vee$, $\neg$, as well as other connectives that are defined via these (such as the material conditional $\supset$ and the biconditional $\leftrightarrow$) are said to create *extensional contexts*. These are contexts where occurrences of two sentences ϕ and ψ with the same extension can be mutually substituted without altering the truth value of the whole sentence.

Modalities that represent intentional mental states such as believing, knowing, imagining, etc. create *hyperintensional contexts* which call for a semantical analysis that is sensitive to more fine-grained individuation of sentences [18, 24, 22]. Within these contexts, the informational content of a sentence is understood to have epistemically relevant dimensions beyond its intention, such as what the sentence is about, or which issues it embeds. Intensional semantics on the other hand individuate sentences up to their truth values, and are not able to account for the hyperintensionality suggested in these contexts. Stalnaker points out that this is mainly due to imposing the logical structure of possible worlds on the semantical analysis of sentences [140]. Essentially, propositions (i.e. intensions) are determined solely based on a set of possible worlds, with a fixed domain of issues. Each sentence is then assumed to be embedding the totality of this domain, as they divide the whole set of possible worlds into two camps, namely *true* and *false*. In other worlds, since possible worlds are internally complete with regards to a given domain, they determine the truth value of each sentence in the language and for each issue in their domain, hence a sentence is *about* everything in that domain. Given a fixed domain of issues, even contingent sentences with the same intension cannot be distinguished further.[2]

While the term *hyperintensional* is coined by Cresswell [36], discussions of hyperintensionality and hyperintensional content go back to [32, 96, 97]. Following Cresswell, hyperintensional contexts are simply defined as contexts where logical equivalences are not respected. Earlier attempts at analysing sentences in hyperintensional context including [32, 96, 97, 36] state that intensions are not proper designations for the meaning of sentences and intensional identity does not account for the identity of meaning. Hyperintensional equivalence is sometimes referred to as synonymy, and it requires a stronger meaning relation between sentences than intensional identity. They try to explain away the problems of intensional identity by analysing the structures of sentences and the meaning of their parts. While [32, 96] introduce a notion of *intensional structure* (although the term is only coined later by Carnap), [97, 36] aim at compositionally building the meaning of a sentence on the intensions of its parts.[3] In the following, only more recent attempts are considered, which try to explicitly capture the epistemically relevant aspects of the content of sentences, enhancing their intensional meaning with *topics* or *subject-matters* embedded in a sentence, that is sometimes called the *aboutness* component [171]. Consequently, they enhance the truth-preservational definition of logical entailment with aboutness-preservation. These

[2]While Stalnaker identifies the problem of logical equivalents and deduction in this way, he also presents a defense of the possible worlds analysis for the representation of intentional mental states in terms of a *causal - pragmatic picture* [140]. Intentional mental states are understood essentially in terms of their pragmatic role as they are used to determine actions. Hence what is necessary is distinguishing between possible alternative outcomes of the agent's actions as possible states of the world. This is exactly what is achieved by the possible worlds analysis. In this pragmatic analysis of propositions, the relevant division is into possibilities where a proposition is true and where it is false. Representing possible alternatives in this way is essential to the role of beliefs and desires, however, the particular way of representation is not. See Chapter 1 in [140].

[3]Carnap states that if two sentences are not only L-equivalent (intensional identity) in the whole but consist of L-equivalent parts, and both are built up out of these parts in the same way, then they have the same intensional structure (i.e., they are intensionally isomorphic) [32, p.59].

are usually frameworks based on possible worlds semantics. These will then be compared to the *state-based* approach presented in Chapter 2. The latter achieves to account for hyperintensional sensitivity by redefining what determines a proposition, and allowing in return more fine-grained individuation of meaning.

Section 3.2 provides a limited classification of hyperintensional approaches to belief revision. These are in particular approaches based on possible worlds semantics which aim to enhance the intensional meaning and truth-preservation-based classical logical consequence relation with an aboutness component, and approaches based on state-semantics which redefine what determines a proposition and replace truth-preservation with support-preservation. The former approach is exemplified below with frameworks based on *world-partitions* [171, 99] and that are based on *topic-sensitivity* [19, 175]. The latter approach is primarily exemplified with the models presented in Chapter 2, after the foundations for state-based semantics is laid by introducing the *situation semantics* of Barwise and Perry [15, 120]. Section 3.3 then proposes a comparative assessment of the introduced approaches based on the issues of topic-transitivity, topic inclusion, and the notion of semantical opposition in relation to the belief revision context.[4]

3.2 Approaches to hyperintensional belief revision

The classification of hyperintensional approaches presented here is similar to the analysis concerning different ways of representing subject-matters as part of the meaning of sentences by Hawke [69]. The first category below corresponds to Hawke's *way-based approaches*. These are reconstructions subject-matters as world-partitions, which are "different ways for a subject to be", following Lewis' definition [99]. The next category includes topic-sensitive frameworks which are classified by Hawke as an *atom-based approach*. These are theories in which the subject-matter of a sentence is identified with a set of the topics based on the atomic claims in the sentence [69, p.698]. Finally, the state-based approaches considered below include the situation semantics of Perry and Barwise [15, 120], which is classified by Hawke within the *subject-predicate approach*. However, the rest of the state-based approaches considered here do not appear in Hawke's analysis since they do not represent subject-matter per se. Another work that is partly concerned with a categorisation of hyperintensional semantics is due to Sedlár [136]. He proposes a categorisation in regard of state-based, syntactic, and structuralist approaches. All frameworks presented below, except for the topic-sensitive framework, can be classified as state-based approaches since they all deviate from the idea that sentences are evaluated solely on possible worlds. Accordingly, world-partitions can be seen as course-grained states rather than possible worlds, and while topic-sensitive frameworks are not mentioned in [136], it might be possible to

[4] Some notable works on hyperintensionality include [77, 99, 140, 171]. The subject has also been studied as theories of topics [24, 69]. The application of hyperintensional logics on knowledge and belief seems to have a more recent and limited literature (compared to the use of intensional semantics in these contexts). Examples include a series of works starting with [19, 29, 175].

reconstruct the idea via Sedlár's hyperintensional models based on content-assignment.[5]

3.2.1 Varying domains of issues

The possible worlds analysis has been commonly used to represent informational and epistemic content. The level of idealisation imposed by this analysis, however, has attracted attention in the literature, and sometimes included under the title of the problem of logical omniscience. The problem is partly due to the result of deductive closure conditions imposed on the representation of epistemic states of agents. Some problematic consequences of this application are as follows. (Suppose ϕ and ψ express sentences in our preferred language.)

1. Closure under logical entailment: if the agent comes to believe that ϕ, and ψ is logically (or intensionally) equivalent to ϕ, then she also believes that ψ.

2. Belief in logical / necessary truths: if ϕ expresses a necessarily true proposition, then the agent necessarily believes that ψ.

3. Consistency and triviality: if an agent comes to believe a necessary falsehood (such as a contradiction), she then believes every sentence in the language, making her belief state trivial.

The problem of deductive closure, at least to some degree, can be analysed in terms of hyperintensional sensitivity and more fine-grained individuation and implication conditions between sentences. Stalnaker proposes that the problem partly arises from the faulty assumption that the possible worlds analysis of intentional mental states is committed to the existence of a unique total set of possible worlds [140]. He points out that possible worlds distinguish between different alternative possibilities, still different sets of worlds can achieve this in varying degrees. The analysis then allows different *domains of issues* to be determined and embedded in the sentences. In particular, a given set of possible worlds might be smaller than another, allowing more fine-grained differentiation of possibilities in the latter. This amounts to saying that the larger set of worlds indeed includes a larger domain of issues. It is moreover possible that different sets of possible worlds include different domains of issues without being subsumed by one another. In terms of intensional equivalents, this means that sentences which determine the same proposition given a domain of possibilities, may determine different propositions given a different domain of possibilities.

This discussion is related to the distinction between the notions of *completeness* and *comprehensiveness* as they apply to possible worlds. Once a domain of issues is set, possible worlds are complete with respect to this domain. They are however, not necessarily comprehensive relative to an external interpretation, which might include issues that fall outside of their domain [119]. A reoccurring example of the application of varying domains

[5] I thank personal interaction with the author for the latter point.

of issues to attitude modalities is due to Stalnaker [140, p. 63]. A more compact version is presented by Perry, as quoted below. Perry uses the terms *total ways* and *partial ways* to refer to the possible worlds and the *situations* (i.e., partial worlds) respectively.

> [S]uppose we are interested in the beliefs of a dog and its master. [T]hey both believe that a bone is buried in the back yard. But the master has the concept of an ersatz bone and the dog does not have this concept. In representing the master's beliefs, we would want to include, among the possibilities his belief might rule in or out, or leave open, the bone in question being ersatz. But we would not want to do this in the case of the dog, or at least might not want to. In the first case, we should take the set of total ways to separate the two cases; in the latter we should not. Given these different sets of total ways, the propositions believed will differ also, since propositions are functions from sets of total ways to truth values. ([119, p.85])

The application of varying domains of issues, although pinpoints the problem successfully, is far from being satisfactory. Both Stalnaker and Hintikka mention that while the possible worlds analysis can be improved to offer a solution to the problem of deductive closure in reasoning, the work has not been done. The problem persists since once a domain of issues is fixed, the framework cannot distinguish between the intensionally equivalent sentences within that domain. In order to make such distinctions possible, Hintikka and Perry separately suggest supplementing the possible worlds analysis with partial worlds. Hintikka suggests that possible worlds should instead be interpreted as *small worlds* with a limited domain of issues, where some worlds are smaller than the others. That is, possible worlds could be models of situations rather than entire universes, and there is nothing in the analysis that prevents such an addition [72, p.161]. Perry on the other hand suggests, that while possible worlds are usually seen as total ways that provide an answer to every issue under consideration (as total functions from sets of issues to answers), the analysis could be supplemented to involve partial ways that instead provide answers to a limited range of issues. Partial ways are then not necessarily disjoint alternatives for how things might be, rather they can be related to one another and to total ways in various forms. In this way, a possible world in the traditional sense can make any number of (partial) ways true, while making only one total way true [119, p.87].

To sum up, once a domain of issues is fixed, even when this domain does not include all logical possibilities, the problem of logical equivalence persists. The semantical analysis of meaning needs to be supplemented further, to be able to account for shifts in subject-matters. On the other hand, the discussion on partial worlds already motivates the development of situation semantics, which is presented below.

3.2.2 Reconstructing subject-matters as world-partitions

Introduced by Lewis in [99], the theory of world-partitions aims to capture hyperintensional sensitivity by reconstructing subject-matters as part of the meaning of the sentences.

World-partitions are defined as "ways of structuring the logical space" by grouping possible worlds to obtain more course-grained worlds. Each way of doing so is said to generate a *partition*, which determines or spells out the relevant issues. Lewis then identifies partitions with *subject-matters* embedded in the meaning of sentences. Reasoning based on these partitions allows agents to ignore the issues or aspects of the world that are not relevant to a given reasoning task.

Our running example above can be formulated using Leitgeb's formalisation in [91, p. 310] of Lewis' world partitions. Let p stand for the fact that "a bone is buried in the backyard", and q for the fact that "the bone in question is an ersatz bone". Consider a set of possible worlds $W = \{w_1, w_2, w_3, w_4\}$ and π_1 and π_2 partitions of Wsuch that

$w_1 = \{p, q\}, w_2 = \{p, \neg q\}$,
$w_3 = \{\neg p, q\}, w_4 = \{\neg p, \neg q\}$, and
$\pi_1 = \{\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}\}$,
$\pi_2 = \{\{w_1, w_2\}, \{w_3, w_4\}\}$.

Given a world-partition semantics, two worlds belong to the same *partition cell* iff they agree on all relevant truth assignments. The partition π_2, for instance, differentiates the worlds in W up to the value of p, however, it does not determine the truth value of q since the worlds in each partition cell of π_2 do not agree on the their assignment of the value of q. We can then state that the subject-matter of the sentence p is *about* the given partition, and the value of q is undetermined. The partition π_2 is then able to represent the beliefs of Fido, since he does not consider whether or not the bone in question is an ersatz bone. The partition π_1, on the other hand, is able to represent the beliefs of Owner. The idea is then, the differences between the worlds which belong to the same partition cell are ignored, and the agents reason with fever possibilities based on the relevant aspects or issues.

In particular, the satisfaction of formulas are two-fold: let w be a possible world in W, π a partition in $\Pi : W \to W^2$, and ϕ a sentence in our language, then $w, \pi \models \phi$ iff

(i) $w \models \phi$, and

(ii) the partition cells of π distinguish the worlds precisely up to the truth values of the propositional letters in ϕ (the *least subject-matter* of ϕ) or in some more fine-grained manner.

The above formalisation can be expanded to include a belief modality B_i, based on an accessibility relation R that picks out the doxastically possible worlds for an agent i:
$w, \pi \models B_i\phi$ iff

(i) for all w' s.t wR_iw', $w' \models \phi$, and

(ii) the partition cells of π distinguish the worlds precisely up to the truth values of the propositional letters in ϕ (the *least subject-matter* of ϕ) or in some more fine-grained manner.

Let the relevant accessibility relation for both Fido and Owner be

$$R = \{\langle w_1, w_1\rangle, \langle w_1, w_2\rangle, \langle w_2, w_2\rangle, \langle w_3, w_3\rangle, \langle w_4, w_4\rangle\}$$

Thus, $w_1, \pi_2 \models B_{Fido}p$, that is at w_1, Fido believes that there is bone buried in the backyard. However, $w_1, \pi_2 \not\models B_{Fido}(q \vee \neg q)$, since Fido does not consider whether the bone in question is an ersatz bone. Whereas the Owner believes that there is a bone buried in the backyard and that it is either a real bone or an ersatz bone: $w_1, \pi_1 \vDash B_{Owner}(p \wedge (q \vee \neg q))$.

The theory of world-partitions enhances the idea of varying domains of issues in an important way, with a *mereological structure* that applies to subject-matters. The general idea is, as the relevant issues change, the partitions change, while the given set of possible worlds is preserved. In other words, the cells in a partition can be further divided, or can be fused together to generate different partitions, based on the same set of possible worlds. As a result, sentences that are intentionally equivalent in one partition can be distinguished in another. Consider again the sentences represented by p and $p \wedge (q \vee \neg q)$ above. Given partition π_1, they express the same intension since for all $w \in W$, $w, \pi_1 \vDash p$ iff $w, \pi_1 \vDash p \wedge (q \vee \neg q)$. Given partition π_2 however, $(q \vee \neg q)$ is not satisfied anywhere, while $w_1, \pi_2 \vDash p$. The truth assignments then depend on the specific partition in question. In this way, the framework allows one to represent different degrees of *content resolution* based on different ways of partitioning one and the same set of possible worlds [91, p.7].

Lewis uses equivalence relations for partitioning the logical space. Consequently, the worlds in a partition cell are exactly alike, with respect to the relevant subject-matters. An equivalence class then specifies how things might be with respect to a subject-matter completely [99, p.162]. He further introduces the notion of *refinement* of a subject-matter. Modelled as partitions, a subject-matter M includes another subject-matter N iff each cell of N is a union of the cells in M. In this case, the larger subject-matter M refines the smaller subject-matter N, by dividing the logical space in a more fine-grained way. Overlapping subject-matters are defined as having a subject-matter as a common part.

Aboutness

The issue of subject-matters and the notion of aboutness has been further analysed by Yablo in [171]. He, however, uses similarity relations rather than equivalence relations to structure the logical space. This allows partitions with non-disjoint cells, where a sentence can be true in more than just one way. The subject-matter of a sentence S is defined as what distinguishes ways of S being true (false). Two worlds then belong to the same cell in a partition when S is true (false) *in the same way*.[6]

Yablo furthermore establishes a robust relation between aboutness and *truth-making*. Commonly known in the literature due to Fine's works starting with [50], truth-makers cover more than truth-conditions of sentences. In particular, they account for what a sentence is about, where aboutness is possibly independent of the truth-conditions [171, p.57]. Fine refers to the truth-conditional content of a sentence, that is given by its

[6]In particular, a Lewisian least subject matter (the lower-bound) for S would be "whether S is true". This means different ways of being true for a sentence are incompatible with each other. Yablo proposes rather "how S is true" as the least subject-matter, that is, the upper bound for the subject-matter of S. If S's ways of being true do not change between two worlds (i.e., they belong to the same partition cell) then a dissimilarity between them is not part of the subject-matter of S. The distinction between the two ways of structuring the logical space is similar to the distinction between the recursive and reductive truth-makers.

intention, as the *thin content* of the sentence, whereas the term *thick content* refers to the meaning of a sentence supplemented with its subject-matter, determined by the pair of its truth-makers and false-makers [51, p.134]. These refer to the ways the sentence can be true and false respectively (in the following, I only refer to truth-makers, however, one can always pair this with the false-makers). The thick content of a sentence is evaluated on a logical space that consists of a family of non-empty sets of worlds, i.e., *divisions* [171]. Divisions are similar to world-partitions, only that they are obtained via similarity relations instead of equivalence relations. A truth-maker semantics can function as an intentional semantics or as a hyperintensional semantics. On the one hand, there are *reductive truth(false)-makers*, which are minimal models (counter-models) of formulas. In this case, the subject-matter of a sentence depends on its truth-conditions, hence classically intentionally equivalent (propositional) sentences are equivalent also in terms of their thick content. On the other hand, the theory of *recursive truth-makers* allows one to make more fine-grained, *non-minimal* individuation of sentences.

Consider the following two formulas: $p \vee (p \wedge q)$ and $p \wedge (q \vee \neg q)$. The minimal models that make both formulas true are models that make p true. The two then have the same reductive truth-makers, and are treated as equivalent. Reductive truth-making takes into consideration only what is required to make a sentence true, and does not account for a stronger distinction based on ways of being true. It can be argued that this minimal content suffices for certain purposes, for instance in a classical propositional setting. Recursive truth making on the other hand, posits a truth-maker for each disjunct. Recursive truth-makers for the above formulas are then $\{p\}, \{p, q\}$ and $\{p, q\}, \{p, \neg q\}$, respectively. This shows that there can be more than one way for a disjunctive sentence to be true. This is possible as truth-makers are non-disjoint alternatives since subject-matter of a sentence is defined as the upper-bound for ways of a sentence to be true.

We will conclude this section with a note on hyperintensionality and entailment relations in hyperintensional contexts. Classical logical entailment is based solely on truth-preservation. In an intensional context, a sentence ϕ entails another sentence ψ iff whenever ϕ is true, ψ is also true. Using possible worlds semantics, this can be represented with an inclusion relation between two propositions: $[\![\phi]\!] \subseteq [\![\psi]\!]$, where $[\![\psi]\!]$ is the set of worlds that makes ψ true. Let us call this *the minimal form of entailment*. In some contexts however, such as the epistemic context, one needs to opt for an entailment relation that accounts for a non-minimal, more fine-grained content preservation. In order to realise this, hyperintensional accounts propose to strengthen the classical entailment relation with an *aboutness-preservation*. To this end, Yablo defines *inclusive entailment*: ϕ (inclusive) entails ψ iff (i) each of ϕ's truth-makers contain as a subset a truth-maker for ψ, and (ii) each of ψ's truth-makers are contained in a truth-maker for ϕ. Consequently, when defined via recursive truth-makers, a sentence might not inclusively entail some of its classical consequences.[7]

[7]Similar consequence relations are studied in the literature. In particular, they relate to what is called *relevant entailment*, and have a broader place in the literature on logical entailment and implication. The discussion is sometimes included also in the subject of *fallacies of relevance*. Early works in the area include [7, 16, 99]. The first part of Yablo's definition of inclusive entailment corresponds to *tautological*

Consider the following example from Yablo [171, p.117-120]. The fact "Z is zebra" (ϕ) truth-conditionally implies that "Z is not a cleverly disguised mule" (ψ). It cannot be said however, that an agent who believes that "Z is a zebra" necessarily believes also that "Z not a cleverly disguised mule", solely based on what they already believe. This is because, the subject-matter of ψ is indeed not included in the subject-matter of ϕ. While ϕ is about zebras, ψ is about everything that is not a clearly disguised mule. For instance, ways of ψ being true includes Z being a lion, whereas this is not true for ϕ. Aboutness-preservation then fails between the two sentences, and the content of ψ is a mere (truth-conditional) consequence of ϕ, which do not carry over as a logical entailment of the content of ϕ in an epistemic context.[8]

Logical closure in terms of inclusive entailment, namely *topical closure*, then diverges from classical logical closure. Yablo notes that in the context of knowledge (and belief) attributions, over and above truth-conditions, the subject-matter, i.e.,"true how" matters. The closure principle for knowledge ascriptions should hence state: "If [an agent] S knows that ϕ, and ψ is part of ϕ, then S knows that ψ" [171, p.117], rather than the classical closure requirement (i.e., *equivalence*): "If S knows that ϕ, and ϕ and ψ are logically equivalent, then S knows that ψ". To sum up, in virtue of aboutness-preservation, failure of intentional equivalence comes about as a truth-value shift between truth-conditionally equivalent sentences, due to a shift in aboutness [171, p.125].

Partitions and Questions

The theories of world-partitions and aboutness suggest a way of reasoning that is relativized to partitions. Partition-dependent reasoning is in fact a form of context-dependent reasoning, where the choice of partition constitutes part of the epistemic context that the agent is in (i.e., *the reasoning context*), which in turn co-determine the content of belief sentences. Leitgeb highlights that, the way that the logical space is partitioned does not affect the logical content of sentences, nor does it mean that the agents *ignore* logical possibilities, rather, only *resolution* of the content of beliefs change as the context shifts [90, p.148]:

> I should add that if the agent asserted one sentence after the other in order to express her beliefs in $A_1, A_2, A_3, \ldots$, then it would still not be the case that any of these sentences would have to express one proposition in one context and a different proposition in another: rather, which proposition is entertainable by the agent may shift from one context to the next.[90, p.145]

entailment in [7].

[8] Lewis also presents an enhancement of classical entailment along similar lines. He defines that, a sentence is about a subject-matter iff the truth value of the sentence *supervenes* on that subject-matter. That is, worlds in each partition cell agree on the truth value of the sentence, and they either imply or contradict it. It follows from this definition that whenever a sentence ϕ implies another sentence ψ it follows that ϕ and ψ are relevant, i.e., they share a subject-matter. Another noteworthy consequence is that the necessary truths are about all subject-matters, meaning that they are implied by all sentences [99, p.169].

Partitions of logical space have also been identified with *questions* by various authors. Levi argues, for instance, that inferences are made within a context that is determined by the underlying question [95]. Leitgeb states hyperintensionality can be understood based on the thought that "hyperintensions are nothing but intensions under a question"[91, p.319]. Elsewhere he states, that rational belief seems sensitive to the agent's underlying questions, when formulated in terms of a semantics of questions, where questions are reconstructed as world-partitions (see below) [90, p.140]. In [172] Yalcin introduces partitions induced by different questions as different resolutions of the logical space, in an attempt to model beliefs as question-sensitive. His framework addresses various concerns related to the problem of logical omniscience, both in terms of hyperintensionality of beliefs and of closure conditions for belief sets. The main motivation behind *inquisitive semantics* is, in fact, the idea that questions can be understood as generating a (reasoning) context via world-partitions. Similar to the above theories, in inquisitive semantics, a proposition determines a context (a set of worlds) with the information it conveys and induces a partition on this context with the issues it raises. In this way, the informative content and the inquisitive content of sentences, i.e., the information conveyed and the issues raised by those sentences, in a uniform account[9].

Based on the partition theory of questions [61] (which pre-dates the introduction of inquisitive semantics) dynamic frameworks of world-partitions have been introduced in the literature [59, 75, 76, 3, 2, 4, 5]. These models do not deal with the dynamics of intentional mental states, but rather with the dynamics of partitions themselves. Based on [75], A context C is modelled as an equivalence relation on a set of worlds $S \subseteq W$. Thus, C induces a partition on S, to create a *discourse context.* S is the domain of C, which consists of all worlds that are compatible with the information so far established in the discourse. The relation C then encodes the issues raised so far by relating two worlds w and v only if they do not differ under the present discourse (i.e., the issues raised so far). Here, individual sentences are identified with their potential to change the discourse context. More specifically, declarative sentences restrict the context S when they are uttered, and questions disconnect the worlds which are initially related by C. Thus, while assertions induce information-directed change, the questions induce issue-directed change. In [60] the authors suggest that using a similarity relation in place of an equivalence relation such as

[9] In particular, inquisitive logic is a conservative expansion of classical logic. A support-based-semantics is introduced on top of the classical possible worlds semantics. The primitives of the support relation are sets of possible words $s \subseteq W$ interpreted as information states. Propositions are defined as downward-closed non-empty sets of information states. The support relation then amounts to set membership between information states and propositions ($s \models P$ iff $s \in P$). Entailment between propositions is then defined as preservation of support (i.e., set-inclusion between propositions). Accordingly, an information state s supports a proposition P iff s settles all the issues raised by P. Truth at a possible world can be defined as follows: $w \models P$ iff $w \in s$ for some $s \in P$. Conversely, $s \models P$ now means for all $w \in s$, $w \models P$. The union of all the information states that support a proposition P is the informative or truth-conditional content of P. For non-inquisitive sentences, the informative content and the inquisitive content coincides [34]. The distinction between the background classical logic of non-inquisitive propositions and the hyperintensional logic of inquisitive propositions is similar to the treatment of intensional and hyperintensional context in topic-sensitive frameworks below.

C, one can obtain a better analysis of different forms of questions, hence different ways of making issue-directed changes in a discourse context. Besides providing formal frameworks for the world-partition systems, these highlight the importance of *issue-directed changes in discourse*, which is similar to a fine-grained analysis of belief change, which accounts both for the informational content and the issue-related-content of sentences in epistemic contexts [39].

3.2.3 Topic-sensitivity

Topic-sensitive frameworks refine possible worlds semantics for belief with sets of topics, and enhance truth-conditional entailment with a *topicality filter*. In particular, *topic-sensitive intensional models* include a set of topics that are compositionally assigned to sentences by an assignment function, and the topic-sensitive entailment requires *topic-containment* between propositions besides classical truth-preservation. Applications to belief revision, knowability and imagining, have been proposed in various works [19, 175, 22, 18, 24, 20]. In the following, I refer to [19] for simple (static) hyperintensional belief revision and to [175] for dynamic and topic-sensitive belief revision which includes a dynamics of topics.

In general, *topicality* is understood to impose an aboutness filter on the meaning of sentences [24]. Formally, topics are represented as members of an abstract set, and they are assigned to atomic sentences (i.e., to atomic parts of the sentences) of a language by a topic-assignment function. They can be defined as states of affairs, issues, circumstances or situations that sentences are about. The basic framework for topic-sensitive hyperintensional belief revision in [19] assumes only two restrictions on the nature of topics. First, they comply with certain mereological conditions, and second, the topic of complex sentences are made up compositionally of the topics of their parts. Within a topic-sensitive context, a sentence ϕ logically entails a sentence ψ iff (i) ϕ truth-conditionally entails ψ, and (ii) the topic of ψ is fully included in the topic of ϕ. The latter part of the definition accounts for a *content-containment requirement*, similar to the semantics introduced in [49]. While the two-fold entailment requirement applies only to the topic-sensitive context, which is generated by an hyperintensional modality such as belief, entailment for propositional formulas is classical. Berto explains hyperintensionality of belief revision as a special case of *framing effects*. That is, intensionally equivalent content of beliefs trigger different revision recipes, depending on how they are presented. Consequently, certain logical closure principles may fail in the context of belief revision, as well in other topic sensitive contexts [19].

The models in [175] can be seen as a generalisation of the simple hyperintensional belief revision introduced in [19]. Both extend the classical Kripke models for belief with a set of topics, a topic assignment function, and a topic fusion function, while the former include also a designated topic for "the topic of the agent's belief state representing the totality of the subject-matter the agent has grasped already". Dynamic hyperintensional belief revision is then defined by expanding these models with a plausibility ordering of worlds and a topic-sensitive lexicographic upgrade operator. For this, the traditional dynamic of lexicographic upgrade (as in [155]) is paired with a dynamics of topics. Belief revision then has two aspects, revising a belief set w.r.t. a new belief ϕ means, (i) the agent has received the information ϕ and (ii) has come to grasp the topic of ϕ. Accordingly, a belief revision formula $[\Uparrow \phi]$ transforms a model M across two components: (i) all ϕ-worlds become more plausible than all $\neg\phi$-worlds with the order within these sets preserved, and (ii) the new designated topic for the agent's total belief state now represents that the agent has grasped the topic of ϕ, adding it to the initial designated topic [20, 145]. Topic-sensitive dynamic belief revision is hyperintensional. Specifically, the replacement rule RE (from $\phi \leftrightarrow \chi$ infer $[\Uparrow \phi]\psi \leftrightarrow [\Uparrow \chi]\psi$) fails when ψ is a doxastic sentence, while a topic-sensitive version of this rule is shown to be valid [175, p.783].

A static conditional belief modality is also introduced based on the two-fold entailment relation and the designated topic for a belief state as described above. Accordingly, a fact ψ is believed by an agent (we omit the agent subscripts) conditional on a fact ϕ (denoted with the formula $B^{\phi}\psi$) at a world w in a model M iff (i) the most plausible ϕ-worlds are ψ-worlds, and (ii) the topic of ψ is in the *fusion* of the topic of ϕ with the designated topic. Plain belief can then be represented by replacing the condition with the truth-constant which has as a fixed topic the whole subject matter the agent is on top of [20].

3.2.4 State-based hyperintensional semantics

State-based semantics include *states* or *situations* rather than possible worlds as designations of sentences and hence as the primary units for representation of their meaning. The content of the sentences are identified with sets of states, which now determine the propositions. An important divergence from the possible worlds-based hyperintensional semantics is that classical logic is not preserved as the background or external logic, however, it can be recovered as special cases.

Situation semantics was first introduced by Barwise and Perry [15, 120] as semantic tools for studying *logic of attitudes*. Situations are defined as configurations of objects and relations, which can be understood as occurring as parts of total worlds.[10] In virtue of being partial, non-disjoint entities and their mereological features, situations are natural ways of modeling information states, and they adequately capture hyperintensional sensitivity. In particular, there can be larger (richer) and smaller (poorer) situations, which can be included in one another or fused to generate new situations. In state-based semantics, the existence of inconsistent states are usually allowed, although not in Barwise and Perry's framework, allowing different ways of being inconsistent.

Looking back at our running example, the designation of Fido's belief sentence is a

[10]Situations are sometimes referred to as *partial worlds* or *small worlds* [72, 119, 141]. They are represented in the literature either as sets of possible worlds, or as primitive entities. The form of representation, however, generates some differences as to how situations function. For instance, in the latter form, situations are consistent and they support (make true) all logical truths that are represented within the relevant domain of issues, while in the latter form, this is not necessarily so. On the other direction, possible worlds can be understood as special cases of situations, where in general situations are partial functions from objects and relations to truth values, and possible worlds are total functions.

situation where the configuration `(buried-in-the-backyard,Bone)` is true. This can be denoted with a situation s_0 with 's_0`(buried-in-the-backyard,Bone)` $= 1$'. Whereas the configuration `(real-or-ersatz,Bone)` is not true of s_0 since neither s_0`(real,Bone)` $= 1$ nor s_0`(ersatz,Bone)` $= 1$ are realised in s_0. When the units of representation are shifted from possible worlds to situations, different subject-matters or different topics that contribute to the meaning of classically logically equivalent sentences may determine different sets of situations. Barwise and Perry state that from the point of view of situation semantics, logical equivalence should mean *being true in the same situations*, rather than in the same (possible) worlds. This is how they solve the problem of logical equivalence, which occurs within the context of logic of attitudes that is generated by taking as designations of sentences their intensions, i.e., the sets of possible worlds that make them true [15, p.676].

Situation semantics is a theory of natural language where situations are informative entities. It can, however, be generalised to obtain other state-based semantics where the units of representation are abstract, mathematical entities, which allow various interpretations for the models. For instance, while in the context of truth-maker semantics, states are fact-like entities which *verify* or *falsify* sentences, and also make up the worlds [50], within an informational semantics such as the HYPE-semantics, they can be interpreted as information states, simply representing collections of information.

Truth-maker semantics

Fine's truth-maker semantics also exemplifies state-based hyperintensional semantics. Applications of it to hyperintensional belief revision are developed in [70, 82], while an application of *exact truth-making* for belief ascriptions is provided by Jago [78]. The truth-maker state-space is a partially ordered set of states. Propositions are identified with a set of *verifying* and *falsifying* states, and two distinct propositions can have common verifiers while having distinct falsifiers. Hence, verification and falsification can be interpreted as ways of a sentence being true and false. Two different levels of truth-making can be defined based on truth-maker semantics. *Exact truth-making* requires that the verifying (falsifying) states are *wholly relevant* to the truth (falsity) of a sentence.[11] *Inexact turth-making* on the other hand, can be defined in relation to exact truth-making, as a looser form of truth-making with a monotonicity condition. Simply put, each exact verifier of a sentence S is an inexact verifier of S, and if a state s is an inexact verifier of S then for all states s' that contain s, s' is also an inexact verifier of S. Inexact verifiers are then *partially relevant* to a sentence.[12] Consider the two classically logically equivalent formulas q and $q \vee (q \wedge r)$. The

[11] This is similar to the notion of tautological entailment by van Fraassen [163]. Fine and Jago state that [163] was the first paper to present *an exact semantics for classical logic* [53]. Jago [78] proposes a belief ascription system based on exact truth-making and entailment between propositions based on *conjunctive parthood*, similar to Yablo's inclusive entailment. He notes that the transitive closure of the rules and axioms of this system suggests a deductive relation between sentences that is identical to Angell's *analytic containment* in [9] [78, p.9].

[12] It is argued later by Leitgeb that exact truth-making can indeed be defined as minimal inexact truth-making, by supplementing Fine's models with non-persistent *totality facts*, and *totality fact operators* indexed to each state [92]. Another attempt to define inexact truth-making as the primary form of truth-

exact truth-makers of the former are the $\{q\}$-states, while the exact truth-makers of the latter are the $\{q\}$-states and the $\{q \cup r\}$-states. The q-states are also inexact truth-makers of the formula $q \vee (q \wedge r)$. Hyperintensionality of inexact truth-making can be exemplified by the following: classically logically equivalent formulas $p \vee \neg p$ and $q \vee \neg q$ has distinct verifiers and falsifiers.

A brief comparison of Fine and Yablo's truth-maker semantics is in order. The main difference is that Fine's truth (false)-makers as ways of being true (false) are objects of their own, rather than structured sets of possible worlds. Fine comments on this by saying that while Yablo's notion of thick content is an enhancement of the intensional content, his is a modification [52, p.134]. This generates notable differences between the two. To begin with, truth-makers as divisions of sets of possible worlds allow the existence of a unique impossible state. This means, the subject-matter of all contradictory sentences are identical, and included in the subject-matter of all sentences (just as the subject-matter of necessary truths include all subject-matters that can be represented in a given model). Truth-makers as partial states allow for a diversity of impossible states, which can be distinguished in terms of the possible states contained in them. In this way, the subject-matter of an inconsistent sentence is not necessarily included in the content of every other sentence.[13] Moreover, using Fine's truth-makers, the mereology of subject-matters can be analysed based on the mereology of states: the subject-matter of a sentence ϕ is the union of all its verifiers. Hence, two sentences share a subject-matter iff they have the same closure of verifiers under a fusion or an inclusion function. Whereas, given Yablo's theory, two sentences are subject-matter-identical iff they have exactly the same verifiers.[14]

A more recent contribution to the literature on state-based hyperintensional semantics is Leitgeb's HYPE-semantics [91], which is the foundation for the hyperintensional belief revision models in Chapter 2. The semantics is built on a space of possibly incomplete (gappy) and inconsistent (glutty) states. A natural interpretation of this semantics is to take the states as collections of information.[15] A HYPE-state-space is characterised with a partial fusion function, which can be interpreted as representing a monotonic information flow between states. HYPE-logical equivalence is defined as support in the same set of states.[16]

making is by Deigan [40].

[13] Distinguishing impossible states or inconsistent pieces of information has been a motivating problem for the literature on hyperintensionality. In particular [23, 77] introduce hyperintensional semantics which enhance possible worlds spaces with impossible worlds, towards a solution to this problem.

[14] Recall that, alternatively the subject-matter of a sentence is defined as the pair of all its verifiers and falsifiers. All definitions presented here could then be adjusted to include this pair instead of only the verifiers.

[15] Leitgeb notes that states can also be interpreted metaphysically as "chunks of reality", and he leaves the interpretation as an open choice.

[16] Leitgeb leaves open the nature of the support relation between the states and formulas. The alternatives include,

- ϕ is satisfied in s, and ϕ's "inexact" subject matter is s,
- s is an "inexact" truthmaker of ϕ,

3.3 Discussion

The two camps of hyperintensional approaches to belief revision to be assessed here are differentiated based on the primary entities of representation employed to specify what determines propositions and hence the meaning of sentences. Based on possible worlds representations, the content of sentences are identified with their classical intension enhanced with the information about what the sentences are about, determined by their topics. This generates a two-fold analysis of meaning and suggests a two-fold entailment relation between sentences. Classical logic is preserved as the external logic, and propositions behave classically outside of the hyperintensional (i.e., topic-sensitive) context. These models validate classical propositional tautologies, however, they do not transfer into the domain of the topic-sensitive modalities. The topics of sentences then do not affect their classical truth-conditions. In other words, and it is important to highlight that, the topics or subject-matter of sentences are *independent* of their truth conditions.

The state-based frameworks identify the meaning of sentences with the proposition they express, however, as they are determined by sets of partial states. The focus of valuation then shifts from classical truth-conditions to *support conditions* on states. While this suggests a one-layered entailment relation between sentences, the analysis of meaning is more fine-grained than classical, as states determine different ways of a sentence being true. [17] Hence, the hyperintensional component of meaning is not separated from truth-conditions of sentences.[18] The differences between the two approaches assessed below mainly stem from whether topics are independent of the truth-conditions.

3.3.1 Topic-transitivity

The role of topics or subject-matters in a formal semantics are to help determine which inferences are warranted in an aboutness-sensitive context. There is a broad literature on the notion of topics and subject-matter, which focuses in part on the criteria for the behaviour of subject-matter in connection to logical connectives [52, 69]. The theory of subject-matters suggests that what a sentence is about is independent from its truth-conditions. Both world-partition frameworks and topic-sensitive semantics reflect this fact. In particular, subject-matters are determined based on how the logical space is partitioned (or divided), and topics are assigned to the sentences compositionally. This brings about *the transitivity of subject-matters and topics* under propositional logical connectives. That

- s is a way of ϕ being true,
- s is semantically committed to the truth of ϕ [91, p.311].

[17]The distinction we make between one-layered and two-fold entailment relations here should not be confused with only verification based vs verification and falsification based entailment relations. There are various one-layered support-based semantics that come with a dual entailment relation for verification and falsification of sentences. Fine's truth-maker semantics is an example of this.

[18]Fine, however, suggests that we can reconstruct subject-matter as the pair of verifiers and falsifiers of a sentence, while the truth conditions are given by the set of verifiers alone.

is, there is no shift in aboutness as a result of being embedded under these connectives. For instance, a sentence ϕ is about what its negation, $\neg\phi$, is about. Consequently, subject-matters and topics are *negation neutral* [52, 69, 24], in other words, negation is a *topic-transitive* connective. This is furthermore accepted as a requirement for a good recursive account of aboutness [24, p.6].[19]

The principle of compositionality for determining the topic of complex sentences (i.e., *congruence* [52] also requires that the topic of a conjunction $(\phi \wedge \psi)$ is identical to the topic of a disjunction $(\phi \vee \psi)$. In both cases, the topic of the complex sentence is the fusion of the topic of its parts: $c(\phi) \oplus c(\psi)$ (the requirement is called *junction* [52]). Hawke on the other hand argues that the topic of a disjunction $(\phi \vee \psi)$ should merely be included in the topic of a conjunction $(\phi \wedge \psi)$ [69]. Hawke's position follows from an epistemic intuition about disjunctions. Knowing or believing that $(\phi \vee \psi)$ might not entail knowing fully about ϕ and knowing fully about ψ, since knowing or believing a disjunctive fact does not necessarily entail knowing or believing either one of its disjuncts. Knowing that $(\phi \vee \psi)$ might be an intermediate state, where the agent partly knows about ϕ and partly knows about ψ. Whereas knowing that $(\phi \wedge \psi)$ entails knowing that ϕ and knowing that ψ. The topic-sensitive belief revision as introduced in [19, 175, 71] require junction, yet still allow that agents believe that a disjunction is true without believing that either disjunct is true, provided that the agent has already fully grasped the topic of both disjuncts.

It is important to note that the topic of a belief sentence $B\phi$ is then identical with the topic of the sentence ϕ. This, however, might not fit with the intuition that the meaning conveyed by a belief sentence is more complex than the meaning conveyed by the sentence it embeds. While the sentence "Snow is white" can be (informally) said to be about "snow being white", the sentence "Alice believes that snow is white" is also about the information or beliefs of Alice.

The criteria introduced for the behaviour of subject-matter and topics do not apply to the analysis of the meaning captured by support-based hyperintensional semantics. In this section, the analysis suggested by the models in Chapter 2 is taken as the exemplifier of this approach, unless stated otherwise. Let us call the relevant meaning of sentences captured by these models, which include information about both the truth-conditions and the aboutness-conditions, *the supported-meaning* of sentences. First, supported-meaning, as opposed to topics, is not negation-transitive: the set of states which support a sentence and the set of states which support its negation are usually distinct sets (otherwise the sentence is a necessary contradiction), furthermore, these two sets might not be compliments of each other in a given model. The same supported-meaning then cannot be assigned to a sentence ϕ and to its negation (unless ϕ necessarily implies its negation). This is reflected also in terms of the verifier and falsifier states of truth-maker semantics.

Sedlar's representation of *hyperintensional meaning* also makes this point [136]. The

[19] Yablo's account of world-divisions and Fine's account of subject-matter satisfies this as well: the (overall) subject-matter of the sentence is the pair consisting of its truth-makers (subject-matter) and false-makers (anti-matter). Note that ϕ and $\neg\phi$ has the same unordered pair as their verifiers and falsifiers since a state w is a truth-maker for $\neg\phi$ iff it makes ϕ false, and a false-maker for $\neg\phi$ iff it makes ϕ true [52, 171].

meaning of sentences are explicitly represented and separated from their intensions. The meaning assigned to the sentences are represented as abstract *contents* that are members of arbitrary sets, a specific theory of meaning hence is not assumed. His framework subsumes state-based, syntactic and structuralist accounts of meaning as special cases that can be obtained by replacing abstract representation of contents with more specific classes of objects. A hyperintensional generalisation of Montague- Scott semantics is used to achieve this. The hyperintensional models $M = \langle W, C, O, N_C, I \rangle$ consist of two non-empty sets W and C, which stand for a set of states and a set of semantic contents, O is a content function from formulas of the language to C which assigns content to sentences, N_C is a function that assigns to every state in W a subset of C, that is a distinguished set of contents in w, N_C then is a property of contents, and I assigns to every $c \in C$ a proposition $I(c) \subseteq W$. It is important that I assigns propositions to contents rather than to formulas directly. The idea is that propositions are determined by contents and not by intentions. Briefly, given a formula F, there is a content function which assigns F a set of contents from C, hence $O(F) \subseteq C$, and I assigns a proposition to this content, hence $I(O(F))$ is a proposition. In order to assign propositions to formulas, he uses a combination of the functions O and I, and a hyperintensional model is obtained by defining the valuation function for a formula F as follows: $[\![F]\!] = I(O(F))$. It is obvious in this case that it cannot be that $O(p) = O(\neg p) = c$ since this means $I(c) = W - I(c)$, but this is not possible since W is non-empty [136, p.940].

Whether negation is a topic-transitive unary connective or it shifts the aboutness conditions of sentences as part of their supported-meaning affects how it behaves when embedded in belief sentences. First note that negation does not behave classically within the context of hyperintensional belief base revision (BBR) models of Chapter 2: while ϕ and $\neg\neg\phi$ are classically and also BBR-equivalent for propositional formulas, $B\phi$ and $B\neg\neg\phi$ are not BBR-equivalent. The failure of logical entailment between the two belief sentences can be explained by the different complexity levels of the formulas, which transforms to their meaning. A similar argument also applies to explain the difference between topic-identical sentences $\neg B\phi$ and $B\neg\phi$ (the topic of both sentences are identical to the topic of ϕ). While the former can be said to be about the absence of a belief, the latter conveys information about an existing belief of a negated sentence. In particular, the former does not require that the agent is familiar with the meaning or the content of the embedded sentence. This adds also to the intuition about the failure of negative introspection in the context of belief. Note that topic-sensitive belief models make the same distinction, as negative introspection fails there as well, since $\neg B\phi$ does not require that the agent is on top of the topic of ϕ while $B\neg\phi$ does require this. These examples are aimed at showing that negating sentences as well as embedding them in belief sentences indeed shift the aboutness-conditions of sentences and add to their complexity levels. This points out that reasoning agents might be more susceptive to the syntactic complexity of information than is assumed in topic-sensitive frameworks.

Other compositionality principles for topics might also fail for supported-meaning. For instance, the set of states that support $(\phi \vee \psi)$ may be distinct from the set of states that support $(\phi \wedge \psi)$. The supported-meaning of $(\phi \wedge \psi)$, however, includes that of $(\phi \vee \psi)$.[20]

3.3.2 Topic inclusion and disjunctive closure

The two-fold topic-sensitive entailment [19] and the inclusive entailment suggested by Yablo [171] require *topic inclusion* between sentences. Consequently, a belief sentence $B\phi$ entails the sentence $B\psi$ only if the topic of the latter is fully included in the topic of the former (more accurately, in the fusion of the topic of ϕ and the topic of the total belief state). For instance, $B^{\phi}\psi$ then does not entail $B^{\phi}(\psi \vee \chi)$ unless topic of χ is also included in the topic of ϕ. Berto states, however, that there are reasons to relax such a constraint, allowing, e.g., partial overlap of topics rather than full inclusion, for various purposes [24, p.21], and explores different ways and consequences of doing so (in the context of imagining) in [20, p.125].

BBR models as well as inexact truth-making on the other hand suggest weaker inclusion containment requirements for logical entailment. Disjunctive closure in the form from $B\psi$ infer $B(\psi \vee \chi)$ is indeed valid for BBR-logic. It can be argued that disjunctive closure of beliefs describe a rather harmless way of expending the information and belief states of agents via inferences from a stronger content to a weaker one. While the new disjunct χ might be about a new subject-matter, it does weaken the overall content of the belief rather than introducing additional content. Recall from the discussion of junction, that for an agent to believe that ψ, she must fully believe that ψ, while to believe that $(\psi \vee \chi)$ she does not need to believe fully either disjunct. This argument then suggests, partial overlap of topics or subject-matter might suffice to make inferences from stronger content to a weaker one.

Allowing disjunctive closure through shared content is indeed not unusual. In describing tautological entailment, Anderson and Belnap state that when the antecedent and the consequent of an implication (in their normal forms) share an atom, this suffices to say that the consequent is contained in the antecedent [7, p.12,23]. Lewis's theory of world-partitions and inclusive entailment also suggest a logical entailment relation that requires partial overlap of subject-matters, since the subject-matter is assumed to be closed under refinement. That is, the subject-matter of a sentence $(\phi \vee \psi)$ overlaps with the subject-matter of ϕ as the latter is a refinement of the former and both are thus included in the closure ϕ^{+}. Lewis' account of subject-matter, however, does not comply with various other requirements describing the behaviour of subject-matters commonly stated in the literature [52, 69], and has been argued to be an inadequate theory of subject-matter in [171] (see section 2.6).

[20] See [136] for principles of weak compositionality and strong compositionality principles in relation to hyperintensional models.

3.3.3 Negation, semantical opposition and incompatibility

The notion of *semantical opposition* describes one of the most basic intuitions behind the concept of negation.[21] It is classically interpreted in terms of truth and falsity. Two sentences are semantically opposed if the truth of one follows from the falsity of the other, and vice-versa. Such as, the negation of a sentence ϕ is true iff ϕ itself is false, and the negation of a sentence ϕ is false iff ϕ itself is true. This is at the root of classical definitions of truth and falsity of a formula dividing the logical space into two opposite camps. It follows that the truth conditions completely determine the falsity conditions for formulas, consider for instance the classical possible worlds semantics.

This, however, is not always the case in the realm of non-classical logics. For instance, truth-conditions of sentences in many valued logics such as the logic of first degree entailment (FDE), and consequently paraconsistent and paracomplete logics such as logic of paradox (LP) and the weak Kleene logic (K3) do not divide the logical space into two opposite camps. Therefore, truth-conditions no longer determine the conditions of falsity for sentences. Truth and falsity need to be treated as two primitive concepts. This is sometimes done by introducing double induction consequence relations that involve a verification relation and a falsification relation separately, e.g., truth-maker semantics, Nelson Logic (N4). Semantical opposition is then interpreted based on the relations of verification and falsification, instead of truth and falsity: a negated sentence $\neg\phi$ is verified at a state s iff ϕ is falsified at that state, and $\neg\phi$ is falsified at a state s iff ϕ is verified at s. The HYPE-semantics on the other hand achieves primitive treatment of falsity by defining negation based on an incompatibility relation between the HYPE-states. Accordingly, the sentence $\neg\phi$ is supported at a state s iff s is such that it is incompatible with all states where ϕ is supported.[22]

To sum up, treating truth and falsity separately is not uncommon in formal models of information and inference, where truth values are sometimes understood as *told values* rather than being based on truth simpliciter [168]. Wansing states that if we understand the study of inference as the study of information flow, and entailment is supposed to preserve the support of truth rather than truth simpliciter, then “the negative extension of a formula ϕ in general is not the boolean complement of its positive extension” [167]. Thus, the interpretation of semantical opposition and consequently the formal and informal characterisation of negation changes as the focus shifts from truth-conditions to support-conditions, or from classical contexts to hyperintensional contexts.[23]

[21] For a general overview of varieties of negation and basic assumptions, and to see related research fields such as paraconsisteny and contra-classicality, see [74, 166].

[22] It is worth noting that the HYPE negation (and possibly other forms of negation as incompatibility) is not a local connective, i.e., not extensional since in order to see if a negated formula is supported at a state, one needs to look at other states. HYPE negation corresponds to a *star negation* (although this fails in the model language of BBR), and it is possible to see whether a negated formula is supported at a state only by looking at its star image. This is not the primary definition of the HYPE negation, however, since the star operator is not a primitive of the models.

[23] The reinterpretation of semantical opposition and the non-classical characterisations of negation changes also what is considered as contradictory and as trivial states. For instance, the logic of FDE

Provided that the context in which we evaluate sentences is sensitive to the more fine-grained meaning of sentences, semantical opposites should be redefined accordingly. An example of this shift occurs in formal models of belief and inference where agents are allowed to hold non-trivial, inconsistent belief states, within different *reasoning contexts*, which is in deed the subject of Chapter 4.[24] These frameworks avoid trivialisation in the face of contradictory beliefs by stipulating that holding contradictory beliefs is problematic only when these beliefs belong to the same context. Reasoning contexts can then be understood as contributing to the meaning of sentences.

3.4 Concluding remarks

I start the concluding remarks with a note on what defines hyperintensional logics. Wansing and Odintsov suggests a characterisation of hyperintensionality based on the notions self-extensionality or congruentiality[115]:

> [A logic] L is said to be self-extensional if for all formulas ϕ, ψ, θ and all propositional variables p from the language of L, $\phi \dashv\vDash \psi$ implies $\theta(\phi/p) \dashv\vDash \theta(\psi/p)$, where $\theta(\phi/p)$ and $\theta(\psi/p)$ are the results of replacing ϕ for p and ψ for p in θ, respectively. Equivalently, the self-extensionality of L can be defined by requiring that for every n-place connective $\#$, the following congruence rule is admissible: if $\phi_i \dashv\vDash \psi_i$ for $i = 1, \ldots, n$, then $\#(\phi_1, \ldots, \phi_n) \dashv\vDash \#(\psi_1, \ldots, \psi_n)$. A connective is said to be congruential according to L, if the congruence rule for it is admissible in L. [115, p.51]

A logic L is then hyperintensional if it is not self-extensional, and an operator is hyperintensional is it is not congruential according to L. According to this definition, propositional HYPE-logic is self-extensional, and the HYPE-conditional, characterised by Leitgeb as a hyperintensional connective when "measured by the standards of classical logic and semantics" [91, p.393], is congruential, since it is an intensional connective within the hype logic. Wansing and Odintsov provide a proof for the self-extensionality of the HYPE-logic, and state their concerns about Leitgeb's take on hyperintensionality. Their first point concerns embedding sentences from classical logic into the HYPE-logic to form a hyperintensionality measure. The two logics, however, have different vocabularies, for instance, the definitions of negation and conditional come apart in the two. Their second point concerns the choice of classical logic as the reference point for hyperintensionality. They leave it open, however, the discussion of adding to HYPE non-congruential modal operators, as is done here,

(as well BBR) does not treat the formula $\neg(\phi \vee \neg\phi)$ as a contradiction, since $(\phi \vee \neg\phi)$ is not a necessary truth. In the same way, a state in a BBR model which supports a formula $(\phi \wedge \neg\phi)$ is not trivial. A contradiction can be defined based on BBR-logic, as a formula in the form $(\phi \rightarrow \neg\phi)$, instead of $(\phi \rightarrow \bot)$, where ϕ entails a unique falsum-state.

[24] A reasoning context is an informal notion, expressing the aspects of a situation an agent deems relevant in a reasoning task, such as what is at stake for her. For instance, making a courtroom statement and a conversation with friends are different reasoning contexts.

possibly to model attitude reports. Note that the formulas ϕ and $\neg\neg\phi$ are HYPE-logically equivalent when ϕ is propositional, while $B\phi$ and $B\neg\neg\phi$ are not equivalent according to the logic of the BBR models.

While "hyperintensionality" is used here as an umbrella term for contexts, semantics and logics where intensional (or classical logical) equivalences are not respected, there are various approaches and modeling solutions in the relatively recent yet dynamic literature. This chapter presents a limited survey of different approaches to hyperintensional belief revision, and points out the different characteristics that result from these approaches. The focal approaches are topic-sensitive frameworks, which aim to explain hyperintensionality of belief revision based on topic-inclusion, and state-based frameworks, which take issue with the designation of the meaning of sentences as possible worlds and suggest instead an analysis of sentences based on states. These, however, do not exhaust the approaches in the literature. For instance, containment logics and relevant logics (and their combinations) also exemplify hyperintensionality based on topic-inclusion [145, 49], while various impossible-worlds approaches can be categorised as non-normal approaches to hyperintensionality [77, 107]. Fagin and Halpern's awareness logic in [45] can also be listed here, as the awareness models avoid closure under classical equivalence and belief of all classical validities by introducing an awareness filter on a possible worlds modeling for *explicit beliefs* of agents. The awareness operator assigns to each possible state, a set of primitive propositions that the agent is aware at that state, and the agents cannot have explicit knowledge about formulas of which they are not aware.

Topic-sensitive and partition-based reasoning introduced here are similar approaches to hyperintensionality in the sense that both introduce so-called *topicality filters* on classical intensional logics. Topics, as well as subject-matters (i.e., partitions), are assigned not only to the sentences of a language, albeit in different ways, but also to the overall discourse context. The topic-sensitive belief revision takes into account what the agent already believes, i.e., the totality of the subject-matter the agent has grasped already, while different ways of partitioning or dividing a logical space generates a discourse context.

The state-based semantic can also be interpreted along similar lines with the topic-sensitive frameworks, with topics or subject-matter of sentences are understood as *context set by information states*. For instance, Barwise suggests a *relational account of information* based on the situation semantics, where the context set by a state goes beyond the intensional meaning of pieces of information supported by a situation. In particular, informational content of a situation s is information *about* something else, determined in particular by the relations and constraints between different situations. In other words, a situation s contains information about another situation [13, p.139]. It follows that the information available to an agent depend in part which of these constraints the agents know [13, p.140], beyond the logical connections between pieces of information. The *information states* introduced in Chapter 2, determined by the parthood and the preference orderings between states, can then be interpreted as *reasoning contexts*. Berto states the importance of explicitly representing such contexts:

> [B]y representing the topicality component of the contents of attitudes sep-

> arately, it allows one to smoothly focus on modeling agents with conceptual limitations, taking these as limitations in the subject matters one is positioned to grasp [20, p.138].

Modelling of reasoning contexts explicitly to enhance intensional meaning within a dynamic informational semantics then remains an exciting research question.

Chapter 4

Fragmented Belief Based on Information Bases

Various sources of logical omniscience can be identified which result from the modeling assumptions made while representing beliefs and belief dynamics. The belief base revision (BBR) models of Chapter 2 challenge multiple standard assumptions of belief ascription and dynamics. In this chapter, the focus is on rejecting various commonly assumed deductive closure requirements for belief sets. In particular, the belief sets obtained via the BBR models are not necessarily closed under conjunctions and believed implications. Moreover, the structure of the information bases allows a natural way of relaxing the consistency requirements of belief sets using a less cautious belief modality. In the following, a survey of the literature on deductive closure principles for belief sets is presented, with the aim to locate the BBR framework within this discussion. Next, a direction towards modeling *fragmented belief states* is proposed, where the total belief set of an agent at a time is fragmented into separate belief states. The focus of this chapter is mainly on the static aspects of belief, while the discussion can be transformed to apply to the dynamics of belief. Recall that the validities and the invalidities of belief revision, discussed in Section 2.5, depend on closure conditions of belief sets (i.e., features of the static aspects of the BBR models).

4.1 Introduction

Use of the possible worlds semantics and the necessity-like treatment of the doxastic modality impose certain requirements on the representation of belief states of agents. Accordingly, beliefs of agents are expected to satisfy various closure conditions, particularly, the agents are expected to believe all logical consequences of their beliefs. These conditions are sometimes attributed as normative conditions for ideal reasoning agents since realisation of them requires the agents to have ideal logical and cognitive capacities (such as unlimited time and memory). The gap between idealised reasoning described and ordinary reasoning by ordinary agents is called *the problem of logical omniscience*. Some of the commonly dis-

C1 If $B\phi$ and $\phi \models \psi$, then $B\psi$	Closure under logical implication
C2 If $\models \phi$, then $B\phi$	Belief of valid formulas
C3 $\models \neg(B\phi \wedge B\neg\phi)$	Non-contradiction
C4 If $B\phi$, and $\phi \dashv\models \psi$, then $B\psi$	Closure under logical equivalence
C5 If $B\phi$ and $\models \phi \rightarrow \psi$, then $B\psi$	Closure under valid implication
C6 If $B\phi$ and $B(\phi \rightarrow \psi)$, then $B\psi$	Closure under believed implication
C7 If $B\phi$ and $B\psi$, then $B(\phi \wedge \psi)$	Closure under conjunction
C8 If $B\phi$, then $B(\phi \vee \psi)$	Closure under disjunction

Figure 4.1: Deductive closure conditions on belief sets

cussed closure conditions in this context are listed in Figure 4.1. For detailed descriptions of all points see also [46, 162] and [23, p.108]. In the following, B can be read as knowledge or belief, $\models$ is logical entailment, $\rightarrow$ is an indicative conditional.

Abandoning the closure conditions completely is not a viable solution to the problem. Belief attributions are used to make predictions about the agents' actions, and as Stalnaker states, agents should not act on single beliefs, entailments of what agents believe are in fact crucial in guiding their actions [140, p.82]. However, for this practical description of belief attributions to apply to ordinary agents, we need to find ways to relax the idealisations suggested by these closure conditions, at least to some degree. Cherniak [33] states the following as reasons for a search for *minimal rationality* standards, as opposed to the idealised rationality standards that suggest the reasoning agents are logically omniscient. First, we do not assume ideal rationality in everyday belief attribution. Ordinary agents might fail to deduce some consequences of their beliefs, or they might have inconsistent beliefs, yet this does not render them irrational. Second, upholding the ideal rationality standards excludes ordinary humans from having rational beliefs, and even from having beliefs at all. Lastly, he argues, that ideal rationality cannot be achieved by ordinary reasoning agents.

One way we can approach to the problem is to identify the different sources of logical omniscience, and offer different modeling solutions accordingly. Fagin et al. state the following as (some) reasons agents might fall short of idealised rationality:

1. Lack of awareness
2. People are resource bounded
3. People don't always know the relevant rules
4. People don't focus on all issues simultaneously [45, p.40-41]

For instance, distinguishing between explicit and implicit beliefs might address the problem of *lack of awareness* of agents. Levesque's implicit-explicit belief modeling [94], and Fagin and Halpern's awareness logic [45] avoid C2 and C5 above by evaluating the explicit beliefs of agents based on possibly incomplete and inconsistent states (while a possible worlds evaluation is preserved for the evaluation of the implicit beliefs), and by introducing an awareness filter, respectively. Topic-sensitive frameworks can also be understood to address the same point through the topicality filters. Balbiani et al. [10] also suggest a distinction between the implicit and the explicit beliefs using neighborhood semantics, this time in an attempt to address the problem regarding resource bounded reasoning. Resource bounded reasoning, in particular, concerns of limited information access and limited memory are also the basis of Cherniak's minimal rationality standards [33]. Minimal rationality differs from ideal rationality as the agents are not expected to deduce and believe all logical consequences of their beliefs, however, they are expected to achieve this sometimes, and for some of their beliefs. In the following, I will mainly focus on the works which address the problem of deductive closure in terms of *shifting focus situations*. Ordinary agents reason by focusing on parts of the totality of the information they have, determined by the issues they focus on at a time. However, since the sources of the problem are not disjoint, there are overlaps with the other points as well. The aspects of the problem of logical omniscience I am mainly concerned here are C3, C6, and C7.

The preferred modeling solution is based on the idea that the belief states of the agents may be *fragmented* into separate, smaller belief states. Different fragments of an agent's belief state may also be called *frames of mind*, or *centers of rationality*. Each fragment functions as a full belief state in guiding the agent's actions. The agents are usually expected to be ideal reasoners within each fragment, however, they may not be able to put together the conclusions of different fragments, hence failing to have deductive closure w.r.t. the totality of their beliefs. This fragmentation theory of mind is as a way to make sense of the closure conditions for the beliefs, rather than rejecting them altogether. It is proposed that these conditions apply only to the fragments of the belief states. The models of fragmented belief states can be combined with other modeling solutions to address multiple aspects of the logical omniscience problem. For instance, a fragmented belief approach based on incomplete information states is proposed here, which addresses both the problem of deductive closure and the problem of logical closure.

Fragmented belief states are usually characterised by the following criteria:

F1. The total set of an agent's beliefs (at time t) is fragmented into separate belief states.

F2. Each belief state (at t) is a fragment whose constituent beliefs are consistent with each other and closed under logical consequence.

F3. The belief states of a single agent (at t) are logically independent: They may not be consistent with each other, and the agent may not believe the consequences of his belief fragments taken together.

F4. Different belief fragments of a single agent (at t) guide the agent's actions in different contexts or situations. [80, p.4]

While there are quite a few open questions concerning the relationship between different fragments such as the rationale behind intra-fragment and inter-fragment consistency, the above characterisation is widely accepted as a starting point in the literature. Some early examples of models of fragmented belief states are from Stalnaker [140], Lewis [98], and Fagin and Halpern [45]. While Stalnaker focuses on the issue of deductive closure, Lewis focuses on a solution to contain inconsistencies of belief that might occur. Both seem to suggest the idea of fragmentation not as a modeling solution, but rather as an explanation of why these particular states of belief might come to be. Fagin and Halpern's work is only one of the various attempts at a solution to the logical omniscience problem informed by the literature on computation and artificial intelligence, which is connected to the semantics and logic of knowledge and belief.

The interest in the theories of fragmented mind has increased recently, leading to the publication of collected works that state various motivations for fragmentation [80]. The common subject across the collection is the rejection of the principle of *unity*, that states, (i) belief sets do not contain any logical contradictions, (ii) they contain all logical implications of all beliefs, (iii) as well as all logical tautologies, (iv) and they are closed under conjunctions [90, p.33]. The principle of unity and the related closure conditions for knowledge and belief have been upheld in the literature for various reasons. These include aiming for a descriptive theory for idealised rationality, theoretical virtues of modeling closed knowledge or belief states such as simplicity, and the idea that the beliefs aim at truth and at an accurate representation of the world which is assumed to be consistent and complete.

In the next section, I will state what fragmentation means in more detail, along with some motivations and modeling approaches from the literature. Besides what is mentioned above, I will add motivations from *contextual* or *situated* reasoning to the list provided so far. Later, I will introduce a direction for modeling fragmented belief based on fragmented information states.

4.2 Fragmented belief in the literature

4.2.1 Stalnaker: separate centers of rationality

In [140], Stalnaker presents a justification for the use of possible worlds semantics to model belief states based on a pragmatic-dispositional account of belief attributions. According to this picture, intentional mental states such as knowledge and belief should be understood by the role they play in explaining and predicting the actions of rational agents. What is essential to rational action is that the agents are confronted with alternative possible outcomes of their possible actions, they have a way of distinguishing these possibilities, and they have attitudes toward these possibilities. The objects of representational attitudes, including the beliefs of the agents, are exactly these possibilities. Possible worlds represent the possible states of the world, and the attitudes of the agents towards these possibilities guide their actions. The pragmatic-dispositional account of belief attributions motivates a possible worlds analysis of propositions. According to this analysis, propositions are ways of distinguishing between alternative possibilities. They are, in this view, not the objects nor components of beliefs, rather, they are used to characterise and express the beliefs of the agents.

There are, however, two implications of the proposed account of belief attributions that need to be addressed for an adequate theory. The first implication concerns necessary equivalents. The possible worlds account of propositions suggests that there is a unique necessary truth since all necessary truths are expressed by the same proposition, i.e., the set of all possible worlds. Similarly, it suggests that there is a unique necessary falsehood that is expressed by the empty set of possible worlds. We have seen in the previous chapter that Stalnaker addresses this point by means of varying domains of issues. The second implication of the proposed possible worlds account is the requirement that the agents believe all deductive consequences of their beliefs. Stalnaker analyses the requirement of deductive closure in terms of three separate closure conditions [140, p.82]:

1. if ϕ is a member of a set of [beliefs], and ϕ entails ψ, then ψ is a member of that set.
2. if ϕ and ψ are each members of a set of [beliefs], then $\phi \wedge \psi$ is a member of that set.
3. if ϕ is a member of a set of [belief], then $\neg\phi$ is not a member of that set.

The first condition follows from the idea that, in terms of guiding the actions of agents, what is important is not the single beliefs of the agents, but rather the consequences of their beliefs, i.e., the totality of their beliefs. This is because, the objects of the beliefs are the possible states of the world, and not single sentences: to be in a state to believe that ϕ is simply to be in a belief state which lacks any possible worlds in which ϕ is false. Stalnaker argues against a *storage model of beliefs* where belief states are modeled as sets of sentences. He argues, instead, that individual beliefs are merely *properties* of the belief states [142].[1] Hence, based on the practical-dispositional account of beliefs, there is no reason to abandon the first condition. The second condition, on the other hand, while it seems similar to the first, is not a reasonable one to impose on the totality of an agent's beliefs:

> [W]hile the [conjunctive closure] condition is a reasonable one, given the pragmatic account, to impose on the propositions determined by a belief state, it is not a reasonable condition to impose on the totality of an agent's beliefs. It is compatible with the pragmatic account that the rational dispositions that a person has at one time should arise from several different belief states. A person may be disposed, in one kind of context, or with respect to one kind of action, to behave that is correctly explained by one belief state, and at the same time be disposed in another kind of context or with respect to another kind of action to behave in ways that would be explained by a different belief state. This need not be a matter of shifting from one state to another or vacillating between states; the agent might at the same time be in two stable belief states,

[1]This motivates Stalnaker's account of tacit beliefs, which he sees as one of the advantages of the possible worlds analysis of belief states.

> be in two different dispositional states which are displayed in different kinds of situations. If what it means to say that an agent believes that P at a certain time is that some of the belief states the agent is in at that time entails that P, then even if every set of propositions defined by a belief state conforms to the [conjunctive closure] condition, the total set of propositions believed by an agent might not conform to that condition [140, p.83].

Moreover, once it is accepted that an agent can be in different belief states at different times and contexts, there are no grounds for requiring that these belief states are always compatible with one another. Thus, the requirement of consistency falls with the requirement of conjunctive closure, in the context of the totality of an agent's belief state. Stalnaker argues that while *ideally* the totality of an agent's beliefs should be integrated into a unique and consistent system of beliefs, failing to do so does not render an agent irrational, but only points out a divergence from the standard of ideal rationality. That is, the fragmented belief states can still be used to explain the rational actions of the agent. He indeed explains the role of deductive inquiry in terms of the fragmentation of belief states and the ultimate goal of reasoning:

> There may be propositions which I would believe if I put together my separate systems of belief, but which, as things stand, hold in none of them. These are the propositions whose truth might be discovered by a purely deductive inquiry. [W]hat one does [in deductive inquiry] is to transform [the information] into a usable form, and that, it seem plausible to suppose, is a matter of putting it together with the rest of one's information [140, p.85-86].

4.2.2 Lewisian fragmentation

Lewis proposes a theory of fragmented belief states as a strategy to contain inconsistent beliefs to separate fragments, motivated by his famous example:

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (By "roughly" I mean "to within 20°".) So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. [M]y system of beliefs was broken into (overlapping) fragments. [T]he first and second sentences in the inconsistent triple belonged to -were true according to- different fragments; the third belonged to both. The inconsistent conjunction of all three did not belong to, was in no way implied by, and was not true according to, any one fragment. [98, p.435-436]

Lewis's suggestion includes redefining truth and falsity according to a corpus of beliefs or information. Accordingly, a sentence is true if it is explicitly stated in the corpus, or it is logically implied by other sentences in the corpus, and a sentence is false according to a corpus if its negation is true according to the corpus. A corpus can include misinformation such that a sentence might be both correct and false according to a corpus, and if the corpus also includes all deductive consequences of the information it contains then becomes trivial. A resolution, Lewis proposes, is via fragmentation. A fragmented belief set is such that each fragment is deductively closed, however, the closure principles such as logical implication, conjunction, non-contradiction, and disjunctive syllogism (from $A \vee B$ and $\neg A$ infer B) fail to preserve truth w.r.t. to the whole corpus. If something is true according to one fragment, it is true of the whole corpus. In this sense, truth according to a corpus serves as a *fallible* guide to truth *simplicitier* and can rationally guide an agent's actions.

Lewis, although informally, states various features of fragmentation of belief states. Particularly, each fragment is deductively closed and consistent, different fragments may overlap with each other, however, they do not appear all at once, and the agents do not make inferences based on mixtures of the fragments. The fragments come into play in different situations, and *steer the actions of agents* at different times:

> All my actions may be rational in that they are directed toward desired ends and guided by coherent conceptions of the way things are even if there is no single conception of the way things are that guides them all. [140, p.85]

> Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. [98, p.435]

Both Stalnaker and Lewis propose fragmentation as an explanation of why certain deductive closure conditions might fail. Similar to Stalnaker, Lewis also states that fragmentation is a divergence from the ideal of rationality, and that once the agent becomes aware of it, she needs to reason to fix it.

Local Reasoning

Fagin and Halpern [45] present a formal model of *local reasoning* that captures Lewis's approach to fragmentation. They introduce a *local reasoning structure* $M = \{S, \pi, C_1, \ldots, C_2\}$ where S is a set of states, π is a truth assignment from each state $s \in S$ to the primitive propositions, and $C_i(s)$ is a nonempty set of subsets of S. The beliefs of an agent at a state s is given by a set $C_i(s) = \{T_1, \ldots, T_k\}$: the agent sometimes considers T_1 as the set most plausible possible states and believes whatever holds at T_1, and she sometimes considers T_2 as the set of most plausible possible states and believes whatever holds at T_2. The structure is informally described as the agent having a *society of minds*. Each of $T_1, \ldots, T_n$ represents a belief state, within each the agent reasons perfectly, and they might still be inconsistent with one another.

Local reasoning models account for the failures of closure under conjunction and closure under material (believed) implication. Since each fragment $T_1, \ldots, T_n$ is internally deductively closed, the agents are still logically omniscient w.r.t. valid formulas and closure under logical implication (i.e., logical equivalents).

The Preface Paradox

The preface paradox, first introduced by Makinson [110] states that the combination of the requirements of deductive closure on the belief sets, in particular, requirements of conjunctive closure and non-contradiction, leads to a paradox. The paradox states an author who rationally believes the truth of each of the assertions $(s_1, ..., s_n)$ he makes in his book. He also rationally believes that not everything asserted in the book can be true, that is, he believes that at least one of the assertions he makes is false (denoted by the sentence s_{n+1}). While it is rational to believe each of these assertions, a set of statements that contains all of them $(s_1, ..., s_n, s_{n+1})$ is inconsistent. It follows that, when the two deduction conditions, conjunctive closure and non-contradiction, are applied to a belief set, in which each belief is reasonably accepted, they lead to a contradiction: believing that s_{n+1} is true is equivalent to believing that $\neg(s_1 \wedge ... \wedge s_n)$ is true, therefore, the conjunction of the two, $s_1 \wedge ... \wedge s_n \wedge s_{n+1}$, classically implies a trivial belief set.[2] A solution to the paradox suggests, we give up either the conjunction principle or the non-contradiction principle. According to the fragmentation approach, we reject that the conjunction principle applies to the totality of the belief set. To avoid the trivialisation of the belief state, it suffices to contain the inconsistent belief in separate fragments of the belief state. Kyburg and Foley (as well as Stalnaker, following Kyburg) thus reject conjunction as a global rule [88, 54, 140]. Kyburg further argues that the paradox stems from the faulty assumption that all of our beliefs can be represented as one fat (conjunctive) consistent statement [88, p. 77]:

> I probably cannot believe a contradiction or act on one. But I certainly can believe, and even act on, each of a set of statements which, taken conjointly, is inconsistent. [88, p.60]

4.2.3 Partition-sensitive fragmentation

Yalcin [173] presents a *resolution-sensitive* model of fragmentation as an *atlas* that guides the actions of an agent, as a response to the idea that "belief is the map by which we steer" by Ramsey. An atlas is a plurality of maps, where each map represents a consistent way the world might be, without a guarantee of cross-map consistency. The formalisation

[2] See [87, 88] for a detailed analysis of how these principles lead to a contradiction. Kyburg's arguments for rejecting conjunctive closure mainly depend on the contradiction that follows from the conjunction principle (given a body of accepted statements S, the conjunction of any finite number of members of S also belongs to this body) taken together with the weak deduction principle (if S is a body of accepted statements, s_1 belongs to S, and s_1 entails s_2 according to the underlying logic, then s_2 also belongs to S), and the weak consistency principle (there is no member of S which entails everything in the language). A generalised version of the argument, that does not depend on the idea that a contradiction implies everything in a given language, follows from the consistency principle: a reasonably accepted body of statements does not include inconsistencies.

of an atlas is based on the Lewisian theory of world-partitions [99]. Yalcin proposes that the fragmentation of belief states is a result of changing issues or questions that guide the reasoning of the agents (see also [172]). Following Stalnaker, he argues that the objects of the beliefs are possible states of the world, however, the domain of issues represented by a given set of possible states is not universal, nor it is maximally comprehensive of all possible issues. A set of possibilities should be modeled in a way that it includes also the information about the degree to which the agent is able to distinguish these possibilities. That is, which questions these alternatives address, and fail to address. This is the basis of a *Resolution-sensitive model* of a belied state:

> A state of belief is representable as a partial function mapping a resolution of logical space (question, subject matter) to a belief partition (answer, information about the subject matter). An agent's accessible beliefs, relative to a resolution, will be those propositions true throughout the belief worlds and foregrounded by the resolution. Those propositions true throughout the belief worlds but backgrounded by the resolution are the agent's inaccessible, implicit beliefs. [172, p.11]

The resolution-sensitive models then introduce a distinction between the explicit and the implicit beliefs of an agent. Different resolutions of the logical space bring forward different sets of questions or issues, that are then mapped to different belief states of an agent. That is, the ways the agent distinguishes between the possibilities differ based on the domain of issues relative to the specific inquiry she is engaged in. Yalcin's account diverges from the above Stalnaker-Lewis accounts, and in general, from the characterisation of the fragmented belief states given at the beginning of this chapter. First, he does not maintain a restriction that only a single fragment steers the actions of an agent at a time. However, he adds that given a partition, the agents cannot consider two sets of possibilities as candidates for actually, that is, once a resolution is fixed, there can be no fragmentation of beliefs. Moreover, two atlases that make exactly the same distinctions, or cases in which one strictly refines the other are not allowed.

Yalcin's model of fragmentation is similar to Leitgeb's partition-based reasoning models [90]. Leitgeb proposes that the resolution of issues depends on an agent's *epistemic context*. The epistemic context describes the agent's attention, interests, perceived stakes, etc., along with her *degree-of-belief function*, i.e., her subjective probability measure [90]. The epistemic contexts in Leitgeb's theory correspond to the world-partitions discussed in the previous chapter, in particular, both structures represent degrees of content resolution. An important aspect of the epistemic contexts is that the possibilities are individuated differently in different contexts, rather than being eliminated or ignored:

> Different contexts of reasoning are available to an agent at a time, but, presumably, at each time only one context is chosen to be active (implicitly or explicitly) and will thus ground the agent's rational all-or-nothing beliefs at the time. That context, or at least certain aspects of it (most importantly, the

> partition), will be maintained for a certain period of time in which the stability of the agent's beliefs will (hopefully) pay off. But at some point the context (and in particular its crucial aspects, such as the partition) will change again due to changing questions, perceived stakes, interests, and the like. [A]ccordingly, while the logic of beliefs does hold locally within every context, logical inferences across contexts are not licensed unrestrictedly. [90, p.146]

Both Yalcin and Leitgeb model cases of fragmentation as a result of different resolutions of content. Based on the Lewis-Stalnaker model of fragmentation, the belief state of an agent can be fragmented within the same state of content resolution. The shifts between fragments are not necessarily connected to different ways of partitioning the logical space, rather, it is a matter of which possibilities the agent focuses on. The two approaches come apart in important aspects such as whether inconsistent beliefs are allowed based on the same resolution of the logical space. Recall that Leitgeb also describes the shifts between fragments based on the changes in content-resolution:

> I should add that if the agent asserted one sentence after the other in order to express her beliefs in $A_1, A_2, A_3, \dots$, then it would still not be the case that any of these sentences would have to express one proposition in one context and a different proposition in another: rather, which proposition is entertainable by the agent may shift from one context to the next.[90, p.145]

4.2.4 Imperfect information access

Various accounts of fragmented belief states focus on the structure of memory and limitations of information access. According to these approaches, given a reasoning task, the agents do not actively access the totality of information available to them as this is usually beyond the cognitive capacities of ordinary reasoning agents. Rather, they focus on smaller sets of alternatives and premises that are relevant in a given context. The fragmentation approach allows that the information available to an agent is structured in a way, that the agent can make informed searches among the fragments of her information state to pick out the clusters of information that are relevant for a given task and a situation. Cherniak states that searching at random within an unstructured information state would be an unreliable method for information recall. He proposes an information access model based on the structure and the organisation of long-term memory, where information retrieval is done not by searching the entire memory of the agent but by narrower searches based on the purposes at hand. The organisation of long-term memory is necessary for efficient recall of the right beliefs to the short-term memory, that are relevant to make *minimally rational* decisions [33].

Another model of imperfect information access is from Elga and Rayo [44]. They maintain that the parts of a body of information can be accessible for some purposes and inaccessible for others. They suggest, moreover, that which parts of a body of information are accessible depends not only on the questions asked but also on how the questions are framed. They propose a theory of information access based on what they call the *access tables*. The access tables show elicitation conditions on the one side and the information accessible relative to these conditions on the other side. In a toy example they present that the accessible information might differ for the following two questions:

i) Asked for an apartment number given a name

ii) Asked for a name given an apartment number

While both questions can be answered in the same way, e.g., “The person in ‘23-H’ is named ‘Beatrice Ogden”, it might be the case that the relevant piece of information is not accessible when the inquiry of the agent is framed in the second way.

Kindermann connects the issue of information access and fragmentation to conversational backgrounds. He proposes that a conversational common ground, that is, the information that is presupposed and mutually taken for granted by the speakers, is available (and unavailable) to them only relative to a given conversational task. [79].

Lewis has a similar notion of context that is the result of excluding irrelevant possibilities. He defines knowing as follows: “[Agent] S knows [that] P iff P holds in every possibility left uneliminated by S's evidence (or equivalently, iff S's evidence eliminates every possibility in which not-P) - except for those possibilities that [S is] properly ignoring” [100, p.554]. Thus, knowing is possible only with proper ignoring, and based on smaller sets of possibilities. What is and what is not ignored, he argues, is a feature of the particular conversational context [100, p.559].[3] The failures of deductive closure originate from shifting conversational context:

> The premise “I know that I have hands" was true in its everyday context, where the possibility of deceiving demons was properly ignored. The mention of that very possibility switched the context midway. The conclusion “I know that I am not handless and deceived" was false in its context, because that was a context in which the possibility of deceiving demons was being mentioned, hence was not being ignored, hence was not being properly ignored. ([100, p. 564])

4.3 Fragmented belief based on information states

An important feature of the BBR models is that the objects of belief are structured collections of information rather than pieces of information, or total belief sets. An agent's total information is then essentially fragmented, where each part of the total information base of an agent can be understood as a fragment. The structures of the information states

[3]Recall, on the other hand, Leitgeb argues that ignoring possibilities is in conflict with the idea that beliefs aim at truth. Instead, epistemic contexts are supposed to determine the underlying partition of the space of possibilities (among other things). Hence, different contexts individuate the possibilities differently and represent different degrees of content resolution based on the agent's practical interests, or what is salient to them.

provide a general flexibility for the models. For instance, they reflect that information is always acquired in a context, accompanied by other information, and not in isolation. The agents are usually confronted with possibly incomplete and inconsistent theories about the world, and they do not always isolate good parts of a theory from the bad parts, hence, they might refute or accept a theory altogether based on a part of it. Consequently, pieces (and similarly collections) of information might stand and fall together, without being logically connected to each other.[4] How the information is *divided and lumped together* affects also its dynamics and the dynamics of beliefs based on this information. Kratzler notes a similar effect concerning how premise sets are structured in premise semantics, stating "[r]epresenting the content of recommendations, claims, beliefs, orders, wishes, etc. as premise sets thus offers the priceless opportunity to represent connections between propositions in a given premise set." [83, p.19] Example 2.3.3 shows how the dynamics of information states might differ in the BBR models solely based on how information is structured.

Fragmentation of information is responsible also for the failure of deductive closure of the belief sets. While each fragment of information, including the maximal fragment of total of information, is deductively closed, this does not necessarily transfer to the level of beliefs. Barwise states that "information travels at *the speed of logic*, genuine knowledge only travels at the speed of cognition and inference" [14, p.762]. While making inferences based on the available information, agents might fail to combine all their inferences, hence, fail to believe all logical consequences of their inferences. The BBR models suggest that the resulting belief sets, while not necessarily logically closed, are unique and consistent.

In order to achieve the fragmentation of belief states based on the BBR models, we start by introducing a diamond-like belief modality $B^{\cup}$ besides the box-like (intersecting) belief modality B of the models. Let us call the latter a *cautious belief* modality and denote it with $B^{\cap}$ for the purposes of this chapter. In comparison, then, $B^{\cup}$ represents an *inconsistency-tolerant belief* modality. We extend the language of the BBR models without the belief modality B, with respect to the new modalities $B^{\cap}$ and $B^{\cup}$, and define the following:

For all formulas ϕ in the extended language $L_{\cup\cap}$, for all BBR models M, and all states $s \in S$,

[4] Consider reading a certain newspaper. Suppose that you are heavily set on your belief that any piece of information given in this newspaper is highly doubtful and generally incorrect. Thus, when encountered with a piece of information ϕ, which looks reasonable, due to the non-logical circumstances around this piece of information, such as the source of it and other information contained in the same source, you do not accept it as a belief. To see the formal possibility of such scenarios, consider a belief base model M on the situation space $S = \{1, 2, 3, 4\}$, on a language whose literals are $l = \{p, \bar{p}, q, \bar{q}, s, \bar{s}\}$. Let $V(1) = \{p, q, \bar{q}\}$, $V(2) = \{p, s\}$, $V(3) = \{p, q, \bar{q}, s\}$, $V(4) = \{p, s, \bar{s}\}$. Let $(1 \circ 2) \circ 3 = 3$, $2 \circ 4 = 4$, $1 \circ 1 = 1, 2 \circ 2 = 2, 3 \circ 3 = 3, 4 \circ 4 = 4$. Finally, let $1 \perp 1, 3 \perp 3, 4 \perp 4, 1 \perp 3, 2 \perp 4, 3 \perp 4$. Thus, it holds that $1^* = 4$ and $2^* = 3$. Given that the current information of the agent is given in situation 1, the agent does not believe that p (their belief set is empty) although p is among the information of the agent. However, if the current information of the agent is given in the situation 3, they believe that p. This is because, in the former case, the information that p is available only as part of an inconsistent theory. Whereas, in the latter case, it is also available as part of a consistent and unrefuted theory.

- $s \models B^{\cap}\phi$ iff for all $A \in Best_M(s)$, there is $s' \in A$, $s' \models \phi$
- $s \models B^{\cup}\phi$ iff for some $A \in Best_M(s)$, there is $s' \in A$ such that $s' \models \phi$.

An agent then accepts as beliefs all accessible information that are not (strictly) suppressed (by other information) based on her preference ordering. The cautious belief operator $B^{\cap}$ contains all inconsistencies that may arise within the level of information, and generates a unique and consistent belief set. The inconsistency-tolerant belief modality $B^{\cup}$, on the other hand, still generates a unique belief set, that might, however, contain inconsistent beliefs. It is important to note that the inconsistent beliefs might occur only as instances of $B^{\cup}\phi \wedge B^{\cup}\neg\phi$, and not as instances of $B^{\cup}(\phi \wedge \neg\phi)$. That is, the inconsistent beliefs are always based on distinct fragments of the information state, and they cannot be combined together.

Proof. Let M be a BBR model based on the language $L_{\cup\cap}$, and let $s \in S$. Assume $s \models B^{\cup}(p \wedge \neg p)$. Then, for some $A \in Best(s)$, there is a situation u in A such that $u \models p \wedge \neg p$. It follows that $u \perp u$, and that A is inconsistent. By definition, for all $A \in Best(s)$ it holds that A is consistent. Therefore, it cannot be the case that $s \models B^{\cup}(p \wedge \neg p)$. $\square$

As noted, the belief sets generated by the operator $B^{\cup}$ are still unique. Explicitly representing the fragments on the level of beliefs requires the introduction of further structure on the models, I leave this for future work. Here, I introduce two intermediate modalities, that pick out whether a pair of beliefs arise from the same (maximally consistent) fragment of an information state, or from distinct fragments. To accommodate these new modalities, we extend the language $L_{\cup\cap}$ with $S(\phi, \phi)$ and $I(\phi, \phi)$, respectively, and call the new language L^{+}. The new modalities are defined as follows:

- $s \models S(\phi, \psi)$ iff there is an $A \in Best(s)$ with $t, t' \in A$ such that $t \models \phi$ and $t' \models \psi$
- $s \models I(\phi, \psi)$ iff $s \models B^{\cup}\phi$ and $s \models B^{\cup}\psi$ and for no $A \in Best(s)$ it holds that $t, t' \in A$ such that $t \models \phi$ and $t' \models \psi$.

The two modalities are not reducible to each other, and it follows from the satisfaction clauses for $S(\phi, \phi)$ and $I(\phi, \phi)$, that

$$I(\phi, \psi) \wedge S(\phi, \psi) \vdash \perp$$

Example 4.3.1. The following example shows the failure of conjunctive closure, closure under believed implication, and non-consistency for a belief set generated by the inconsistency-tolerant belief modality $B^{\cup}$.

Let M be a fragment of a BBR model (as shown in Figure 4.2) with a state-space $S = \{1, 2, 3, 4, 5\}$ based on the language L^{+} with $l = \{p, \bar{p}, q, \bar{q}, r, \bar{r}\}$. Let the incompatibility relation between the states be such that for all $s, s' \in S$, $s \perp s'$ iff for some $v \in l$, $s \models v$ and $s' \models \neg v$. Let $2 \circ 3 = 3$, $2 \circ 5 = 5$, $4 \circ 5 = 5$, $(3 \circ 4) \circ 1 = 1$ and $1^* = 2, 3^* = 3, 4^* = 5$.

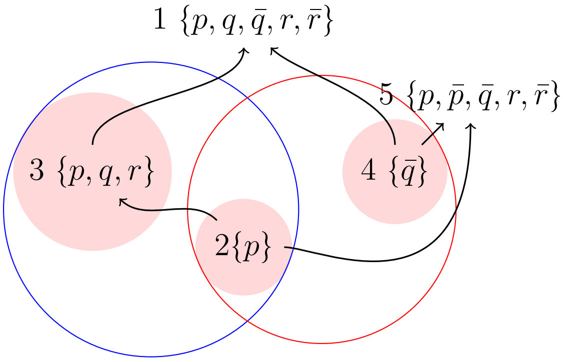

Figure 4.2: Fragment of a BBR model M

Finally, let for all $A, B \subseteq S$ it holds that $A \leqslant_M B$. Then, the following are true in M: since $Best(1) = \{\{2,3\}, \{2,4\}\}$,

$$1 \vDash B^{\cup} q \wedge B^{\cup} \neg q, 1 \nvDash B^{\cup}(q \wedge \neg q)$$
$$1 \vDash B^{\cup} p \wedge B^{\cup}(p \to \neg r), 1 \nvDash B^{\cup} \neg r$$

4.4 Concluding remarks

I conclude this section with a note on the rationality of fragmented belief. The fragmented belief approach shifts the rationality criteria that traditionally apply to the belief sets, to the belief fragments. That is, they are now interpreted as intra-fragment rationality criteria. A full-fledged theory of fragmentation, however, still needs inter-fragment rationality criteria. It is possible only then to account for cases of irrationality, that are not warranted cases of fragmentation. Borgoni suggests [28] inter-fragment rationality to be based on a notion of *epistemic rationality*, as opposed to procedural rationality that is based on coherence. Epistemic rationality refers to the sensitivity of fragments to each other in epistemically relevant ways other than coherence. For instance, the fragments that are sensitive to the same evidence sets are also sensitive to each other, which is called responsiveness to evidence. While the literature on fragmentation of beliefs goes back at least to Stalnaker and Lewis, the discussion on the rationality of fragmentation is fairly new, and presents more questions than it answers.

Chapter 5

Consistency-Sensitive Epistemic Modalities In Information-Based Semantics

In this chapter, an alternative formalisation of the information states is explored, and two novel epistemic modalities are defined. In particular, a framework of information-based semantics for intuitionistic logic is extended with a paraconsistent negation and consistency-sensitive epistemic modalities. In this framework, information states represent information collected from various sources and as such they can be inconsistent because they receive contradictory information either from a single inconsistent source or from various mutually incompatible sources. The modalities reflect only those sources that are consistent and trusted.

Various aspects of the framework presented below overlap with the semantics and logic of the belief base revision (BBR) models, and with the HYPE-semantics. The information-based semantics that is introduced here is a framework for intuitionistic logic enriched with a paraconsistent negation. In [91] it is stated that every Kripke model for intuitionistic logic can be extended to a HYPE-model, and that HYPE contains intuitionistic logic as a subsystem if intuitionistic negation is reconstructed as $A \to \bot$, that is, $A \to \neg\top$, where $\to$, $\neg$, and $\top$ are logical primitives in HYPE [91, p.318].[1] The information-based semantics, however, diverge from the Kripke semantics for intuitionistic logic by allowing the existence of non-prime states: a state can support a disjunction even if it does not support any of its disjuncts. By allowing non-prime states, we can model information states that support

[1] It is shown that every Kripke model of intuitionistic propositional logic may be viewed as the partial part of a so-called intuitionistic propositional HYPE-model. An intuitionistic propositional HYPE-model is constructed based on a sublanguage of propositional HYPE, by dropping the HYPE-negation except for formulas in which it occurs in a $\bot$-context, which captures the intuitionistic negation. The state-spaces of such models consist of two disjoint parts, W and W^+, where all states in W are glut-free (for all formulas A, $s \not\models A \wedge \neg A$). The partial part of intuitionistic propositional HYPE-model means the part that consist only of the states in W. For the construction see [91, p.380].

disjunctive information without supporting any of its particular disjuncts.[2]

The paraconsistent negation of information-based semantics is defined based on a compatibility relation, and is equivalent to the HYPE-negation. The implication ($\rightarrow$) of information-based models also overlaps with that of HYPE's, and it is defined in the fashion of definition 2.2.4. The interpretation of the disjunction, however, diverges from the HYPE-models. The information-based semantics includes a join operation ($\sqcap$), that gives the *common part* of two states, besides a meet operation, which gives the fusion of two states. Either of these operations can be used to determine a partial order on the states. Disjunction is defined based on the common part of two states: for s, t, u states in a model, and for all formulas α, β,

$$s \models \alpha \vee \beta \text{ iff } s = t \sqcap u \text{ for some } t,\, u \text{ such that } t \models \alpha \text{ and } u \models \beta$$

The other differences between the following models and the BBR models include that in the former, the meet and join operations are defined for any two states, that frames are distributive lattices with a top element (that is the trivially inconsistent state), and that the existence of the star images of states are not assumed. The epistemic modalities introduced below are different from the modalities studied in Chapter 2. The evidence modality E reflects the information of an agent that comes from the trusted sources, and the modality B is a *persistent* belief modality, which again reflects the trusted sources of information.

The content of this chapter is joint work with Vít Punčochář, Marta Bílková, and Thomas M. Ferguson and it is based on parts of the paper "Consistency-Sensitive Epistemic Modalities in Information-Based Semantics", published in Studia Logica [126]. Parts of the published paper are not included in this paper. In the omitted parts, the authors prove completeness, the disjunction property, the finite model property and various other results. Furthermore, they consider an alternative treatment of negation, based on a characterisation of negation in the style of the bilateralist setting, following e.g., the work of David Nelson on constructible falsity [114].

5.1 Introduction

In this paper we study some epistemic operators in the context of a semantic framework for intuitionistic logic enriched with a paraconsistent negation. The general framework is based on structures interpreted as algebras of information states as in [124, 129]. Negation is defined in terms of a compatibility relation among the states in the style of [42]. We assume that there is one trivially inconsistent state that supports every formula. However, we also allow for states that are non-trivially inconsistent in the sense that they support a formula as well as its negation without collapsing into the trivial state.

In contrast to Kripke semantics [84], the states of our semantics are not always prime: a state can support a disjunction even if it does not support any of the disjuncts. This

[2] Recall that the BBR states can support disjunctive beliefs without supporting any of its disjuncts.

is the distinguishing feature of the "information-based" semantics, which is a framework that has been most significantly used in the context of inquisitive logic [34, 35] for a formal representation of questions, and, under the name "team semantics", in dependence logic [153, 174] for a logical representation of functional dependence. A generalized information-based semantics was introduced in [124, 125] in order to provide a framework for logics of questions based on various non-classical logics of declarative sentences. Negation and disjunction in such a generalized setting were recently explored in [128, 127]. In the present paper we use this generalized version of information-based semantics as a basic framework in which we introduce and study epistemic modalities.

Disjunction of intuitionistic logic is usually interpreted in a constructive way, and thus it might seem that its semantics should allow only for prime states (like Kripke semantics does). However, this is a mistaken view if we take into account that information states may involve also information that is merely assumed. By making a hypothetical assumption we move from one state into a new one. For example, if an intuitionist assumes, for the sake of an argument, the following piece of information:

$$x = 5 \text{ or } x = 6$$

she moves to a state that supports neither $x = 5$ nor $x = 6$. We should also not exclude non-prime states, if we allow (as we do in this paper) information states to be formed by information that is received from a source without a proper justification. In this way, one can receive disjunctive information that does not support any of the disjuncts. This picture is fully compatible with intuitionistic logic which is reflected by the general fact that the following schema is not intuitionistically valid:

$$(\alpha \to (\beta \vee \gamma)) \to ((\alpha \to \beta) \vee (\alpha \to \gamma)).$$

Since the states of this framework are not prime, the semantics is more similar to the so-called Beth semantics [25] than to Kripke semantics. It is even more similar to the framework described in [37]. In our setting, disjunction is defined as what two states have in common if they respectively support the disjuncts. For this characterization of disjunction we need an operation assigning to any two states s, t their common content $s \sqcap t$. This operation determines a partial order $\leqslant$ among states ($s \leqslant t$ iff $s \sqcap t = s$) which can be viewed as a part-whole relation (s is a part of t iff the common content of s and t is the whole s). The same order can be determined also by join: $s \leqslant t$ iff $s \sqcup t = t$, where $\sqcup$ can be viewed as fusion of information. Thus every state can be viewed as having its own internal mereological structure determined by the states below it. As indicated in Fig. 5.1, a given model also determines the possible extensions of the state.

Strongly inspired by [27, 26, 29] we conceive of an information state s as a site where information from various sources has been collected. The information received from a particular source forms a part of s, i.e. a state below s. A peculiar feature of our framework is that even if the state s is viewed as the state of an agent, it does not represent the agent's actual beliefs. According to this picture, s is just the accumulation of all incoming information from all sorts of sources and the agent must be typically very selective about

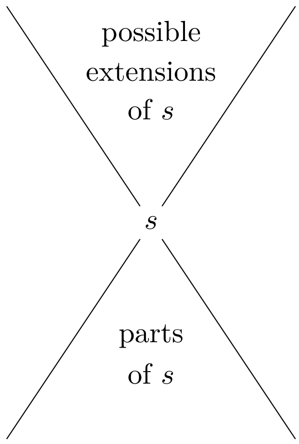

Figure 5.1: States with a mereological structure

it. Not every source is reliable and it is not reasonable to accept every incoming piece of information. After all, if we just collect the information that is coming from a variety of sources, the result will typically be an inconsistent body of information. The process of forming safe beliefs on the bases of information coming from a variety of different channels is highly important. The modalities studied in this paper qualify the collected pieces of information in a way that captures some significant formal and normative features of this process.

We do not describe a mechanism that would allow us to separate the "good" sources from the "bad" ones. That would be a great achievement but it would be a different project. In this paper, we just specify some basic normative constraints that any such mechanism must respect, including the requirement that the good sources are consistent (but not necessarily mutually compatible). Our task will be to describe the logic of some modal operators defined in terms of the selection mechanisms satisfying these constraints.

In this sense we will work with a very basic and simplistic picture. In reality there might be much richer structure on sources of incoming information. We consider just a natural starting point, based on the assumption that there is some kind of distinction between "good" sources and the "bad" ones. To this effect, we introduce a relation S among the states and we read tSu as "t is a part of u consisting of information received from a trusted source". We will simplify this more accurate formulation by saying just that t is a trusted source of u.

Any part of a trusted source will be regarded as trusted. Moreover, consistency of t will be regarded as necessary (but not sufficient) condition for tSu. So, the relation S selects for each state s a downward closed set of consistent parts of s as illustrated in Fig. 5.2.

We extend the framework of information-based semantics for intuitionistic logic with a weak, paraconsistent negation and two epistemic operators E and B. The intuitive meaning of these operators can be described as follows: (a) $E\alpha$ says that α is supported

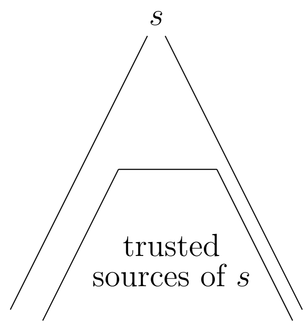

Figure 5.2: Trusted sources of s selected by the relation S

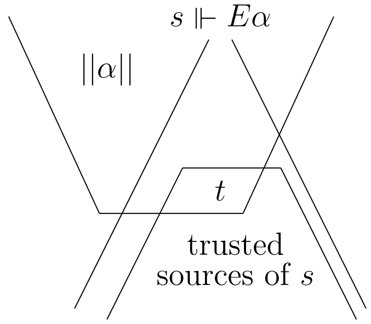

Figure 5.3: s supports $E\alpha$ iff a trusted source of s supports α

by a trusted source; (b) $B\alpha$ intuitively says that α can be safely accepted because it is not in conflict with any potentially reliable information.

E is just a backward looking existential modality defined in terms of the relation S. Its meaning is illustrated in Fig. 5.3. Its logical characterisation ("factive", consistent, monotonic) suggests that E can function as a guiding modality in formally describing doxastic attitudes such as acceptance (as belief) and safe acceptance. The semantics of B is a bit more complicated. The support of $B\alpha$ means that for any extension of the current state, every trusted source is contained in a trusted source that supports α. Note that requiring that α is supported by every possible trusted source (i.e. every trusted source of any extension of the current state) would be too strong. There might be good sources of information that are completely unrelated to α. But $B\alpha$ is supported (i.e. α can be safely accepted) only if no good source is in conflict with α in the sense that every such source is contained in a good source that already supports α, as illustrated in Fig. 5.4. We will see that in the models where every trusted source is included in a maximal trusted source (which must be the case for example in every finite model) the characteristic clause for B can be expressed more concisely: $B\alpha$ is supported by s iff α is supported by each maximal trusted source of every extension of s.

The modal operators B and E are rather non-standard. Before we present them purely

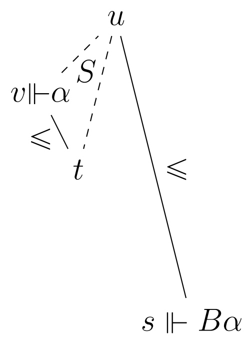

Figure 5.4: s supports $B\alpha$ iff for every extension of s, every trusted source is contained in a trusted source that supports α

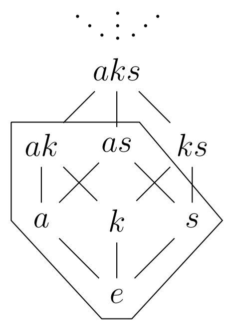

Figure 5.5: The state aks represents the information state formed by collecting the information from Alan, Kate and Shawn. We assume $a \Vdash \neg h$, $k \Vdash h \vee l$, $s \Vdash g$. We assume that k is not compatible with s. The delimited area represents the trusted sources of aks.

formally, let us illustrate their behaviour with a simple informal example. Assume we want to find out where our friend Jane is. She does not pick up our calls and does not react to our messages. We start collecting information about where she might be. We call her friend Kate who tells us that Jane is studying at home or in the library. Then we call her partner Alan who tells us that Jane is not at home. We decide to go to the library but then we meet Shawn who claims that Jane is in the gym. We trust all these people but Kate's and Shawn's claims are mutually incompatible. In our framework we can represent the situation as in Fig. 5.5 (the picture does not represent a fully specified model of our semantics, we just describe some of the relations among the relevant states).

The states a, k, s represent the information collected respectively from Alan, Kate, and Shawn. The state e would represent the common content of a, k and s, which basically corresponds to the disjunction of the pieces of information provided by Alan, Kate and Shawn. The state aks is our current state given by a fusion of these three states. The state ak is the fusion of the information provided by Alan and Kate, and so on. We assume that k and s are mutually incompatible states and so the fusion of these states is not regarded as a trusted source. The area delimited in the picture is the collection of the trusted sources of aks.

We assume that a supports the information that Jane is not at home ($a \Vdash \neg h$), k the information that Jane is at home or in the library ($k \Vdash h \vee l$) and s the information that Jane is in the gym ($s \Vdash g$). We can further assume that $ak \Vdash l$. Then we obtain that aks supports e.g. $l \wedge g$, and also El and Eg separately, but not $E(l \wedge g)$ (we have some positive evidence that Jane is in the library as well as some positive evidence that she is in the gym but that does not mean that we have a positive evidence that she is in the library *and* in the gym).

The information that Jane is not at home is supported by ak as well as by as, i.e. by all maximal trusted sources of aks, and if we assume that this feature is preserved by all expansions of aks present in the model then $aks \Vdash B\neg h$ (but note that $\neg h$ is not supported by all trusted sources of aks). The fact that expansions are taken into account can be illustrated as follows. We can further assume that s is its own unique maximal trusted source, and thus all maximal trusted sources of s support g. However, since aks is in the model regarded as a possible expansion of s, we have $s \nVdash Bg$ because aks has a maximal source (namely ak) that does not support g.

In the rest of this paper we will proceed as follows. We will introduce the information-based semantics for intuitionistic logic in the next section. In Section 5.3 we will extend this semantics with a weak, paraconsistent negation and the two epistemic operators E and B. We will axiomatize the logic of these operators in Section 5.4. We will also define (in Section 5.5) and axiomatize (in Section 5.6) a conditional modification of B. This will result in a rather unusual epistemic conditional construction $\alpha \rhd \beta$ that is roughly interpreted as follows: β is not in conflict with any potentially reliable information supporting α.

5.2 Information-based semantics for intuitionistic logic

Let us start formulating the whole framework more precisely. We first introduce the "information-based semantics" for intuitionistic logic which we will build on in the subsequent sections. It is based on [124]. A very similar semantics was developed in [37].

Consider the language L_0 of intuitionistic logic:

$$\alpha ::= p \mid \bot \mid \alpha \to \alpha \mid \alpha \vee \alpha \mid \alpha \wedge \alpha$$

We assume that the language is based on a countable infinite set of atomic formulas At. Let us define $\top$ as the formula $\bot \to \bot$, and $\alpha \leftrightarrow \beta$ as $(\alpha \to \beta) \wedge (\beta \to \alpha)$. In the semantics, *frames* are distributive lattices (representing algebras of information states) with the top element i (representing trivially inconsistent state). A *filter* in such a frame $\mathcal{F} = \langle A, \sqcap, \sqcup, i \rangle$ is a non-empty subset of A which is upward closed (w.r.t. the lattice ordering) and closed under $\sqcap$. A *model* $\mathcal{M}$ is a frame $\mathcal{F}$ equipped with a *valuation* V defined as a function assigning to each atomic formula a filter in $\mathcal{F}$. The semantic clauses for the logical symbols of the language L_0 define a support relation $\Vdash$ between states of any given model and L_0-formulas. These clauses are defined as follows:

- $s \Vdash p$ iff $s \in V(p)$, for every atomic formula p,
- $s \Vdash \bot$ iff $s = i$,
- $s \Vdash \alpha \to \beta$ iff for all t, if $s \leqslant t$ and $t \Vdash \alpha$ then $t \Vdash \beta$,
- $s \Vdash \alpha \vee \beta$ iff $s = t \sqcap u$ for some t, u such that $t \Vdash \alpha$ and $u \Vdash \beta$,
- $s \Vdash \alpha \wedge \beta$ iff $s \Vdash \alpha$ and $s \Vdash \beta$.

One can easily show that in every model each L_0-formula expresses a filter, by which we mean that the set of states supporting the formula forms a filter.

We say that an L_0-formula is *logically valid* if it is supported by every state of every model. It can be proved (see, e.g., [124]) that the set of logically valid L_0-formulas corresponds to intuitionistic logic.

5.3 Paraconsistent negation and consistency-sensitive modalities

We now extend the propositional language of intuitionistic logic with a paraconsistent negation $\neg$ (which is not defined in the intuitionistic way as implication of contradiction) and the two epistemic operators E and B:

$$\alpha ::= p \mid \bot \mid \alpha \to \alpha \mid \alpha \vee \alpha \mid \alpha \wedge \alpha \mid \neg\alpha \mid E\alpha \mid B\alpha$$

The resulting language will be called L_1. For this language we extend the semantics presented in the previous section. It is based on the following class of models.

Definition 5.3.1. An info-frame is a tuple $\mathcal{F} = \langle A, \sqcap, \sqcup, i, C, S\rangle$, where $\langle A, \sqcap, \sqcup\rangle$ is a distributive lattice with the top element i (the trivially inconsistent state), C and S are binary relations on A (representing respectively the compatibility relation and trusted source relation), for which we assume number of constraints:

(a) C is symmetric;

(b) there is no t such that tCi;

(c) $tC(u \sqcap v)$ iff tCu or tCv;

(d) if tSu and $u \neq i$ then tCt;

(e) iSi;

(f) if tSu then $t \leqslant u$;

(g) if tSu and tSv then $tS(u \sqcap v)$;

(h) if tSu, $t' \leqslant t$ and $u \leqslant u'$, then $t'Su'$.

A valuation in an info-frame $\mathcal{F}$ is a function that assigns a filter in $\mathcal{F}$ to each atomic formula. An info-model is a pair $\langle \mathcal{F}, V \rangle$, where $\mathcal{F}$ is an info-frame and V is a valuation.

The constraints (a)-(h) play particular technical roles but they also have a clear intuitive meaning. In natural language, these can be summarized as follows: the constraint (a) simply acknowledges that compatibility is symmetric; the constraint (b) says that the trivially inconsistent state is compatible with no other state; the constraint (c) says that for any two sources u and v, either is compatible with a state t precisely when their common content is compatible with t (see [125] for a detailed justification); the constraint (d) says that only consistent sources of the non-trivial states are trusted; (e) says that the trivial state regards even itself as a trusted source; (f) says that the body of information received from a particular trusted source forms a part of the body of information collected from all sources; (g) says that if t is a common trusted source for u and v then t is also a trusted source for the state formed as the common informational content of u and v; (h) says that whenever t is a trusted source for u, it remains trustworthy for all information states extending u, as does any part of the information collected by t.

Given an info-model $\mathcal{M} = \langle \mathcal{F}, V \rangle$ we define the relation of support $\Vdash$ between the states in $\mathcal{M}$ and L_1-formulas. The semantic clauses for the intuitionistic connectives are defined as in the previous section. The new operators are semantically characterized in the following way:

- $s \Vdash \neg\alpha$ iff for all t, if tCs then $t \nVdash \alpha$,
- $s \Vdash E\alpha$ iff for some t, tSs and $t \Vdash \alpha$,
- $s \Vdash B\alpha$ iff for all $u \geqslant s$ and tSu there is v such that $t \leqslant v$, vSu and $v \Vdash \alpha$.

Let $||\alpha||_{\mathcal{M}}$ denote the set $\{s \in A \mid s \Vdash \alpha \text{ in } \mathcal{M}\}$ (the subscript will often be omitted if clear from the context). The set $||\alpha||$ is always a filter.

Proposition 5.3.1. $||\alpha||_{\mathcal{M}}$ is a filter, for every L_1-formula α and every info-model $\mathcal{M}$.

Proof. By induction. Let us prove just the inductive steps for E and B. Assume that $||\beta||$ is a filter. We will prove that $||E\beta||$ and $||B\beta||$ are also filters.

Clearly $i \Vdash E\beta$, because of Definition 5.3.1-(e), and the inductive assumption. $||E\beta||$ is upward closed because of Definition 5.3.1-(h). Assume $s \Vdash E\beta$ and $t \Vdash E\beta$, i.e. there is $s'Ss$ such that $s' \Vdash \beta$ and there is $t'St$ such that $t' \Vdash \beta$. By the inductive assumption $s' \sqcap t' \Vdash \beta$. Moreover, by Definition 5.3.1-(g) and (h), $(s' \sqcap t')S(s \sqcap t)$. Hence, $s \sqcap t \Vdash E\beta$.

It holds $i \Vdash B\beta$, because of Definition 5.3.1-(e), and the inductive assumption. $||B\beta||$ is clearly upward closed. We will check that it is also closed under $\sqcap$. Assume $s \Vdash B\beta$ and

$t \Vdash B\beta$. In order to prove that $s \sqcap t \Vdash B\beta$, take any u, v such that $s \sqcap t \leqslant u$ and vSu. Take $s' = s \sqcup u$ and $t' = t \sqcup u$. Since $s \sqcap t \leqslant u$, and $\langle A, \sqcap, \sqcup \rangle$ is a distributive lattice, we obtain $s' \sqcap t' = u$. Moreover, from Definition 5.3.1-(h), we obtain vSs' and vSt'. Since $s \Vdash B\beta$ and $t \Vdash B\beta$, there are s'', t'' that both support β and such that $v \leqslant s''Ss'$ and $v \leqslant t''St'$. By Definition 5.3.1-(g) and (h), we obtain $v \leqslant (s'' \sqcap t'')S(s' \sqcap t')$. Hence, there is w (namely $w = s'' \sqcap t''$) such that $v \leqslant wSu$, and $w \Vdash \beta$. Hence, $s \sqcap t \Vdash B\beta$. $\square$

The logic determined by this semantics might be called *paraconsistent intuitionistic logic with consistency-sensitive modalities*. In this paper, we will denote it simply as $\mathsf{L1}$. We say that an L_1-formula α is $\mathsf{L1}$-*valid*, and we write $\vDash_{\mathsf{L1}} \alpha$, if α is supported by every state in any info-model. It is easy to check that for example all the instances of the schemata A1-A12 from Fig. 5.7 in the next section are $\mathsf{L1}$-valid. Importantly, the following two schemata are not $\mathsf{L1}$-valid:

$$(E\alpha \wedge E\beta) \to E(\alpha \wedge \beta), \;\; E(\alpha \vee \beta) \to (E\alpha \vee E\beta)$$

In order to verify this, consider the info-model in Fig. 5.6. The dashed arrows lead from each state to its maximal trusted sources. So, for example, i is related via S to all states in the model and $s \sqcup t$ is related to s, t and $s \sqcap t$. The dotted ellipses show that $V(p)$ is the filter generated by s and $V(q)$ is the filter generated by t. The compatibility relation will not play any role in our example but we can fix for instance: uCv iff $u \neq i$ and $v \neq i$. In this model we have $s \sqcup t \Vdash Ep \wedge Eq$ but $s \sqcup t \nVdash E(p \wedge q)$. We also have $s \sqcap t \Vdash E(p \vee q)$ but $s \sqcap t \nVdash Ep \vee Eq$.

In [27, 26] a modality analogous to our E was introduced with the same semantic characterization as a backward looking existential modality relative to a source relation. However, since disjunction in [27, 26] was characterized by the standard semantic clause (disjunction is supported iff a disjunct is supported) the schema $E(\alpha \vee \beta) \to (E\alpha \vee E\beta)$ was forced to be valid. This is a problematic feature regarding the intuitive meaning of E and fails to comport with natural intuitions. In general, information sources need not be prime, *e.g.*, a report that one of two candidates will win an election need not indicate which candidate will win. (See [129] for further discussion of failures of this type of distribution.) But if the information states are not necessarily "prime", like in our current setting, one can avoid this consequence in an elegant and natural way.

Concerning the informal interpretation of the modality E, one can notice that it is to some extent similar to the "evidence modality" from evidence logic [160, 157]. In evidence logic it is also possible to have conflicting evidence from different sources in the sense of having evidence for α and evidence for β without having evidence for $\alpha \wedge \beta$. There are however, many technical differences between the two frameworks. For instance, in contrast to our setting, evidence logic is based on classical logic and the evidence modality is semantically defined in terms of neighbourhood models. It would be interesting to compare our framework with that of evidence logic but it would require too much space and we have to leave it for another occasion.

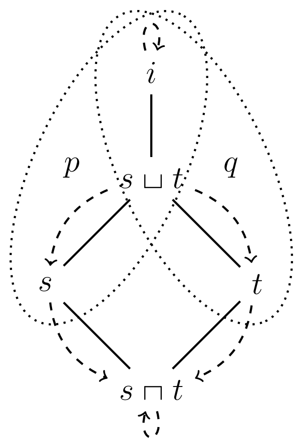

Figure 5.6: A counterexample to $E(\alpha \vee \beta) \to (E\alpha \vee E\beta)$ and $(E\alpha \wedge E\beta) \to E(\alpha \wedge \beta)$

5.4 Axiomatic characterization of L1

We will show that L1-validity can be completely axiomatized by the system in Fig. 5.7. The axioms A1-A9 and the rule R1 is just an axiomatization of intuitionistic logic, so the specific axioms and rules of our system are A10-A12 and R2-R4. We will use this notation: $\vdash_{\mathsf{L1}} \alpha$ means that α is provable in the system from Fig. 5.7; $\alpha \vdash_{\mathsf{L1}} \beta$ means that $\vdash_{\mathsf{L1}} \alpha \to \beta$; $\alpha \equiv_{\mathsf{L1}} \beta$ means that $\alpha \vdash_{\mathsf{L1}} \beta$ and $\beta \vdash_{\mathsf{L1}} \alpha$.

Let us explain the motivation behind the axiom A12 and rule R4. Informally, A12 can be justified by considering that when $B\beta$ holds at s, every trusted source for s itself is *a fortiori* compatible with β. If α is supported by a trusted source t, then, the compatibility between t and β entails that t can be extended—perhaps with "virtual" information—to a further source supporting both α *and* β that continues to be trusted at s.

More formally, we can observe that if we had propositional quantification in the language we would obtain the following relation between E and B:

$$B\beta \leftrightarrow \forall p(Ep \to E(\beta \wedge p))$$

where it is assumed that p does not occur in β. This equivalence is not expressible in L_1 but note that in the context of the left-to-right implication, instantiation of the universal quantifier amounts to

$$B\beta \to (E\alpha \to E(\beta \wedge \alpha))$$

which can be more compactly expressed as the axiom A12. Moreover, in the context of the right-to-left implication, generalization for universal quantifier amounts to the rule

$$\alpha \to (Ep \to E(\beta \wedge p))/\alpha \to B\beta$$

where p does not occur in α and β. This can be reformulated more compactly as the rule R4.

Axioms

A1	$\bot \to \alpha$
A2	$\alpha \to (\beta \to \alpha)$
A3	$(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$
A4	$\alpha \to (\alpha \vee \beta)$
A5	$\beta \to (\alpha \vee \beta)$
A6	$(\alpha \to \gamma) \to ((\beta \to \gamma) \to ((\alpha \vee \beta) \to \gamma))$
A7	$\alpha \to (\beta \to (\alpha \wedge \beta))$
A8	$(\alpha \wedge \beta) \to \alpha$
A9	$(\alpha \wedge \beta) \to \beta$
A10	$E\alpha \to \alpha$
A11	$E(\alpha \wedge \neg\alpha) \to \bot$
A12	$(E\alpha \wedge B\beta) \to E(\alpha \wedge \beta)$

Rules

R1	$\alpha, \alpha \to \beta / \beta$
R2	$\alpha \to \neg\beta / \beta \to \neg\alpha$
R3	$\alpha \to \beta / E\alpha \to E\beta$
R4	$(\alpha \wedge Ep) \to E(\beta \wedge p) / \alpha \to B\beta$
	assuming that p does not occur in α and β

Figure 5.7: A Hilbert style axiomatization of L1

As an illustration of a simple use of these specific rules we will show that every instance of $(B\alpha \wedge B\beta) \to B(\alpha \wedge \beta)$ is provable in the system from Fig. 5.7. Assume that p is an atomic formula that does not occur in α and β.

1. $(Ep \wedge B\alpha) \to E(p \wedge \alpha)$, by A12

2. $(E(p \wedge \alpha) \wedge B\beta) \to E(p \wedge \alpha \wedge \beta)$, by A12

3. $(Ep \wedge B\alpha \wedge B\beta) \to E(p \wedge \alpha \wedge \beta)$, from 1. and 2. by intuitionistic logic

4. $(B\alpha \wedge B\beta) \to B(\alpha \wedge \beta)$, from 3. by R4

We will also show that necessitation, i.e. the rule $\alpha/B\alpha$, is admissible in the system, that is, if α is provable then $B\alpha$ is also provable:

1. α, assumption

2. $p \to (\alpha \wedge p)$, from 1. by intuitionistic logic

3. $Ep \to E(\alpha \wedge p)$, from 2. by R3

4. $(\top \wedge Ep) \to E(\alpha \wedge p)$, from 3. by intuitionistic logic

5. $\top \to B\alpha$, from 4. by R4

6. $B\alpha$, from 5. by intuitionistic logic

Note that the negation $\neg$ is very weak. It satisfies only a form of contraposition expressed in the rule R2, and as such it is a paraconsistent negation: $(\alpha \wedge \neg\alpha) \to \beta$ is not a valid schema. The rule R2 could be replaced, without any impact on the set of provable formulas, by double negation introduction $\alpha \to \neg\neg\alpha$ and a rule of contraposition in this modified form: $\alpha \to \beta/\neg\beta \to \neg\alpha$. This is sometimes regarded as a minimal characterization of negation. For example, according to [17], "nothing can be called a negation properly if it does not satisfy (Minimal) Contraposition and Double Negation Introduction". This might be however, viewed as too harsh, given that there are some intriguing non-classical logics in which negation does not validate the rule of contraposition.

5.5 A conditional counterpart of B

In this section we will introduce a conditional counterpart $\rhd$ of the modality B. We will employ a language L_2 in which B is replaced with $\rhd$ (we will see that B will be definable by $\rhd$):

$$\alpha ::= p \mid \bot \mid \alpha \to \alpha \mid \alpha \vee \alpha \mid \alpha \wedge \alpha \mid \neg\alpha \mid E\alpha \mid \alpha \rhd \alpha$$

Informally speaking, $\alpha \rhd \beta$ says that for every extension of the current state every trusted source that supports α is contained in a trusted source that supports β. This is illustrated in Fig. 5.8 and defined more precisely by this semantic clause:

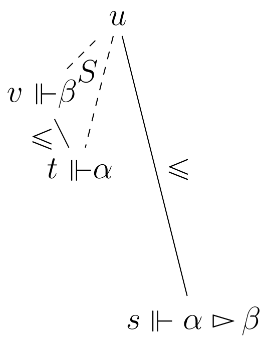

Figure 5.8: s supports $\alpha \rhd \beta$ iff for every extension of s, every trusted source supporting α is contained in a trusted source supporting β

- $s \Vdash \alpha \rhd \beta$ iff for all $u \geqslant s$ and tSu such that $t \Vdash \alpha$ there is v such that $t \leqslant v$, vSu and $v \Vdash \beta$.

Note that $B\alpha$ can now be defined as $\top \rhd \alpha$. So, the conditional $\rhd$ is a natural conditional counterpart of B that allows us to regard B as a defined connective. One could proceed the other way around and define an interesting conditional connective in terms of B, for example by the formula $B(\alpha \to \beta)$ or by $B\alpha \to B\beta$. We plan to explore the properties of these conditionals in future research.

An important feature of $\rhd$ is that, on the algebraic level, it is an operation on filters.

Proposition 5.5.1. $||\alpha||_{\mathcal{M}}$ is a filter, for every L_2-formula α and every info-model $\mathcal{M}$.

Proof. By induction. The inductive step for $\rhd$ is analogous to the step for B in the language L_1 (see the proof of Proposition 5.3.1). $\square$

One can observe that the semantics of $\rhd$ in a sense generalizes intuitionistic implication, by reflecting the structure of trusted sources. Consider the special cases of info-models in which sCt iff $s \sqcup t \neq i$, and sSt iff $s \leqslant t$, that is, s is compatible with t iff the fusion of s and t does not collapse into the trivially inconsistent state, and each part of any state is considered as trusted.

It can be checked that if C and S are defined in this way in a distributive lattice with the top element i then the resulting structure is indeed an info-model. Moreover, in this model negation collapses into the intuitionistic negation, i.e. for every state s, $s \Vdash \neg\alpha$ iff $s \Vdash \alpha \to \bot$. Moreover, the epistemic modalities become trivial in this way: for every state s, $s \Vdash E\alpha$ iff $s \Vdash B\alpha$ iff $s \Vdash \alpha$. In these particular cases, the implication $\rhd$ becomes equivalent to the intuitionistic implication: for every state s, $s \Vdash \alpha \rhd \beta$ iff $s \Vdash \alpha \to \beta$.

We say that an L_2-formula α is $\mathsf{L2}$-*valid*, and we write $\vDash_{\mathsf{L2}} \alpha$, if α is supported by every state of every info-model. Here are some examples of general logical principles governing $\rhd$.

Proposition 5.5.2. For any L_2-formulas α, β, if $\alpha \to \beta$ is $\mathsf{L2}$-valid then $\alpha \rhd \beta$ is $\mathsf{L2}$-valid. Moreover, the following formulas are $\mathsf{L2}$-valid, for all L_2-formulas α, β, γ:

(a) $(\alpha \wedge \neg\alpha) \rhd \beta$,

(b) $((\alpha \rhd \beta) \wedge (\beta \rhd \gamma)) \to (\alpha \rhd \gamma)$,

(c) $(\alpha \rhd (\beta \rhd \gamma)) \to ((\alpha \wedge \beta) \rhd \gamma)$,

(d) $(\alpha \rhd (\alpha \rhd \beta)) \to (\alpha \rhd \beta)$,

(e) $((\alpha \rhd (\beta \rhd \gamma)) \wedge (\alpha \rhd \beta)) \to (\alpha \rhd \gamma)$,

Proof. (a) Assume $s \leqslant t$, uSt and $u \Vdash \alpha \wedge \neg\alpha$. This can happen only if $u = i$. Then $u \Vdash \beta$.

(b) Assume $s \Vdash \alpha \rhd \beta$ and $s \Vdash \beta \rhd \gamma$. Take any $t \geqslant s$ and uSt such that $u \Vdash \alpha$. Then there is $v \geqslant u$ such that vSt and $v \Vdash \beta$. Then there is $w \geqslant v$ such that vSt and $w \Vdash \gamma$. So, $s \Vdash \alpha \rhd \gamma$.

(c) Assume $s \Vdash \alpha \rhd (\beta \rhd \gamma)$. Take any $t \geqslant s$ and uSt such that $u \Vdash \alpha \wedge \beta$. Then there is $v \geqslant u$ such that vSt and $v \Vdash \beta \rhd \gamma$. Since $v \leqslant t$, uSt and $u \Vdash \beta$, it holds that there is $w \geqslant u$ such that wSt and $w \Vdash \gamma$. So, $s \Vdash (\alpha \wedge \beta) \rhd \gamma$.

(d) Similar to (c). Just replace β with α and γ with β.

(e) Assume $s \Vdash \alpha \rhd (\beta \rhd \gamma)$ and $s \Vdash \alpha \rhd \beta$. Take any $t \geqslant s$ and uSt such that $u \Vdash \alpha$. Then there is $v \geqslant u$ such that vSt and $v \Vdash \beta \rhd \gamma$, and there is $w \geqslant u$ such that wSt and $w \Vdash \beta$. Since $t \geqslant v$ and wSt, there is $x \geqslant w$, xSt, and $x \Vdash \gamma$. Since $x \geqslant u$, we have shown that $s \Vdash \alpha \rhd \gamma$. □

Note that the inverse implications of (c) and (d) from the previous proposition do not hold generally.

5.6 Axiomatic characterization of L2

An axiomatization of the logic L2 is formulated in Fig. 5.9. If an L_2-formula α is provable in this system, we write $\vdash_{\mathsf{L2}} \alpha$. This system modifies the system for L1 in that it replaces the axiom A12 and the rule R4 by the axioms A12*, A13* and the rule R4*. Note that in analogy to the case of the modality B, the new principles together encode specification and generalization of the universal propositional quantifier in the following equivalence that is itself not expressible in L_2:

$$\alpha \rhd \beta \leftrightarrow \forall p(E(\alpha \wedge p) \to E(\beta \wedge p)).$$

5.7 Normal models

In this section we are concerned with "normal" info-models, by which we mean info-models in which every trusted source (of any state) is contained in a maximal trusted source (of the same state). We do not have a general philosophical justification of this property (and thus we do not require that it must be generally satisfied) but it is worth noticing that in

Axioms

A1	$\bot \to \alpha$
A2	$\alpha \to (\beta \to \alpha)$
A3	$(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$
A4	$\alpha \to (\alpha \vee \beta)$
A5	$\beta \to (\alpha \vee \beta)$
A6	$(\alpha \to \gamma) \to ((\beta \to \gamma) \to ((\alpha \vee \beta) \to \gamma))$
A7	$\alpha \to (\beta \to (\alpha \wedge \beta))$
A8	$(\alpha \wedge \beta) \to \alpha$
A9	$(\alpha \wedge \beta) \to \beta$
A10	$E\alpha \to \alpha$
A11	$E(\alpha \wedge \neg\alpha) \to \bot$
A12*	$(E\alpha \wedge (\alpha \rhd \beta)) \to E\beta$
A13*	$(\alpha \rhd \beta) \to ((\alpha \wedge \gamma) \rhd (\beta \wedge \gamma))$

Rules

R1	$\alpha, \alpha \to \beta / \beta$
R2	$\alpha \to \neg\beta / \beta \to \neg\alpha$
R3	$\alpha \to \beta / E\alpha \to E\beta$
R4*	$(\alpha \wedge E(\beta \wedge p)) \to E(\gamma \wedge p) / \alpha \to (\beta \rhd \gamma)$
	assuming that p does not occur in α, β and γ

Figure 5.9: A Hilbert style axiomatization of $\mathsf{L2}$

models satisfying this property the semantic clauses for the epistemic operators B and $\rhd$ can be simplified and presented in a more familiar way. In particular, in normal info-models we can define an accessibility relation on states and characterize B as a box modality and $\rhd$ as a strict implication relative to this accessibility relation.

Since not every info-model is normal, a natural question arises whether the logic of all info-models is the same logic as the logic of all normal info-models.[3]

For any state u of any info-frame $\mathcal{F} = \langle A, \sqcap, \sqcup, i, C, S\rangle$, we define the set of its maximal trusted sources:

- $MS(u) = \{v \in A \mid vSu$, and there is no $w \neq v$ such that $v \leqslant w$ and $wSu\}$.

Instead of $t \in MS(u)$, we will also write $tS^m u$. Moreover, we define an "accessibility relation" R_S among states of $\mathcal{F}$:

- $sR_S t$ iff there is $u \geqslant s$ such that $tS^m u$.

We say that $\mathcal{F}$ is *normal* if every trusted source is contained in a maximal trusted source, i.e. the following condition is satisfied for every state u:

- if tSu then $t \leqslant v$ for some $v \in MS(u)$.

An info-model is *normal* if it is based on a normal info-frame. One can observe that in normal info-models E can be equivalently characterized in terms of maximal trusted sources: $s \Vdash E\alpha$ iff there is $tS^m s$ such that $t \Vdash \alpha$. More interestingly, the semantics of B and $\rhd$ can be simplified in the following way:

Proposition 5.7.1. Let s be a state of a normal info-model. Then

(a) $s \Vdash B\alpha$ iff for all t, if $sR_S t$ then $t \Vdash \alpha$,

(b) $s \Vdash \alpha \rhd \beta$ iff for all t, if $sR_S t$ and $t \Vdash \alpha$ then $t \Vdash \beta$.

Proof. To prove (a) assume first that $s \Vdash B\alpha$ and take any $u \geqslant s$ and $tS^m u$. Then there is $v \geqslant t$ such that vSu and $v \Vdash \alpha$. But since t is a maximal trusted source of u we obtain $v = t$. We have proved for all t, if $sR_S t$ then $t \Vdash \alpha$. For the opposite direction, assume that for all t, if $sR_S t$ then $t \Vdash \alpha$. Take any $u \geqslant s$ and vSu. Since the model is normal, v is contained in a maximal trusted source of u which, by the assumption, supports α. We have proved that $s \Vdash B\alpha$. The claim (b) is proved analogously. □

5.8 Concluding remarks

In this paper we have introduced a semantic framework in which models are based on a collection of potentially non-trivially inconsistent information states, and on a "trusted source relation" that selects among the sources of each state those that are consistent

[3] The question is resolved positively in the published manuscript by proving the finite model property.

and regarded as trusted. On the basis of this relation, we have semantically defined two epistemic modalities and a conditional connective, and we have characterized the logic governing these operators.

In future research we would also like to explore various extensions of the current framework. For example it seems natural to introduce some non-persistent operators into the language, e.g. the non-persistent version of B with a semantic clause quantifying only over the trusted sources of the current state (and not over the trusted sources of all its extensions).

We also plan to refine the framework to make the structure of trusted sources more fine-grained. For example, instead of the binary relation S we could consider a ternary relation T. Then, instead of having just tSu (t is a source of u that is trusted) we would have $Tstu$, meaning that s, t are sources of u, and t is more trusted by u than s. It would be interesting to consider epistemic operators that could be defined in terms of such a comparative trusted source relation.

Chapter 6

Taking Up Thagard's Challenge: A Formal Model Of Conceptual Revision

Both belief revision theories and scientific change theories seek to offer rational ways for theory change in light of new evidence, whether it is the subjective belief states of agents or a corpus of scientific data and hypotheses. Despite this very substantial link between belief revision and scientific theory change, there have been few attempts in the literature to bridge the two fields. A recent collection of papers edited by Martin and Osherson [117] can be cited as part of such endeavor, where various authors reinterpret the AGM belief revision as a way of rational hypothesis selection. We aim to contribute to this endeavor by responding to a challenge posed in [149]. Thagard presents a framework for conceptual change in science based on conceptual systems, and he challenges belief revision theorists, claiming that traditional belief-revision systems are able to model only the two most conservative types of changes in his framework, but not the more radical ones. Here we take up Thagard's challenge, presenting a belief-revision-like system that is able to mirror radical types of conceptual change. We propose a conceptual revision system, i.e., a belief-revision-like system that takes conceptual structures as units of revisions. We show how our conceptual revision and contraction operations satisfy analogous of the AGM postulates at the conceptual level and are able to mimic Thagard's radical types of conceptual change.

Besides suggesting similar methodologies for theory change, both belief change and conceptual change can be characterised as ways of *learning*, which take place on different levels. Bridging the two then not only highlights their common features but also opens the way for their combination in the direction of a holistic theory of learning. The approach of the BBR models for structuring information states provides an important advantage as we can model changes on both the belief level and the conceptual level. Specifically, we can model a fine-grained structure for conceptual hierarchies rather than representing them as single, fat sentences. Moreover, the fusion function designed to apply to the information structures can be applied to the conceptual structures, to model a natural dynamics of concepts. Few other works in the literature follow a similar approach of bringing closer belief change and conceptual change by modeling the two in the same fashion, Corina Strosner's recent work on predicate change aims to deal with conceptual change in the fashion of dynamic epistemic logic [144], and Andreas Holger proposes a dynamization of classical structuralism through a synthesis of the structuralist theory of science with the theory of base-generated revisions [8] are some of them.

A holistic approach that connects conceptual revision in particular with hyperintensional belief revision might also help address the logical omniscience problem. Specifically, some aspects of the logical omniscience problem, characterised by the closure conditions C1-C8 in Figure 4.1 can be explained by reference to *concept possession* by reasoning agents. Jago considers that a plausible explanation for the failure of closure under disjunction introduction (from $B\phi$ infer $B(\phi \vee \psi)$) is that it might be unwarranted to ascribe to an agent beliefs that go beyond her *conceptual repertoire* [78, p.16]. Following Yalcin [172], Jago understands concepts as *abilities of a certain kind*, where to possess a concept is to draw certain distinctions in the world [78, p.17]. Now, Yalcin understands concepts as world-partitions, which are identified with subject-matters in the tradition following Lewis and Yablo. His definition, however, fits very well with the conceptual structures introduced below. While concepts and subject-matters come apart both in Jago's proposal and in the view proposed in this dissertation (that is, concepts are not reasoning contexts set by information states), the following might be an initial step for aligning concepts in the sense of Yalcin, with subject-matters understood as reasoning contexts. To summarise, a reasoning approach that aligns a concept possession model with a hyperintensional belief revision model might improve our understanding of aspects of the logical omniscience problem related to concept possession.[1]

The following chapter adds to the flexibility claim of the BBR models by showing a successful application that goes beyond their initial purpose. What is achieved is also interesting for understanding and modeling conceptual change, as we can further investigate a belief-revision-like theory of how to revise inconsistent as well as fragmented (i.e., *patchwork*) concepts with multiple related meanings, while taking advantage of various features of the BBR models.

The content of this chapter is a full collaboration with Matteo De Benedetto and it is based on the paper "Taking Up Thagard's Challenge: A Formal Model of Conceptual Revision", published in Journal of Philosophical Logic [30].

[1] Jago also proposes a direction for combining belief ascription with conceptual competence, based on the exact truth-maker semantics for belief ascription. Accordingly, the conceptual competence of an agent is understood as her ability to identify states of the world that exactly correspond to a certain conceptual content. These are the states that *exactly* decide for an x whether x is F (i.e., a state that is an exact truth-maker for $Fx \vee \neg Fx$) [78, p.18].

6.1 Introduction

Thagard [149] developed a fine-grained cognitivist model of scientific theory change centered around transformations in conceptual systems. Conceptual systems are complex structures similar to frames [113, 57]. They are made of concepts and objects nodes connected via different kinds of links such as kind-links, instance-links, rule-links, and part-links. Changes in science then correspond to different modifications of these links. Specifically, scientific revolutions involve major transformations in part-links and in kind-links inside a conceptual system.

Thagard defended his concept-based model and the autonomy of conceptual change arguing that these revolutionary changes cannot be modeled by belief-revision theories. This supposed impossibility of modeling radical conceptual change within a belief-revision framework has been dubbed *Thagard's challenge* [118]. Specifically, Thagard's challenge claims that strong kinds of conceptual change are irreducible to belief-revision types of changes, because the former involves holistic recombinations of links and nodes in a given conceptual system that cannot be modeled by any piece-meal belief-revision operation. This irreducibility shows for Thagard how frame-based representation of knowledge, despite being expressively equivalent to first-order logic, is procedurally different [146, 147].

Despite the enormous expansion of the belief-revision literature in the last thirty years [68] and recent work connecting it with philosophy of science [117], Thagard's challenge has not received so much attention.[2] The main aim of this work is to suggest a way of taking up Thagard's challenge by developing a belief-revision-like framework capable of modeling the radical types of conceptual change described by Thagard. Specifically, we will present a conceptual revision framework in which we can revise and contract conceptual structures, i.e., set-theoretic representations of Thagard's conceptual systems. Our change operations will be reminiscent of the ones used in base-generated belief change theories [132, 68], but working on conceptual structures instead of belief bases.

Our choice of units of revision, i.e., conceptual structures, makes our system differ from other applications of belief-revision to the problem of scientific change. Traditionally, belief-revision theories deal with piece-meal changes in a belief set similar to the kind of changes happening in normal science (cf. [121]). In applying these theories to the problem of scientific change, logicians have focused on mirroring changes in scientific theories as changes in (usually structured) belief sets [111, 38, 66, 8, 144]. This belief-centered take on scientific change is exactly the reason why Thagard claims that belief revision theories are not adequate for representing conceptual change [147, 148]. We instead chose to model conceptual change at its native level of abstraction, without any reference to the belief level. We achieve this by lifting the methodology of belief revision theories to the conceptual level. As a result, the aim of our change operations will then be the preservation of the consistency of conceptual structures. This consistency is understood as the satisfaction of some structural constraints on the components of a conceptual structure that ensure the

[2] For a couple of exceptions that suggest ways of expanding belief revision to treat certain aspects of scientific conceptual change, and thus could be considered implicit partial replies to Thagard's challenge, see [135] and [67].

overall consistency of the knowledge represented by it. The knowledge represented via our conceptual structures is similar to the content represented by description logics [170], since they also represent knowledge about concept hierarchies.[3]

We will show how eight out of Thagard's nine degrees of conceptual change can be adequately represented in our conceptual revision framework. Specifically, we will mirror all changes that reflect transformations in the structure of conceptual systems, leaving out what Thagard calls *tree switching*, i.e., a more radical kind of change involving a gestalt-like switch in the external interpretation of a given conceptual system. We will demonstrate how each of these eight degrees of conceptual change is mirrored by a specific case of our conceptual revision and contraction operations.

By taking up Thagard's challenge, we intend to show the limitations of belief-centered approaches to scientific change. Our framework shows that, in order to have a satisfactory belief-revision-like account of conceptual change, we have to work at the conceptual level. The extension of traditional belief revision systems into our conceptual revision framework involves rethinking some of the core notions of belief revision such as consistency and completeness. We move from a set-theoretic notion of consistency (and completeness) to a structural one, and we show how these conditions of consistency organise conceptual knowledge. By shifting to the conceptual level we therefore create new opportunities for more general systems of belief revision which can work on different levels of reasoning. Furthermore, using belief revision as a background framework allows us to bridge the logic-oriented approaches to conceptual change with the more cognitive-oriented ones, therefore achieving a logical taxonomy of types of conceptual change.

In Section 6.2, we will present Thagard's account of scientific conceptual change. In Section 6.3, we will present our belief-revision-like model of conceptual revision. More specifically, we will present a revision and a contraction operation that work on conceptual structures. In Section 6.4, we will show how our conceptual revision model satisfies several rationality postulates analogous to the AGM ones for belief revision theories [1]. In Section 6.5, we will demonstrate how our revision and contraction operations are able to mirror several kinds of conceptual changes depicted in Thagard's framework. Finally we will draw some general conclusion on the results and limitations of the present article and we will sketch some directions for future work.

6.2 Thagard's model of scientific conceptual change

Thagard's model of conceptual change in science is built upon the notion of a conceptual system [149]. A conceptual system is a set of nodes interconnected via various kinds of links, a structure that closely resembles frames [113, 57]. More specifically there are two kinds of nodes and four kinds of links that can figure in a conceptual system. Nodes can be concept nodes or object nodes, mirroring respectively concepts and objects. Concept nodes

[3] AGM-style and base-generated revision theories in description logics are also proposed in [130] and in [131].

can be connected with other concept nodes via three kinds of links (kind-links, part-links, rule-links) and with other object nodes via another kind of links, i.e., instance-links:[4]

- Kind-links (from concepts to concepts): intuitive reading 'is a kind of', example 'the canary is a kind of bird'.
- Part-links (from concepts to concepts): intuitive reading 'a whole has a given part', examples 'the beak is a part of birds', 'fins are part of fishes'.
- Rule-links (from concepts to concepts): intuitive reading as expressing generic relations between concepts, example 'canaries are yellow'.
- Instance-links (from objects to concepts): intuitive reading 'is an instance of', example 'Tweety is a canary'.

The most important kinds of links are the ones between conceptual nodes. Kind-links and part-links specify what the constituents of (a part of) the world are according to a given conceptual system. Concepts within conceptual systems are organised in kind-hierarchies and part-hierarchies, i.e., sets of kind-links and part-links that are constrained in a tree-like form in order to give a consistent picture of (a part of) the world. Rule-links instead represent factual information and default reasoning mechanisms codified within the conceptual system. They are not organised in a hierarchy, but they can be divided between weak-rules and strong-rules depending on the strength of the information they represent.

Conceptual changes on a given conceptual system are then ordered by Thagard [149, p.35] in terms of how radical they are, from the least to the most radical:

1. Instance-addition: adding an instance relation saying that a given individual is an instance of a given concept, e.g., 'that blob in the distance is a whale'.
2. Rule-addition: adding a rule relation, e.g., 'whales can be found in the Arctic ocean' or 'whales eat sardines'.[5]
3. Part-addition: adding a new part-relation, e.g., 'whales have spleens'.
4. Kind-addition: adding a new kind-relation, e.g., 'a dolphin is a kind of whale'.
5. Concept-addition: adding new concept, e.g., 'sound-wave' or 'narwhal'.

[4]Note that Thagard in presenting his framework mentions also a fifth kind of link, property-links [149, p.31]. This kind of links is supposed to mirror the information of a given object possessing a given property, but it does not seem to play any role into Thagard's model of conceptual change. It is in fact not mentioned in his abstract presentation of the model [149, p.34-39] nor in any of the case studies [149, p.131-224]. We chose therefore to omit this kind of link from our discussion.

[5]Note that Thagard actually divides the rule-addition kind of conceptual change in two distinctive sub-types: weak-rule and strong-rule addition. Since Thagard's distinction between weak and strong rules is entirely pragmatical [149, p.35], being it based on the problem-solving power of a rule, we collapse in our framework these two types of changes in one.

6. Kind-collapse: collapsing part of a kind-hierarchy, abandoning a previous distinction, e.g., when Darwin collapsed species and varieties within a species distinction.

7. Hierarchy-reorganization: shifting concepts or parts of the kind and part-hierarchies to another part of the hierarchies, i.e., *branch-jumping* such as Darwin's shift of humans to the animal-mammal part of the kind-hierarchy. It may also involve transformation of part-relations onto kind-relations and *vice versa*.

8. Tree-switching: changing the organisational principle of the kind-hierarchy, e.g., Darwin's switch from a morphological kind-hierarchy to an evolutionary one.

The aforementioned Thagard's challenge consists of the claim that belief revision systems can model just the first two degrees of conceptual change, i.e., instance-addition and rule-addition, but not the other six [149, p. 36]. Both instance-addition and rule-addition represent in fact piecemeal additions that do not involve any recombination in the part- and kind-hierarchies of a given conceptual system. These two kind of changes can then be adequately mirrored as changes at the belief-level, revising for instance the extension of a predicate and its prototypical instances [144]. The other six, more radical kinds of conceptual changes are more holistic types of changes, since they involve the adjustment of the part- and kind-hierarchies (as well as rule and instance-relations) of the whole conceptual system. These changes represent in fact how scientists in revolutionary times add new concepts, delete old concepts, drastically reorganise kind and part-hierarchies, and sometimes they even change the organisational principle of the hierarchical tree. Due to their holistic character, these changes cannot be easily mirrored as changes at the belief level like the first two. These revolutionary changes, then, are for Thagard [149, p.28] evidence that conceptual change is irreducible to belief-revision.

6.3 A conceptual revision model

In the previous section we described Thagard's model of scientific conceptual change. In this section, we present a formal model of conceptual revision that is able to model the kind of changes described by Thagard, including the more radical ones. Our system is equipped with a change mechanism similar to the one of base-generated belief revision frameworks, but the units of change are structure mirroring Thagard's conceptual systems rather than belief bases. In this way, we can mirror Thagard's changes at their native level of abstraction, namely the conceptual level.

Our framework takes as its units of changes set-theoretic entities which we call *conceptual structures*. We define two different domains, one for concepts and one for individual objects, as the primary elements of a conceptual structure. Our conceptual structures enrich these two basic domains with different relations between elements of these domains. Mirroring Thagard's system, we define three two-place relations between elements of the concept domain (kind-relation, part-relation, rule-relation) and one two-place relation between elements of the object domain and elements of the concept domain (instance-relation).

Conceptual Structures and Conceptual Hierarchies Formally, a *conceptual structure* is defined as follows:

Definition 6.3.1. $CS = \langle \mathcal{C}, \mathcal{O}, K, P, R, I \rangle$ is a conceptual structure iff,

- $\mathcal{C}$ and $\mathcal{O}$ are (possibly empty) finite domains of (respectively) atomic concepts and individual objects.
- $K = \{\langle x, y \rangle, \dots\}$ and $P = \{\langle x, y \rangle, \dots\}$ are two-place irreflexive relations between elements of the concept domain such that $x, y \in C$ and $\langle x, y \rangle$ is an ordered pair. They represent respectively Thagard's kind and part links between concept nodes. If $\langle x, y \rangle \in K$, we write $x \sqsubset_K y$ (same for $x \sqsubset_P y$, if $\langle x, y \rangle \in P$).
- $R = \{\langle x, y \rangle, \dots\}$, with $x, y \in C$ and $\langle x, y \rangle$ is an ordered pair, is a two-place anti-symmetric relation between elements of the concept domain. It represents Thagard's rule links between concept nodes.
- $I = \{\langle a, x \rangle, \dots\}$ with $a \in O$ and $x \in C$ and $\langle a, x \rangle$ is an ordered pair, is a two-place anti-symmetric relation between elements of the object domain and elements of the concept domain. It represents Thagard's instance links between object and concept nodes.

We can then single-out specific kind-relations and part-relations through a tree-like structural requirement. Relations satisfying this requirement are then called respectively *kind-hierarchies and part-hierarchies.* This requirement is our way of rationally reconstructing Thagard's implicit structural requirements on conceptual systems. Similarly, we introduce criteria to single out certain rule and instance relations as *consistent* rule and instance relations. With these further criteria we mirror common constraints on how knowledge is represented in a consistent way by frames (cf. [6, 57]). Then, a conceptual structure is a *conceptual hierarchy* iff its kind relation is a kind-hierarchy, its part relation is a part-hierarchy, its rule relation is a consistent rule relation, its instance relation is a consistent instance relation, and all the concepts and objects occurring in its relations are members respectively of the concept domain or the object domain.

Definition 6.3.2. $CH = \langle \mathcal{C}, \mathcal{O}, K_h, P_h, R_{cons}, I_{cons} \rangle$ is a conceptual hierarchy iff,

- $\mathcal{C}$ and $\mathcal{O}$ are (possibly empty) finite domains of respectively concepts and objects, which include all the concepts and objects that appear in the relations.
- K_h is a kind-hierarchy, i.e., a transitive kind-relation $K = \{\langle x, y \rangle, \dots\}$ that, if non-empty, has a top element and from any other element of the ordering there exists a unique path to this top element modulo transitivity.
- P_h is a part-hierarchy, i.e., a transitive part-relation $P = \{\langle x, y \rangle, \dots\}$ that, if non-empty, has a top element and from any other element of the ordering there exists a unique path to this top element modulo transitivity.

- R_{cons} is a consistent rule-relation, i.e., a rule-relation $R = \{\langle x, y\rangle, \ldots\}$ such that $\forall x, y, z \in C$ if $\langle x, y\rangle \in R_{cons}$ and $z \sqsubset_K x$, then $\langle z, y\rangle \in R_{cons}$.
- I_{cons} is a consistent instance-relation, i.e., an instance-relation $I = \{\langle a, x\rangle, \ldots\}$ such that $\forall x, y \in C$ and $\forall a \in O$ if $\langle a, x\rangle \in I_{cons}$ and $x \sqsubset_K y$, then $\langle a, y\rangle \in I_{cons}$; $\forall x, y \in C$ and $\forall a \in O$ it holds that $\langle a, x\rangle \in I_{cons}$ and $\langle a, y\rangle \in I_{cons}$ only if $x \sqsubset_K y$ or $y \sqsubset_K x$.

The transitivity requirement says about a kind-hierarchy and a part-hierarchy that $\langle x, y\rangle, \langle y, z\rangle \in K_h(P_h)$ implies that $\langle x, z\rangle \in K_h(P_h)$. We define the *top element in a kind-relation* (*part-relation*) as follows: given a conceptual structure H, the concept domain $\mathcal{C}_H$ and the kind-relation K_H (part-relation P_H) of H, a concept $a \in \mathcal{C}_H$ is a top element in K_H (P_H) iff for all concepts $t \neq a$ in $\mathcal{C}_H$ which occur in a pair in K_H (P_H), it holds that $\langle t, a\rangle \in K_H$ ($\in P_H$). Hence, an ordering on concepts determined by a kind-hierarchy (part-hierarchy) is upward closed. By *a unique path to the top element from any other element modulo transitivity* we mean that, given a is a top element in a kind-relation (or in a part-relation), for all $t \neq a$ which occur in a pair in the kind-relation (part-relation), if $\langle t, y\rangle$ and $\langle t, z\rangle$ are pairs in the kind-relation (part-relation), and $y \neq z$, then either $\langle z, y\rangle$ or $\langle y, z\rangle$ is also a pair in the same relation. In other words, kind-hierarchies and part-hierarchies do not allow upward branchings (Figure 1). The requirements on the consistent rule-relation and the consistent instance-relation say that these relations are transitively closed modulo the relevant kind-hierarchies. We have also required that a consistent instance-relation does not allow upward branching.

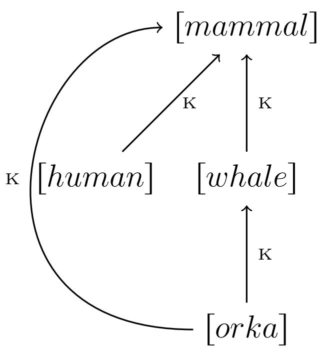

Figure 6.1: A conceptual hierarchy made of four concepts and four kind-links between them.

6.3.1 Revision on conceptual structures

In this section we will describe how a conceptual structure should be revised in our framework. Revising a conceptual structure means adding new elements (of a conceptual structure) to an existing conceptual structure, while preserving (or restoring) the consistency of the revised conceptual structure. The information we want to add (or delete) can be a concept, an object, a kind-relation, a part-relation, a rule-relation, or an instance-relation.

Consistency is characterised by the idea of conceptual hierarchies, i.e., by structural restrictions on the different kinds of relations connecting concepts and objects within a conceptual structure. Therefore, the goal of our conceptual revision framework is to define change operations which preserve these structural limitations.

We start with identifying the form of the eligible arguments for a revision. Suppose we want to revise an existing conceptual structure with an instance (i.e., an instance-addition in Thagard's framework). Suppose we want to add that Bob is an orka ($I\langle Bob, orka\rangle$). If the existing structure does not already contain the concept of an orka and the object Bob, simply adding the instance link would not make sense. Hence we require that the arguments of conceptual revision are formulated as proper conceptual structures. That is, as an argument of a revision, we express the above instance link as a conceptual structure (let us call it C) which consists of the following: $\mathcal{C}_C = \{orka\}, \mathcal{O}_C = \{Bob\}, I_C = \{\langle Bob, orka\rangle\}, K_C = P_C = R_C = \varnothing$.

Next, we have to choose what kind of revision operation we want in our framework. The consistency of the revised (conceptual) structure could for instance be preserved while making the additions, or it could instead be restored after the addition process [67]. The former approach is typical of the AGM belief revision paradigm [1], while the latter is common amongst base-generated revision theories [68, 132]. Our revision operations on conceptual structures will resemble the approach of base-generated revisions. Given a conceptual structure and an argument of revision, we will first add them on top of each other, while restoring the transitivity of the relation (by adding new links) in the resulting conceptual structure. Next, we will retrieve the consistent parts of this structure to build the revised (consistent) conceptual structure.

In base-generated belief revision, one adds the new information to the existing body of beliefs via a set-theoretical union operation.[6] The potential inconsistency in the expanded belief set is caused by too much information. To eliminate this inconsistency, the (less entrenched) beliefs responsible for it are deleted from the new belief set.

In our framework, the inconsistency of a conceptual structure may be caused by too much information or by too little information. In fact, the pivotal requirement of transitivity for the relations of a conceptual hierarchy could be lost during revision. In order to restore the consistency of a conceptual structure, we need to eliminate the inconsistent parts and to repair the transitivity of its relations. We will deal with the transitivity issue in the first step of our revision, i.e., the addition of new information to a conceptual structure (which constitutes the operation of conceptual expansion). In the second step of our revision, we will instead deal with the problem of inconsistent parts, proposing a mechanism that retrieves consistent parts of the expanded conceptual structure.[7]

[6] We significantly simplify the mechanism of base generated revisions. In fact the new information is added on top of a structure called a *belief base* which is the foundation of an agent's beliefs. Moreover, the addition of the new information is set-theoretical only if we are dealing with *flat* belief bases which are not ordered by a preference or entrenchment relation. Full theories of belief revision usually include such orderings to account for the rationality of changing beliefs. Adding beliefs in the AGM paradigm also goes further than the union operation as it involves taking the deductive closure of the new belief set.

[7] While we could keep the expansion process simple and deal with the (possibly lost) transitive closure

Conceptual expansion We will define conceptual expansion as the process of adding new information to a conceptual structure, while restoring the transitivity of the relation in the resulting conceptual structure. More specifically, conceptual expansion will be performed via the fusion models.

Definition 6.3.3. A tuple $CS^{\oplus} = \langle \mathcal{CS}, \oplus \rangle$ is fusion model on conceptual structures iff $\mathcal{CS}$ is a set of conceptual structures that is closed under the total conceptual fusion function $\oplus$ from $CS \times CS$ to CS, uniquely determined by the following:

- $\mathcal{C}_{A \oplus B} = \mathcal{C}_A \cup \mathcal{C}_B$
- $\mathcal{O}_{A \oplus B} = \mathcal{O}_A \cup \mathcal{O}_B$
- $K_{A \oplus B} = TC(K_A \cup K_B)$
- $P_{A \oplus B} = TC(P_A \cup P_B)$
- $R_{A \oplus B} = TC|K_{A \oplus B}(R_A \cup R_B)$
- $I_{A \oplus B} = TC|K_{A \oplus B}(I_A \cup I_B)$

TC stands for the transitive closure operation on our sets of pairs. For instance, the transitive closure of a kind-relation K is the smallest transitive set of pairs that contains K such that if $\langle a, b \rangle$ and $\langle b, c \rangle$ is in $TC(K)$ then $\langle a, c \rangle \in TC\{K\}$. Transitive closure on rule-relations and instance-relation are via transitivity modulo the kind-relation. Thus, an instance-relation I is transitively closed modulo the relevant kind-relation K ($TC|K$) iff given $\langle b, c \rangle \in K$ and $\langle a, b \rangle \in I$, then also $\langle a, c \rangle$ is in I.

The above model specifies how to add a full conceptual structure on top of another one.

We show with an example how conceptual expansion via conceptual fusion models works. Let H be the conceptual hierarchy depicted in Figure 1 such that

$$\mathcal{C}_H = \{[mammal], [human], [whale], [orka]\}, \mathcal{O}_H = P_H = R_H = I_H = \varnothing,$$
$$K_H = \{\langle human, mammal \rangle, \langle whale, mammal \rangle, \langle orka, whale \rangle, \langle orka, mammal \rangle\}.$$

Let A be a conceptual structure consisting of: $\mathcal{C}_A = \{[fish], [orka]\}$, $\mathcal{O}_A = \{(Bob)\}$, $K_A = \{\langle orka, fish \rangle\}$, $I_A = \{\langle Bob, orka \rangle\}$, $P_H = R_H = \varnothing$. Then, by the fusion model, we can obtain the conceptual structure $H \oplus A$ determined by the following elements:

in the process of retrieving information, we believe it is more natural to restore the transitivity required for consistency as part of the expansion operation. One reason is that restoring transitive closure will be done by adding new links, and keeping all additions as part of the expansion and limiting the process of retrieval of information to elimination of some (less entrenched) parts of the expanded structure which contribute to the inconsistency allows simpler definitions for the two processes. Another reason is that this allows us to characterise a conceptual expansion operation which results in structures which resemble conceptual hierarchies to an extent that they are somehow useful in practice.

$$\mathcal{C}_{H\oplus A} = \{[mammal], [human], [whale], [orka], [fish]\}, \mathcal{O}_{H\oplus A} = \{(Bob)\}$$
$$K_{H\oplus A} = \{\langle human, mammal\rangle, \langle whale, mammal\rangle, \langle orka, whale\rangle,$$
$$\langle orka, mammal\rangle, \langle orka, fish\rangle\}, P_{H\oplus A} = R_{H\oplus A} = \varnothing,$$
$$I_{H\oplus A} = \{\langle Bob, orka\rangle, \langle Bob, fish\rangle\langle Bob, whale\rangle, \langle Bob, mammal\rangle\}.$$

The instance pairs $\langle Bob, mammal\rangle$, $\langle Bob, whale\rangle$ and $\langle Bob, fish\rangle$ in $I_{H\oplus A}$ are additions to the simple union of I_H and I_A via the transitive closure operation.

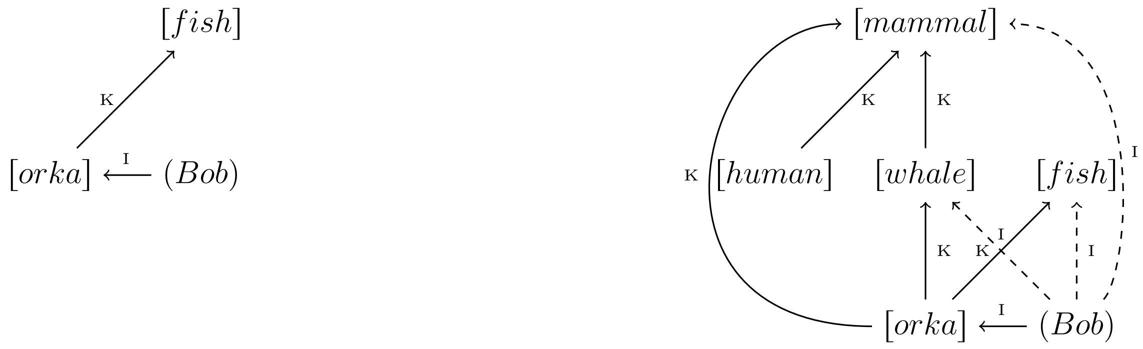

Figure 6.2: The conceptual structure A (on the left) and the conceptual structure $H \oplus A$ (on the right).

Conceptual revision Conceptual expansion may not always produce a conceptual hierarchy. Since our aim is to obtain conceptual hierarchies as the result of revisions, we propose a consistency check mechanism for restoring the consistency of conceptual structures. This mechanism is a modified version of the consolidation operation described in relation to the base generated revisions [132, p.40]. In what follows, we use the notions of a *substructure* and a *maximal hierarchy within a conceptual structure*:

Definition 6.3.4. Given two conceptual structures H, H', we say that H is a substructure of a conceptual structure H' (denoted by $H \subseteq H'$) iff the domains and relations defined in H are subsets of the respective domains and relations defined in H'.

Definition 6.3.5. Given two conceptual structures H, H', we say that H is a maximal hierarchy within H' iff H is a substructure of a conceptual structure H' and any non-trivial expansion of H within H' is not a conceptual hierarchy.

A substructure of a conceptual structure is then a conceptual structure, the components of which (such as objects, concepts, and relations) are subsets of the respective components of the other conceptual structure. For instance, consider the conceptual structure $H \oplus A$ in figure 2 above. The conceptual structure below is a substructure of $H \oplus A$:

$$\mathcal{C} = \{[mammal], [whale], [orka]\}, \mathcal{O} = \{(Bob)\},$$
$$K = \{\langle whale, mammal\rangle, \langle orka, whale\rangle, \langle orka, mammal\rangle\},$$
$$I = \{\langle Bob, orka\rangle, \langle Bob, whale\rangle, \langle Bob, mammal\rangle\}, P = R = \varnothing.$$

A substructure that cannot be (non-trivially) expanded to a conceptual hierarchy, within the given conceptual structure, is a maximal hierarchy. For instance, the above example of a substructure is not a maximal hierarchy, within $H \oplus A$. This is because it can be expanded to a larger hierarchy that includes the concept $[human]$ and the kind link $\langle [human], [mammal]\rangle$. This expanded structure is instead a maximal hierarchy within $H \oplus A$.

While revising a conceptual structure, we first expand it with the argument of the revision. Since it might be the case that the expansion operation fails to produce a conceptual hierarchy, we need a mechanism which marks off the best candidates for the revised conceptual structure. In order to meet some rationality criteria, such selection mechanisms usually rely on an ordering of the alternatives based on the preferences of the selecting agents. Hence we include in our conceptual revision model a preorder among conceptual structures:

Definition 6.3.6. A conceptual revision model is a tuple $CS^{\oplus\leqslant} = \langle CS, \oplus, \leqslant\rangle$such that

- $\langle CS, \oplus\rangle$ is a fusion model on conceptual structures, and
- $\leqslant$ is a connected preorder on a set of conceptual structures.

The preorder between conceptual structures is a preference ordering. If X, Y are conceptual structures in a set, and $X \leqslant_{CS^{\oplus\leqslant}} Y$, we say the conceptual structure X is at least as preferred as the conceptual structure Y given the model $CS^{\oplus\leqslant}$. The most preferred conceptual structures in a set are the ones that are minimal under $\leqslant$, i.e., X is most preferred with respect to $\leqslant$ in a set S of conceptual structures iff for all Y in S, it holds that $X \leqslant_{CS^{\oplus\leqslant}} Y$.

We propose that during the revision operation, the preference ordering on a set of conceptual structures changes as follows: let $\mathcal{CS}$ be a non-empty set of conceptual structures, and $H \in \mathcal{CS}$ the argument of a revision, and let $\leqslant$ be the pre-revision preference ordering on $\mathcal{CS}$ and $\leqslant'$ be the revised preference ordering, then,

- for all $A, B \in \mathcal{CS}$, if $H \subseteq A$ and $H \nsubseteq B$ then $A \leqslant' B$ and $B \nleqslant' A$, and if $H \subseteq B$ and $H \nsubseteq A$ then $B \leqslant' A$ and $A \nleqslant' B$,
- otherwise, $A \leqslant' B$ iff $A \leqslant B$.

In our model, then, we give priority to new data, i.e., the argument of the revision. More specifically, given the revised preference ordering, the parts of the expanded conceptual structure which include the new data are strictly more preferred to the parts which exclude the new data. Apart from this change, the preference ordering remains the same. The revised conceptual structure will then be given by the intersection of the most preferred maximal hierarchies within the expanded conceptual structure, according to the revised preference ordering. The maximality principle concerning these hierarchies is assumed in order to preserve as much information as possible while revising.

A preference ordering may, in fact, rate multiple conceptual hierarchies as the best ones. In the belief revision literature, these cases are commonly solved by taking the intersection of the selected alternatives, following the *partial meet contraction and revision* operations introduced within the AGM paradigm. However, as we will show with an example, intersecting multiple conceptual hierarchies may generate inconsistent conceptual structures. As a solution to this problem, we propose a repetitive revision operation, where the intersection mechanism is repeated until a conceptual hierarchy is obtained.

Since our revision operation involves changing the preference order in a revision model, it is essentially a model-changing operation. Therefore, even if the expansion of the initial conceptual structure with the argument of the revision is a conceptual hierarchy, the revision operation does not reduce to expansion, since changes on the preference ordering are significant for iterating any change operation on the new conceptual structures.

Before moving on to a simple example of a conceptual revision, let us say more about the preference ordering for conceptual structures that we assume in our model. In the scientific context, there are many different factors that might be considered for determining such a preference ordering. One could, for instance, establish a preference ordering based on corroboration, where more empirically confirmed parts of a conceptual structure are preferred (cf. Hansson's scientific corpus model, [67]). Alternatively, another option could be to prioritize parts of a conceptual structure involving empirical concepts in comparison to parts involving theoretical terms, obtaining a sort of empiricist conceptual revision. It could also be possible to held a specific part of a conceptual structure as the most preferred, identifying it with the constitutive part of a scientific theory à la Friedman [55] or with what Lakatos called the hard-core of a scientific research program (cf. [89]). Moreover, a suitable preference ordering could be determined virtually by any specific bundles of theoretical values discussed in debates over theory choice in science, such as simplicity, fruitfulness, empirical adequacy, and the like. In the related literature in philosophy of science, one finds contrasting bundles of epistemic values defended (e.g., [152, 105]).

Given this plethora of possible criteria for determining a preference ordering for our conceptual structures, we decided here to stay neutral on which specific criteria we prefer. Consistently, we will assume an arbitrary preference ordering for each set of conceptual structure prior to revision and focus on how such an ordering change when the conceptual structure is revised.

Since we are dealing with revolutionary changes in particular, an important assumption of our model is, as we have already mentioned, to give priority to new data. In revising a given conceptual structure with some new information, our framework will re-order the preference ordering by giving priority to the new information, i.e., to the argument of the revision. This is, of course, not the only viable option. In the related belief revision literature, one could find several preference re-ordering operations, ranging from extremely conservative options to truly radical ones (cf. [133]).

Now, in the scientific context, just like we found many different possible factors for determining a preference ordering, we can find several, often contrasting, factors for choosing a specific way of re-ordering the preference ordering of our conceptual structures. Our choice of giving priority to new information could be, for instance, generally supported by a Popperian confutationalist stance (cf. [123]) or by the general idea that conservativity is an enemy of scientific progress. Several scholars, including Thagard [151], have in fact argued that the principle of minimal change should be abandoned in the scientific context in favour of a more radical revision that prioritises new information. However, a different picture of scientific activity and rationality could favour a radically different preference re-ordering operation. A conservative operation, prioritising the old parts of a conceptual structure, could perhaps be preferred by philosophers of science influenced by Kuhn [86] and Lakatos [89], who believed that mature science always includes some degree of conservativity. Various finer-grained distinctions on this re-ordering could also be made, such as restricting this conservativity to a specific part (e.g., the hard-core of a theory) of the conceptual structure or to specific concepts (e.g., the empirical ones).

In general, the choice of the preference re-ordering operation, just like the choice of the criteria determining such a preference ordering, seems extremely dependent on the specific view of scientific dynamics that one favours. We chose here to prioritise novel information, because this seems to us the choice closer to the spirit of Thagard's framework.[8] That said, it should be clear to the reader that this is by no means the only interesting way in which the preference ordering of a conceptual structure can be re-ordered.

To see how the revision operation is applied to conceptual structures, consider the conceptual structures H and A in the above example. Suppose we want to revise H with A instead of simply expanding the former with the latter. The first step of the conceptual revision process consists of the aforementioned expansion $H \oplus A$ (depicted in Figure 2). Note that $H \oplus A$ is not a conceptual hierarchy. This is because, its kind-relation is not a kind-hierarchy and its instance-relation is not a consistent instance-relation. The kind-relation of $H \oplus A$ is not a kind-hierarchy because: i) although it is non-empty, it does not have a top element, since the concept $[mammal]$ and the concept $[fish]$ are the top elements of two distinct conceptual substructures; ii) the kind-links $\langle orka, whale\rangle$ and $\langle orka, fish\rangle$, both of which occur in the kind relation of the structure, generate an upward branching. The instance-relation of $H \oplus A$ is not a consistent instance-relation because the pairs of instance-links $\langle Bob, whale\rangle$, $\langle Bob, fish\rangle$ and $\langle Bob, mammal\rangle$, $\langle Bob, fish\rangle$ generate upward branchings.

Since $H \oplus A$ is not a conceptual hierarchy, we continue the revision operation by identifying and intersecting the best conceptual hierarchies within $H \oplus A$, determined by the

[8] We should note that technically Thagard's framework does not involve any preference ordering. Although, it has an evaluation mechanism between different conceptual systems based on the notion of explanatory coherence [150]. This mechanism is, however, used only to compare fully finished conceptual structures after the changes have taken place. We leave the study of the relations between Thagard's evaluation mechanism and our preference ordering for future work.

revised preference ordering. As the new data in this revision is the conceptual structure A, after the revision, the conceptual hierarchies of which A is a substructure are strictly more preferred over the ones which exclude some part of A. Then, an easy way to identify the best maximal hierarchies within $H \oplus A$ is to start with A as the base structure and expand it within $H \oplus A$ until reaching a maximal conceptual hierarchy. However, it might be the case that there are no maximal hierarchies within the expanded structure that include the new data. Then, one identifies all maximal hierarchies within the expanded structure; their preference ordering is based on the 'otherwise' clause in our definition above.

Once we accept A within $H \oplus A$, we can only expand it with the instance link $\langle Bob, fish \rangle$. This is in particular due to accepting that orkas are fishes. This information is inconsistent with the information that orkas are whales, and that orkas are mammals, since fishes are neither whales nor mammals given the expanded structure. The part of $H \oplus A$ concerning the kind-links $\langle human, mammal \rangle$ and $\langle whale, mammal \rangle$ does not directly contradict with the information given in A. However, accepting also this part would mean that the resulting structure does not have a top element since fishes are not mammals, and mammals are not fishes. Consequently, the following is the unique conceptual hierarchy (M (Figure 3)) within $H \oplus A$ which fits the description: $\mathcal{C}_M = \{[fish], [orka]\}$, $\mathcal{O}_M = \{(Bob)\}$, $K_M = \{\langle orka, fish \rangle\}$, $I_M = \{\langle Bob, orka \rangle, \langle Bob, fish \rangle\}$, $P_M = R_M = \varnothing$. The revision of H with A finalises here.[9]

$$\begin{tikzcd}
 & {[fish]} \\
{[orka]} \arrow[ur, "\mathrm{K}"] & (Bob) \arrow[u, "\mathrm{I}"'] \arrow[l, "\mathrm{I}"']
\end{tikzcd}$$

Figure 6.3: The conceptual hierarchy M.

In the case that there are more than one maximal hierarchies in the expanded structure which includes the argument of the revision, we pick up the most preferred among these, and intersect them. It might be the case that the first iteration of this last step do not result in a conceptual hierarchy. In that case, we repeat by intersecting the most preferred maximal hierarchies within the conceptual structure we have obtained as the result of the latest iteration. This is done until one reaches a conceptual hierarchy. Our next example

[9] This example of conceptual revision reveals a significant amount of information loss as a result. This is connected to the revolutionary aspect of the scientific changes we want to represent. As it was famously stressed by Kuhn [86], scientific revolutions involve often the loss of information in the transition from one scientific theory to its successor, a phenomenon commonly known in philosophy of science as Kuhnian loss. A more conservative conceptual revision may also be defined in our framework. We could, for instance, weaken our requirements for conceptual hierarchies, eliminating the necessity of having a top element while maintaining the ban of upward-branching. This change would allow a conceptual hierarchy to consist of several disconnected conceptual hierarchies. This alternative definition would make the revised conceptual structure in the above example to consist of the original result M and the following conceptual hierarchy: $\mathcal{C} = \{[human], [mammal], [whale]\}$, $K = \{\langle human, mammal \rangle, \langle whale, mammal \rangle\}$, $\mathcal{O} = P = R = I = \varnothing$.

shows how the revision operation is applied repetitively.

Suppose we want to revise a conceptual structure X such that $\mathcal{C}_X = \{a, b, A, B\}$, $K_X = \{\langle a, A\rangle, \langle b, A\rangle, \langle a, B\rangle, \langle b, B\rangle, \langle B, A\rangle, \langle A, B\rangle, \}$ and $\mathcal{O}_X = P_X = R_X = I_X = \varnothing$ with the empty conceptual structure $\{\varnothing\}$. There are two maximal hierarchies within the expanded conceptual structure $X \oplus \{\varnothing\}$ whose kind relations are the following, $K_{X1} = \{\langle a, A\rangle, \langle b, A\rangle, \langle a, B\rangle, \langle b, B\rangle, \langle B, A\rangle\}$ and $K_{X2} = \{\langle a, A\rangle, \langle b, A\rangle, \langle a, B\rangle, \langle b, B\rangle, \langle A, B\rangle\}$. Since the empty conceptual structure is a substructure of both, the preference ordering remans untouched and therefore the two conceptual hierarchies are equally preferred. Then, their intersection includes a kind-relation which does not have a top element, i.e., $K_{X1 \cap X2} = \{\langle a, A\rangle, \langle b, A\rangle, \langle a, B\rangle, \langle b, B\rangle\}$, that cannot be the kind-relation of a conceptual hierarchy. We therefore repeat the revision operation, first determining the best maximal hierarchies within the conceptual structure $X1 \cap X2$. These are the hierarchies with the kind-relations $K_Y = \{\langle a, A\rangle, \langle b, A\rangle\}$ and $K_Z = \{\langle a, B\rangle, \langle b, B\rangle\}$. If they are preferred equally, then the revised conceptual hierarchy has in its concept domain only the concepts a and b, together with an empty kind-relation

We can then define our conceptual revision operation as follows:

Definition 6.3.7. Given a conceptual revision model $CS^{\oplus \leqslant} = \langle CS, \oplus, \leqslant\rangle$, and H, A conceptual structures, H revised with A (let us denote it with $H * A$, and the preference ordering after revision with $\leqslant'$) is determined by the following cases:

1. $H * A \equiv H' = \cap\{B : B$ is a maximal hierarchy within $H \oplus A$ and for all maximal hierarchies $I \subseteq H \oplus A$ it holds that $B \leqslant' I\}$, if H' constitutes a conceptual hierarchy;

2. $H * A \equiv H'' = \cap\{C : C$ is a maximal hierarchy within H' and for all maximal hierarchies $I \subseteq H'$ it holds that $C \leqslant' I\}$, if H' does not constitute a conceptual hierarchy and H'' constitutes a conceptual hierarchy;

3. repeat case 2 with the maximal hierarchies within the resulting conceptual structure (e.g., starting with H'' in the first repetition) until reaching a conceptual hierarchy as the result of the intersection, if otherwise.

6.3.2 Contraction on conceptual structures

Contracting a conceptual structure means eliminating a part of it. While our contraction operation is defined based on conceptual revision, it differs from revision significantly in terms of how the argument of a contraction should be formulated or expressed. Suppose we want to contract an instance-pair $\langle x, y\rangle$ from a conceptual structure. In regard to the arguments of revision, we required that they are formulated as conceptual structures. An analogous way of formulating the argument of contraction would be the following: $C = \{y\}, O = \{x\}, I = \{\langle x, y\rangle\}, K = P = R = \varnothing$. However, it is not (always) necessary to eliminate the concept and the object in order to eliminate the instance-link. Hence, a well-formed argument for our contraction operation does not have the limitations we proposed for revision. An argument of contraction can be a part of a conceptual structure as well.

In order to formalize conceptual contraction, we introduce in our revision models a set-theoretical elimination operation $\ominus$:

Definition 6.3.8. Given that A and B are conceptual structures, $A \ominus B = A/B$such that

- $\mathcal{C}_{A\ominus B} = \mathcal{C}_A/\mathcal{C}_B$
- $\mathcal{O}_{A\ominus B} = \mathcal{O}_A/\mathcal{O}_B$
- $K_{A\ominus B} = K_A/K_B$
- $P_{A\ominus B} = P_A/P_B$
- $R_{A\ominus B} = R_A/R_B$
- $I_{A\ominus B} = I_A/I_B$

Our elimination operation simply eliminates B from A. For simplicity, we define our elimination and contraction operations on conceptual structures. When the argument of the contraction is not a (complete) conceptual structure, the excluded elements of a conceptual structure will be regarded as empty. For instance, if an argument of a conceptual contraction A consists solely of an instance relation, then K_A, P_A, R_A, C_A and O_A are taken to be empty. Note that the elimination operation does not include taking the transitive closures of the resulting relations (as opposed to the expansion operation). Our contraction operation is defined as follows:

Definition 6.3.9. A conceptual revision and contraction model is a tuple $CS^{\oplus\ominus\leqslant} = \langle \mathcal{CS}, \oplus, \ominus, \leqslant \rangle$such that

- $\langle \mathcal{CS}, \oplus \rangle$ is a fusion model on conceptual structures,
- $\ominus$ is the set-theoretical elimination operation on conceptual structures, and
- $\leqslant$ is a connected preorder on a set of conceptual structures.

We propose that during the contraction operation, the preference ordering on a set of conceptual structures changes as follows: let $\mathcal{CS}$ be a non-empty set of conceptual structures, and $H \in \mathcal{CS}$ the argument of a contraction, and let $\leqslant$ be the pre-contraction preference ordering on $\mathcal{CS}$ and $\leqslant'$ be the new preference ordering, then,

- for all $A, B \in \mathcal{CS}$, if $H \nsubseteq A$ and $H \subseteq B$ then $A \leqslant' B$ and $B \nleqslant' A$, and if $H \nsubseteq B$ and $H \subseteq A$ then $B \leqslant' A$ and $A \nleqslant' B$,
- otherwise, $A \leqslant' B$ iff $A \leqslant B$.

Note that, after the elimination operation, the remaining part of the conceptual structure does not have any substructures that include the argument of the contraction. Thus, when we focus on the relevant portion of the new preference ordering, that is the portion which orders the conceptual structures within the eliminated conceptual structure, the new preference ordering looks identical to the initial preference ordering, due to the second clause of our description.

Our contraction operation is quite radical, since we not only decrease the preference ordering of the conceptual structures that include the argument of the contraction, but we eliminate the argument from the new conceptual structures irrevocably. Just like in the revision case, this specific contraction operation is of course not the only possible one. In the belief revision literature, we can find more conservative ways of contracting information, such as operations that abstain from irrevocable elimination and give priority to parts of a structure that do not include or imply the argument of the contraction over the parts that do. One could also imagine more radical ways of reordering the preferences by a contraction operation, such as requiring that the parts of a structure that are consistent with the contracted information are at most as preferred as the the parts which are not consistent with it.[10] The choice between these different contraction operations appears, analogously to the one between different revisions operations, strongly dependent on the specific views on scientific rationality that one favours. It is easy, in fact, to envisage philosophical criteria that would justify any of the alternatives above. Thus, we decided to stay as neutral as possible here, choosing the contraction operation most similar to our revision operation.

As in the case of revision, in our framework the outcome of a contraction operation on a conceptual structure ought to be a conceptual hierarchy. It should also be the case that contraction operation do not expand the contracted structures with novel relations, concepts or objects.[11] As we will see, even if nothing is added to a conceptual structure through contraction, the hierarchical structure may be lost. For instance, contracting a structure with respect to a kind-link may affect the transitivity of the kind-relation hence breaking the hierarchical structure. We restore the consistency of contracted conceptual structures as we did for revised structures. That is, we pick up the most preferred maximal conceptual hierarchies within the eliminated conceptual structure, according to the new preference ordering, and apply the intersection mechanism, just as described for revisions, until we obtain a conceptual hierarchy.

As an example of a contraction, consider the conceptual hierarchy H' such that

[10] To see the counterparts of these contraction operations, compare for instance, moderate contraction, conservative contraction and severe withdrawal in [133].

[11] This is another reason in favour of keeping the operation of adding relations or links to recover transitivity as part of the conceptual expansion. This way, we can use the exact process defined for revisions in order to retain consistency after conceptual contraction without making any additions to the conceptual structure.

$$\mathcal{C}_{H'} = \{[mammal], [whale], [orka], [narhwale]\},\ \mathcal{O}_{H'} = I_{H'} = P_{H'} = R_{H'} = \varnothing$$
$$K_{H'} = \{\langle whale, mammal\rangle, \langle orka, mammal\rangle, \langle orka, whale\rangle,$$
$$\langle narwhale, whale\rangle, \langle narwhale, mammal\rangle\}.$$

Consider also the part of a conceptual structure A': $\mathcal{C}_{A'} = \{[orka], [narwhale]\}$, $K_{A'} = \{\langle narhwale, whale\rangle$, $\langle orka, whale\rangle, \langle orka, mammal\rangle, \langle narwhale, mammal\rangle\}$. Suppose we want to contract H' with respect to A'. We start with the simple elimination of A' from H', obtaining $H' \ominus A'$such that

$$\mathcal{C}_{H' \ominus A'} = \{[mammal], [whale]\}, K_{H' \ominus A'} = \{\langle whale, mammal\rangle\},$$
$$\mathcal{O}_{H' \ominus A'} = R_{H' \ominus A'} = P_{H' \ominus A'} = I_{H' \ominus A'} = \varnothing.$$

The output of this particular contraction $H' - A'$ (Figure 4) is equal to $H' \ominus A'$, aside from the changes in the preference ordering. This is because, the latter is a conceptual hierarchy, and the relevant portion of the new preference ordering is determined completely by the *otherwise* clause in our description of preference reordering by the contraction operation. If it were the case that $H' - A'$ is not a conceptual hierarchy, the consistency of the resulting conceptual structure would be recovered by iteratively intersecting the most preferred maximal conceptual hierarchies within $H' \ominus A'$, according to the new preference ordering, until reaching a conceptual hierarchy.

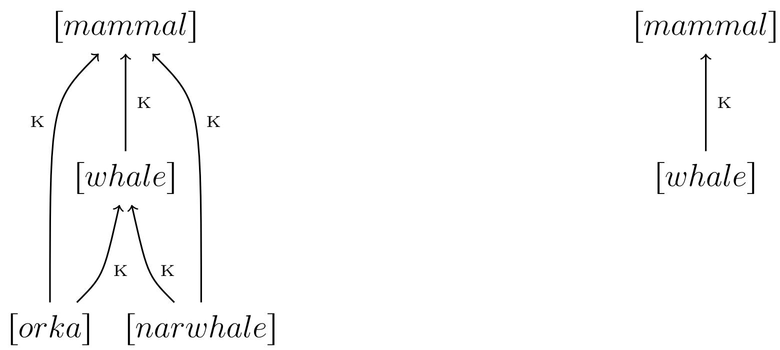

Figure 6.4: The conceptual hierarchy H' (on the left) and the conceptual hierarchy $H' - A'$ (on the right).

Definition 6.3.10. Given a conceptual revision and contraction model $CS^{\oplus\ominus\leqslant}$, and given H is a conceptual structure and A is a (part of a) conceptual structure, H contracted with A (let us denote it with $H - A$, and the preference ordering after contraction with $\leqslant'$) is determined by the following cases:

1. $H - A \equiv H' = \cap\{B : B$ is a maximal hierarchy within $H \ominus A$ and for all maximal hierarchies $I \subseteq H \ominus A$ it holds that $B \leqslant' I\}$, if H' constitutes a conceptual hierarchy;

2. $H - A \equiv H'' = \cap\{C : C$ is a maximal hierarchy within H' and for all maximal hierarchies $I \subseteq H'$ it holds that $C \leqslant' I\}$, if H' does not constitute a conceptual hierarchy and H'' constitutes a conceptual hierarchy;

3. repeat case 2 with the maximal hierarchies within the resulting conceptual structure (e.g., starting with H'' in the first repetition) until reaching a conceptual hierarchy as the result of the intersection, if otherwise.

6.4 Rationality postulates for conceptual change

In this section we will show how our conceptual revision models satisfy several rationality postulates analogous to the AGM ones for belief revision [1]. Since our system works at the conceptual level of abstraction, we cannot straightforwardly apply the AGM postulates to it. Thus, for each AGM revision postulate we will try to develop an analogous postulate at the conceptual level. We will also discuss rationality postulates for conceptual contraction, trying to comprehend the counterparts of the conceptual revision ones.

First, we show that a conceptual counterpart of the AGM closure and consistency postulates for revision is satisfied in our framework.[12] We will call this first conceptual revision postulate the *hierarchy postulate*. This postulate amounts to the claim that a conceptual revision operation always results in a conceptual hierarchy. Recall in fact that for a conceptual structure to be a conceptual hierarchy, the information represented by the relations of the structure should not be contradictory. A conceptual hierarchy is furthermore closed, in the sense that none of the links needed for the transitive closures of the relations is missing. Hence, in our framework the consistency of a conceptual structure is intertwined with its completeness. Our framework satisfies this postulate thanks to the conjunction of the following properties: all conceptual structures have at least one maximal conceptual hierarchy as their substructure (due to their finiteness), the preference ordering always yields some minimal (most preferred) conceptual hierarchy (due to its connectedness), and a conceptual hierarchy can always be reached in finitely many iterations of our revision operation.

Theorem 6.4.1. For all conceptual structures H, A, the product of revising H with A $(H * A)$ is a conceptual hierarchy.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. We need to show that, for all conceptual structures H, A, after expanding H with A $(H \oplus A)$ and reordering the preferences, we can obtain a conceptual hierarchy in finitely many steps, based on the definition of conceptual revision operation. To do this, we show that given an arbitrary conceptual structure, the operation of intersecting the most preferred maximal conceptual hierarchies within H (call this operation $\cap^*$ and the result of it on H with

[12]Our consistency claim is stronger than what is required by the AGM consistency postulate. The AGM postulate assumes that the new information is not a contradiction.

$f(H)$) yields a conceptual hierarchy in finitely many iterations on H (that is, performing $\cap^*$ on the resulting conceptual structure $f(H)$, and on $f(f(H))$, etc.).

(1) We start with showing that for all conceptual structures H, there is an $A \subseteq H$ s.t A is a maximal conceptual hierarchy within H. To show this, the following suffices:

i) The empty conceptual structure ($\varnothing$) is a conceptual hierarchy. This directly follows from the definition of a conceptual hierarchy.

ii) For all conceptual structures H, the empty conceptual structure is a substructure of H ($\varnothing \subseteq H$). This directly follows from the definition of a substructure.

iii) For all conceptual structures H, it holds that H is finite. To show this, we define the size of a conceptual structure as follows: $\#H = \#C_H + \#O_H + \#K_H + \#P_H + \#R_H + \#I_H$ where $\#C_H$ and $\#O_H$ are to the number of elements in C_H and in O_H respectively. $\#K_H, \#P_H$,
$\#R_H, \#I_H$ are the number of pairs in respective relations of H (i.e, we equate the size of a conceptual structure with the sum of the cardinalities of its elements). Since C_H and O_H are assumed to be finite by the definition of a conceptual structure, and since all relations in a conceptual structure are products of these domains and hence finite themselves, it follows that all conceptual structures are finite.

(2) Next we show that for all conceptual structures H, there is an $A \subseteq H$such that A is a maximal conceptual hierarchy within H and A is at least as preferred as all $B \subseteq H$, where B is a maximal conceptual hierarchy within H. It follows from (1) above that for all conceptual structures H, there are finitely many maximal conceptual hierarchies within H, and there is at least one maximal conceptual hierarchy within H. By the model assumptions for the preference ordering, namely by its connectedness, it holds that for all $A, B \subseteq H$ with A, B maximal conceptual hierarchies within H, it holds that either $A \leqslant B$ or $B \leqslant A$. These two facts establish our claim.

(3) Lastly, we show that for all conceptual structures H, finitely many iteration of the operation $\cap^*$ can be performed on H before reaching a conceptual hierarchy as the product ($f(H)$). To achieve this, we refer to the description of size of a conceptual structure from point (1) above, and prove the following:

iv) If S is the empty conceptual structure ($\varnothing$), f(S)=S=$\varnothing$. Since $\varnothing$ is a conceptual hierarchy, $\cap^*$ can be performed exactly $n = 1$ times before reaching a conceptual structure.

v) If S is a non-empty conceptual hierarchy, $f(S) = S$. This is because S is the unique maximal hierarchy within S (for all $A \subseteq S$ such that A is a hierarchy and $A \neq S$, it is possible to expand A into S, hence A is not maximal). Hence, $\cap^*$ can be performed exactly $n = 1$ times before reaching a conceptual structure.

vi) If S is a conceptual structure, that is not a hierarchy, it holds that $f(S) \subset S$, hence $\#f(S) < \#S$. Since $\#S$ is finite, we can iterate the operation $\cap^*$ on S at most $n = \#S$ times before reaching a conceptual hierarchy within S. To see this, let A, B the best preferred maximal conceptual hierarchies within S. Given that S is not a conceptual hierarchy, it holds that $A, B \subset S$. Then, $A \cap B \subset S$ by the definition of substructure. It follows that $\#(A \cap B) < \#S$. Since S is arbitrary, $\#f(S) < \#S$ holds for all such S.

vii) Since $\#S$ is finite, in $n \leqslant \#S$ iterations of $\cap^*$ on S, we obtain a conceptual hierarchy, in the limit case we obtain the empty conceptual structure, with $\#\varnothing = 0$, which is a conceptual hierarchy.

Combining what is established above with the definition of our conceptual revision operation, we conclude that the product of a conceptual revision is always a conceptual hierarchy.

□

Next, we show that our framework satisfies a *success postulate*, i.e., the claim that if the argument of a conceptual revision is a conceptual hierarchy, the argument becomes a substructure of the revised conceptual structure. This postulate corresponds to a weakened version of the AGM success postulate for revisions.[13] For the satisfaction of this postulate, it suffices that the argument of the revision is among the minimal conceptual structures in the (revised) preference ordering. This is achieved since our revision mechanism involves exactly this reordering of the preferences when revising a conceptual structure.

Theorem 6.4.2. For all conceptual structures H and for all conceptual hierarchies A, $A \subseteq H * A$.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. Let H be conceptual structure and let A be a conceptual hierarchy. Suppose we revise H with A $(H * A)$. By the definition of expansion, $A \subseteq H \oplus A$. Let H' be the conceptual structure $H \oplus A$ with the revised preference ordering in accordance with revision of H with A. It follows via (1) above, that there is a maximal conceptual hierarchy A' within H', and since A is a conceptual hierarchy, it holds that $A \subseteq A'$. It follows from the description of reordering the preferences by the conceptual revision operation, that for all $B \subseteq H'$ such that B is a maximal hierarchy within H', it holds that $B \leqslant' A'$ (after the revision) only if $A \subseteq B$. Thus, for all $B \subseteq H'$ that are picked up for intersecting, it holds that $A \subseteq B$. It follows that, $A \subseteq \cap\{B : B$ is a maximal hierarchy within H' and for all maximal hierarchies $I \subseteq H'$ it holds that $B \leqslant' I\}$.

[13]The AGM success postulate requires inclusion of the new belief without an antecedent that says it is a consistent belief. On the other hand, the success postulate required for base-generated beliefs by Rott [132] and Hansson [68] has that antecedent. We consider the weaker version of this postulate due to the strong consistency claim we established. Otherwise we have a contradiction saying the result of a conceptual structure is always consistent and if we revise a conceptual structure with a contradiction, the contradiction is part of the revised structure.

Similarly, for all $n \geqslant 1$ and for all B' such that $B' = f^n(H')$ (i.e., B' is the result of n-times iterating $\cap^*$ on H'), it holds that $A \subseteq f^n(H')$. Since A itself is a conceptual hierarchy, it is not possible that $f^n(H') \subset A$. Therefore, $A \subseteq H * A$.

□

The third rationality postulate we consider is the *vacuity postulate*, i.e., the requirement that if the expansion of a conceptual structure is a conceptual hierarchy, this expansion is equal to the output of the revision process, aside from the reordering of the preference relation. This requirement corresponds to the vacuity postulate in the AGM theory and it is satisfied by our framework.[14]

Theorem 6.4.3. For all conceptual structures H, A, it holds that $H * A \subseteq H \oplus A$ and $H \oplus A \subseteq H * A$ given that $H \oplus A$ is a conceptual hierarchy.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. Let H, A be conceptual structures, and suppose $H \oplus A$ is a conceptual hierarchy. Let $(H \oplus A)'$ be $H \oplus A$ with the revised preference ordering. It holds that $H * A = f^n((H \oplus A)')$, i.e, $H * A$ is the result of n-times iterating $\cap^*$ on $(H \oplus A)'$. By (3) in theorem 6.4.1, it holds that $f(S) \subseteq S$ for all conceptual structures S. Therefore, $H * A \subseteq (H \oplus A)'$. Since $H \oplus A$ and $(H \oplus A)'$ are equivalent besides the preference ordering, it holds that $H * A \subseteq H \oplus A$. Note that this holds even when $H \oplus A$ is not a conceptual hierarchy.

The other direction holds since $H \oplus A$ is the unique maximal hierarchy within itself, and $f((H \oplus A)') = (H \oplus A)' = H * A$ as shown in (v) above. Since $H \oplus A$ and $(H \oplus A)'$ are equivalent besides the preference ordering, it holds that $H \oplus A \subseteq H * A$ □

Next, we consider the *inclusion postulate*, i.e., the requirement that the outcome of a conceptual revision is a substructure of the expansion of the original conceptual structure with the argument of the revision. This postulate corresponds to the AGM inclusion postulate for revisions. This requirement makes sure that a conceptual structure is not expanded further than what is needed to consistently include the argument of the revision. This postulate is satisfied in our framework since all the steps of our revision operation involve only substructures of the expanded conceptual structure.[15]

Theorem 6.4.4. For all conceptual structures H, A, it holds that $H * A \subseteq H \oplus A$.

[14] In the AGM theory, one initially starts with a belief state and a preference ordering, yet as the result of revision or contraction, obtains a new belief set. The result of the change operations do not include a preference ordering. For this reason, while constructing our version, we state that the equality holds, aside from the changes in the preference ordering. The same reasoning applies to the vacuity postulate for contractions and to the recovery postulate, both stated below.

[15] It should be noted that there are three other basic AGM rationality postulates we did not discuss here. One is the extensionality postulate which states that revision of a belief set with classically logically equivalent arguments lead to logically equivalent revised belief sets. Since we did not comment on identity principles concerning the conceptual structures, we cannot map this requirement to our framework for now. The other two postulates are about revisions with conjunctions. We do not consider these as relevant for our current conceptual revision framework, since we did not discuss relations between structures which would correspond to logical connectives.

Proof. See the proof for Vacuity above. Note, however, the result of the operation $H \oplus A$ is different from the basic union of H and A ($H \cup A$) since conceptual expansion fixes transitivity of the relations. It does not hold that $H * A \subseteq H \cup A$. $\square$

After we mapped and analyzed some rationality postulates for conceptual revision framework, let us briefly discuss the corresponding contraction postulates. The first conceptual contraction postulate requires the result of a conceptual contraction to be a conceptual hierarchy. Since conceptual contraction involves the same consistency-recovery mechanism of conceptual revision, this principle is satisfied for reasons analogous to the revision case.

Theorem 6.4.5. For all conceptual structures H and A (A might also be a part of a conceptual structure), the product of contracting H with A ($H \ominus A$) is a conceptual hierarchy.

Proof. We need to show that, for all conceptual structures H, after eliminating a conceptual structure (or a part of a conceptual structure) A from H ($H - A$), we can obtain a conceptual hierarchy in finitely many steps, based on the definition of conceptual contraction operation. The proof similar to that of theorem 6.4.1. $\square$

The success postulate for conceptual contraction requires the argument of the contraction (a conceptual structure or a part of one) to not be a substructure of the the result of the contraction. A weaker version of this principle, which limits the argument of the contraction to non-empty conceptual structures or their parts, is satisfied in our framework. This is because, once the argument of the contraction is deleted from the initial conceptual structure, nothing is added to the resulting structure while rebuilding consistency.

Theorem 6.4.6. For all conceptual structures H and for all non-empty conceptual structures A (A might also be a part of a conceptual structure), it holds that $A \nsubseteq H \ominus A$.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. Let H be a conceptual structure and let A be a non-empty (part of a) conceptual structure. By definition of conceptual elimination we know that $A \nsubseteq H - A$. By the definition of contraction, we have $H \ominus A = f^n(H - A)$, according to the new preference ordering. By (3) in theorem 6.4.1, it holds for all conceptual structures S, $f(S) \subseteq S$, hence $H \ominus A \subseteq H - A$. It follows that $A \nsubseteq H \ominus A$. $\square$

We can also show that the operation of contracting a conceptual structure does not expand the initial conceptual structure in any way. This is a counterpart of the AGM inclusion postulate for contractions.

Theorem 6.4.7. For all conceptual structures H and A (A might also be a part of a conceptual structure), it holds that $H \ominus A \subseteq H$.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. Let H be a conceptual structure and let A be a (part of a) conceptual structure. By the conceptual elimination operation we defined in definition 10, it holds that $H - A \subseteq A$. Since our contraction operation functions via intersections ($\cap^*$), and with (3) in theorem 6.4.1 (for all conceptual structures S, it holds that $f(S) \subseteq S$), it holds that $H \ominus A \subseteq A$. $\square$

The vacuity postulate for conceptual contraction states that, if the argument of the contraction does not occur in the initial conceptual structure, then no changes are made to this structure. In our framework, this requirement is not satisfied, since it is possible that the initial conceptual structure changes in the process of consistency-recovery. A weaker version of this requirement, assuming that the initial conceptual structure is a conceptual hierarchy, is, however, satisfied, aside from the reordering of the preferences.

Theorem 6.4.8. For all conceptual hierarchies H and for all conceptual structures A (A might also be a part of a conceptual structure), if $A \nsubseteq H$ then $H \ominus A \subseteq H$ and $H \subseteq H \ominus A$.

Proof. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision contraction model. Let H be a conceptual hierarchy and let A be a (part of a) conceptual structure. Suppose $A \nsubseteq H$. It follows that $H - A = H$. Since H is the unique maximal conceptual hierarchy in H, it holds that $f(H - A) = H$ see point (v) in theorem 6.4.1 above. Thus, $H \ominus A$ differs from H only in terms of the preference ordering between conceptual structures in the model. Therefore, our claim holds.

Let us also show that this does not hold unless the initial structure is a conceptual hierarchy. Let H be a conceptual structure and let A be a (part of a) conceptual structure. Suppose $A \nsubseteq H$. It follows that $H - A = H$. Given that H is not a conceptual hierarchy, intersecting the most preferred maximal hierarchies within H yields a proper substructure of H. Hence, it holds that $f(H \ominus A) \subset H$, see point (vi) in theorem 6.4.1 above. □

Next, we consider a counterpart of the AGM recovery postulate for contractions. Our version of the postulate requires that the result of the contraction operation is such that if it is expanded with the argument of the contraction, the initial conceptual structure is recovered. As we will see this requirement does not hold in our framework. This should not come as a surprise. Counterparts of this principle have in fact been widely rejected in alternative theories of belief revision, including base-generated revisions [68] and belief withdrawals [108], and in application of such belief revision strategies to scientific change, for example in abductive belief revision in science [135]. In order for the recovery principle to be satisfied, a rational change theory should first and foremost adhere to the principle of minimal change. In particular, the principle states that so much information is retained in a contracted theory, up to the point that the initial theory can be recovered by a simple expansion with the argument of the contraction. It is often argued that there are other important rationality postulates, concerning the preferences of the agents, which should not be overwhelmed by the minimal change principle (cf. [134]). The failure of this postulate is then again in line with our stance concerning the revolutionary nature of scientific change modeled in this paper. In our framework, given that a conceptual structure H is contracted by a conceptual structure (or a part of a conceptual structure) A, it might be the case that $H \ominus A \subset H - A$ (see theorem 6.4.6), and expanding $H \ominus A$ with A does not necessarily recover the information that is lost in the transition from $H - A$ to $H \ominus A$.

Theorem 6.4.9. Given that H is a conceptual structure and A is a (part of a) conceptual structure, it might not hold that $(H \ominus A) \oplus A \subseteq H$ and $H \subseteq (H \ominus A) \oplus A$.

Proof. We show with a counterexample. Let $CS^{\oplus\ominus\leqslant} = \{\mathcal{CS}, \oplus, \ominus, \leqslant\}$ be a conceptual revision and contraction model, with $\leqslant$ such that for all $A, B \in \mathcal{CS}$, $A \leqslant B$ iff A includes the instance pair $\langle Bob, mammal\rangle$. Let $H = \langle C_H = \{[orka], [mammal]\}, O_H = \{(Bob)\}, K_H\{\langle orka, mammal\}, I_H = \{\langle Bob, orka\rangle, \langle Bob, mammal\rangle\}, P_H = R_H = \varnothing\rangle$ be a conceptual structure in $\mathcal{CS}$. Let $A = \{\langle orka, mammal\rangle\}$ be a part of a conceptual structure in $\mathcal{CS}$. Thus, $H \ominus A = \langle C_{H-A} = C_H, O_{H-A} = O_H, K_{H-A} = \varnothing, I_{H-A} = \{\langle Bob, mammal\rangle\}, P_{H-A} = P_H, R_{H-A} = R_H\rangle$.

On the other hand, $H' = (H \ominus A) \oplus A = \langle C_{H'} = C_H, O_{H'} = O_H, K_{H'} = K_H, I_{H'} = \{\langle Bob, mammal\rangle\}, P_M = R_M = \varnothing\rangle$. Note that $H \nsubseteq H'$ since the instance pair $\langle Bob, orka\rangle$ is included in H but not in H'. Therefore, our claim holds. $\square$

Lastly, we will mention the *Levi-identity*, which reduces AGM belief revision to an operation of AGM belief contraction followed by an expansion. In particular, Levi-identity says that revision of a theory H with a piece of information A can be performed by first contracting H with the negation of A, hence making space for consistent incorporation of A, and then expanding this contracted theory with A. In our framework, an analogous identification of conceptual revision with a sequence of conceptual contraction and expansion is not possible. This can be shown with the help of a simple informal example. Suppose we want to revise a conceptual structure that includes the kind-relation representing the information that orkas are fishes, with the new, contrasting, kind-relation expressing the information that orkas are mammals. Assuming the Levi-identity, we would need to first remove from the initial theory whatever contradicts with this latter kind-relation. This might include not only the kind-relation between orkas and fishes, but also (possibly) other related parts of a conceptual structure, such as rules relations encoding important information about fishes. This additionally contracted parts are not necessarily recovered by an expansion operation.[16]

This alleged failure of Levi-identity in our framework should not be surprising. Similar to the case for the recovery postulate above, while contracting a conceptual structure, we may end up eliminating more information than what is required for the consistent incorporation of the new information. Moreover, in the scientific context, it can be argued that Levi-identity should not hold. For instance, Schurz [135] states that, in the context of his abductive belief revision framework, combining ordinary belief expansion and abduction generation based on a contracted theory does not describe abductive revision in science. This is because the information provided by the initial theory that gets lost in the contraction is not necessarily recovered in this way. The same rationale explains the failure of (an alleged translation of) the Levi-identity in our model of conceptual revision.

[16]We conclude our discussion of the failure of the Levi-identity here, since a formal counterexample requires formal tools we have not introduced, such as negating a conceptual structure or a part of it.

6.5 Taking up Thagard's challenge

We have now presented our conceptual revision model and we have shown how our revision and contraction operations satisfy several rationality postulates for conceptual change. In this section, we will show how we can mirror the dynamics of Thagard's conceptual systems in our system. Specifically, we will demonstrate how almost every kind of change described by Thagard can be adequately represented in our framework via a suitable (combination of) change operation(s) on conceptual structures.

6.5.1 Mirroring Thagard's kinds of changes in our conceptual revision model

As we saw in Section 6.2, Thagard described a hierarchy of nine degrees of changes applicable to conceptual systems, ordered by their increasing strength: instance-addition, rule addition, part addition, kind addition, concept addition, kind collapse, hierarchy reorganization, and tree switching.

In what follows, we will discuss each of these degrees of change one by one, from the weakest to the most radical one. With the exception of tree-switching, whose case will be completely different from all the others, the structure of our discussion will take the following form. We will first present how a given kind of change operates on one of Thagard's conceptual systems. Then, we will explain informally how this kind of change can be represented in our framework. After that, we will give a formal definition of the degree of change under focus, showing how it can be seen as a special case of (a series of applications of) our revision and/or our contraction operations. Finally, we will present a toy-example of this kind of change in our framework in order to make clearer our proposed formalisation.

Instance-addition. The addition of an instance-link is the least radical kind of change described by Thagard. It consists in the addition of a single instance link between one object node and one conceptual node of a given conceptual system, representing the information that a given individual is an instance of a given concept.

In our framework, we can mirror instance-addition via our conceptual revision operation, revising a given conceptual structure with another conceptual structure that includes a non-empty instance-relation. In particular, we can define three different forms of instance-addition as three different constraints on the argument of revision. The most general form, what we will call general instance-addition, consists of requiring the argument of the revision to include a non-empty instance relation. A more specific form of instance-addition, i.e., pure instance-addition, requires the argument of the revision to have instance-relation as its only non-empty relation (concept and object domains can be non-empty as well). Finally, we have an atomic instance-addition when the argument of the revision of a pure instance-addition has a single instance-pair as its instance-relation. This last form corresponds to (our interpretation of) Thagard's understanding of instance-addition.

More formally, a conceptual revision operation $H * A$ represents a *general instance-addition* iff $I_A \neq \varnothing$. A conceptual revision operation $H * A$ represents a *pure instance-addition* iff $I_A \neq \varnothing$ and $K_A = P_A = R_A = \varnothing$. A conceptual revision operation $H * A$ represents an *atomic instance-addition* iff $|I_A| = 1$ and $K_A = P_A = R_A = \varnothing$. For an example of a general instance-addition, see the conceptual revision example presented in Section 3.1.

Rule-addition. The second kind of change described by Thagard consists in adding a rule-link between two concepts nodes of a given conceptual system. This change represents adding the information that a generic holds between two concepts.

In our framework, rule-addition is represented similarly as we treated instance-addition, i.e., by requiring the argument of our revision operation to include a non-empty rule-relation. As in the previous case, three different forms of rule-addition can be defined, differing in terms of generality: general rule-addition, pure rule-addition, and atomic rule-addition.

More formally, a conceptual revision operation $H * A$ represents a *general rule-addition* iff $R_A \neq \varnothing$. A conceptual revision operation $H * A$ represents a *pure rule-addition* iff $R_A \neq \varnothing$ and $K_A = P_A = I_A = \varnothing$. A conceptual revision operation $H * A$ represents an *atomic rule-addition* iff $|R_A| = 1$ and $K_A = P_A = I_A = \varnothing$.

As a simple example of rule-addition, let H be composed by:

$$\mathcal{C}_H = \{[mammal], [whale], [orka]\}, \mathcal{O}_H = I_H = P_H = R_H = \varnothing$$
$$K_H = \{\langle whale, mammal\rangle, \langle orka, mammal\rangle, \langle orka, whale\rangle\},$$

and let A be composed by $\mathcal{C}_A = \{[mammal], [air]\}$, $\mathcal{O}_A = K_A = P_A = I_A = \varnothing$, $R_A = \{\langle mammal, air\rangle\}$ (intuitive interpretation: mammals breath air).[17] The output of this revision operation expands the rule-relation of H with R_A and the pairs $\langle whale, air\rangle, \langle orka, air\rangle$. We then have $H \oplus A = M$ where:

$$\mathcal{C}_M = \{[mammal], [whale], [orka], [air]\}, \mathcal{O}_M = P_M = I_M = \varnothing$$
$$K_M = \{\langle whale, mammal\rangle, \langle orka, whale\rangle, \langle orka, mammal\rangle\}$$
$$R_M = \{\langle mammal, air\rangle, \langle whale, air\rangle, \langle orka, air\rangle\}.$$

Since M is a conceptual hierarchy, we have $H * A = M$, modulo the revised preference ordering (Figure 6.5).

Part-addition. The third kind of change described by Thagard is called part-addition or decomposition. It consists in adding a part-link between two concept nodes of a given

[17] Note that it would be possible in our framework to differentiate rules in terms of their intended interpretation, so that for instance the rule *breath* is represented differently from other rules (e.g., *swim*) that may be added to a given conceptual structures. We decided to follow Thagard in leaving the interpretation of the rules outside our framework, considering all rules as uninterpreted rule-pairs.

conceptual system, representing the information that a relation of part-hood holds between the concepts denoted by these nodes.

In our framework, part-addition is represented similarly as we treated instance-addition and rule-addition, i.e., by requiring the argument of our revision operation to include a non-empty part-relation. As in the previous cases, three different forms of part-addition can be defined, differing in terms of generality: general part-addition, pure part-addition, and atomic part-addition.

More formally, a conceptual revision operation $H * A$ represents a *general part-addition* iff $P_A \neq \varnothing$. A conceptual revision operation $H * A$ represents a *pure part-addition* iff $P_A \neq \varnothing$ and $K_A = R_A = I_A = \varnothing$. A conceptual revision operation $H * A$ represents an *atomic part-addition* iff $|P_A| = 1$ and $K_A = R_A = I_A = \varnothing$.

As a simple example of part-addition, take H to be such that:

$$\mathcal{C}_H = \{[mammal], [whale], [orka]\}, \mathcal{O}_H = P_H = R_H = I_H = \varnothing$$
$$K_H = \{\langle whale, mammal\rangle, \langle orka, mammal\rangle, \langle orka, whale\rangle\}.$$

Let A be composed by $\mathcal{C}_A = \{[mammal], [lungs]\}$, $P_A = \{\langle mammal, lungs\rangle$ (intuitive interpretation: mammals have lungs), and $K_A = R_A = I_A = \mathcal{O}_A = \varnothing$. The output of this revision operation expands the part-relation of H with P_A and the pairs $\langle whale, lungs\rangle, \langle orka, lungs\rangle$, in order to recover transitivity. We then have $H \oplus A = M$ where:

$$\mathcal{C}_M = \{[mammal], [whale], [orka], [lungs]\}, \mathcal{O}_M = R_M = I_M = \varnothing$$
$$K_H = \{\langle whale, mammal\rangle, \langle orka, mammal\rangle, \langle orka, whale\rangle\}$$
$$P_M = \{\langle mammal, lungs\rangle, \langle whale, lungs\rangle, \langle orka, lungs\rangle\}.$$

Since M is a conceptual hierarchy, we have $H * A = M$ (Figure 6.5).

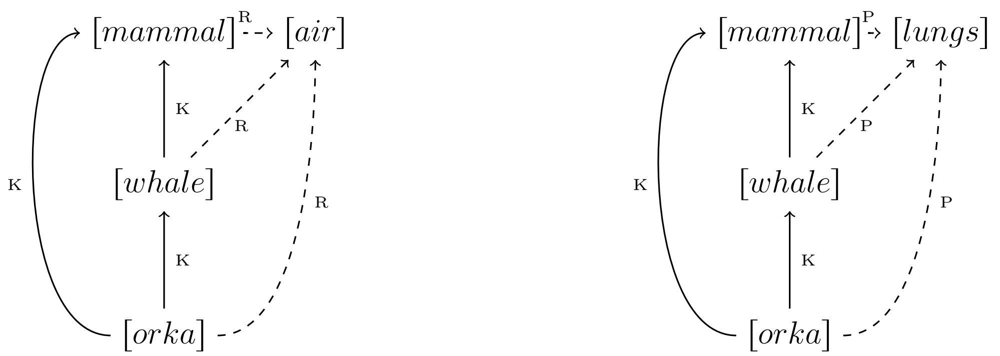

Figure 6.5: The output of the rule-addition example (on the left) and the output of the part-addition example (on the right).

Kind-addition. The fourth kind of change described by Thagard consists in adding a kind-link between two concept nodes of a given conceptual system, representing the information that a relation of kind-hood holds between the concepts denoted by these nodes. Furthermore, Thagard, following Carey's terminology for conceptual change in child psychology [31], distinguishes two special cases of (series of) kind-addition(s): *coalescence* and *differentiation*. The former type of kind-addition happens when we add a superordinate conceptual node linked via a series of kind-links with some concept nodes that had no superordinate kinds before. The latter denotes instead the addition of some subordinate conceptual nodes connected via a series of kind-links with a conceptual node that before had no subordinate kinds.

In our framework, kind-addition is represented by requiring the argument of our revision operation to include a non-empty kind-relation. As in the previous cases, three different forms of kind-addition can be defined, differing in terms of generality: general kind-addition, pure kind-addition, and atomic kind-addition. Coalescence and differentiation can then be represented as specific cases of general or pure kind-addition.

Formally, a conceptual revision operation $H * A$ represents a *general kind-addition* iff $K_A \neq \emptyset$. A conceptual revision operation $H * A$ represents a *pure kind-addition* iff $K_A \neq \emptyset$ and $P_A = R_A = I_A = \emptyset$. A conceptual revision operation $H * A$ represents an *atomic kind-addition* iff $|K_A| = 1$ and $P_A = R_A = I_A = \emptyset$. Furthermore, a general or pure kind-addition $H * A$ is a *coalescence* iff there exists a $x \in C_A$ such that $\langle y, x \rangle \in K_A$ and there is no w such that $\langle y, w \rangle \in K_H$. A general or pure kind-addition $H * A$ is instead a *differentiation* iff there is an $x \in C_A$ such that $\langle x, y \rangle \in K_A$ and there is no w such that $\langle w, y \rangle \in K_H$. For an example of a general kind-addition, see the conceptual revision example presented in Section 3.1.

Concept-addition. The fifth kind of change described by Thagard consists in adding a new concept node to a given conceptual system. This type of change represents the addition of a new concept to a given scientific theory.[18]

In our framework, concept-addition is represented by requiring the argument of our revision operation to include a new concept. Several further restrictions can be imposed. For instance, we present here two more specific forms of concept-addition: unique concept-addition and connected concept-addition. We have a unique concept-addition when there is only one new concept in the argument of the revision (it may also include non-empty relations). We have a connected-concept addition when each new concept in the argument figures in at least one relation.

Formally, a conceptual revision operation $H * A$ is a *concept-addition* iff there is an $x \in C_A$ such that $x \notin C_H$. A concept-addition $H * A$ is then a unique concept-addition iff there is only one $x \in C_A$ such that $x \notin C_H$. A concept-addition $H * A$ is then a connected concept-addition iff for all $x \in C_A$ such that $x \notin C_H$ there exists a $y \in C_A \cup O_A$ such

[18] Thagard also stresses how concept-addition sometimes involves combining two simple concepts into a complex one [149, p.35-36]. This combination aspect of concept-addition is outside the scope of the present version of our framework, since we assumed for simplicity that the concept universe is constant.

that $\langle x, y\rangle$ or $\langle y, x\rangle$ is in $K_A \cup P_A \cup R_A \cup I_A$. For an example of a unique and connected concept-addition see the conceptual revision example in Section 6.3.1, the rule-addition example, or the part-addition example above.

Kind-collapse. The sixth change described by Thagard is kind-collapse, i.e., the removal of a (series of) kind-link(s) from a given conceptual system. More specifically, Thagard says that kind-collapse is the inverse change of differentiation, so that kind-collapse denotes removing all subordinate kinds of a given conceptual node.

In our framework, kind-collapse is a specific case of our contraction operation, namely, the contraction of a given conceptual structure with respect to a set of kind-pairs all of which have the same element as their second element and such that in the contracted structure this element has no subordinate kinds.

Formally, a conceptual contraction operation $H - A$ is a *kind-collapse* iff $\exists x \in C_H$ such that $K_A = \{\langle j_1, x\rangle, \ldots, \langle j_n, x\rangle\}$ and $\neg\exists y \in C_H \cup C_A$ such that $\langle y, x\rangle \in K_{H-A}$. This definition of a kind-collapse makes it the inverse process of a differentiation, just like in Thagard's system. For an example of a kind-collapse, see the contraction example in Section 6.3.2.

Hierarchy-reorganization. The seventh kind of change in Thagard's theory is the general process of hierarchy-reorganisation or *branch-jumping*, i.e., moving a set of concept and object nodes from one part of a conceptual system to another one, thus changing (some of) their relations. This change is typical of many scientific revolutions, such as the Copernican revolution in which the earth branch-jumped from being a unique entity to a kind of planet.

In our framework branch-jumping is a specific series of our contraction and revision operations that does not involve changes to the concept-domains of the conceptual structures involved. The output of such combination is the transportation of certain parts of a given conceptual structure to a different part of it, involving some change in its relations.

Formally, we say that the sequence of contraction and revision operations $(H - A_1) * A_2$ represents a *hierarchy-reorganisation* iff $\mathcal{C}_H = \mathcal{C}_{(H-A_1)*A_2}$, $\mathcal{O}_H = \mathcal{O}_{(H-A_1)*A_2}$ and either $K_H \neq K_{(H-A_1)*A_2}$ or $P_H \neq P_{(H-A_1)*A_2}$ or $R_H \neq R_{(H-A_1)*A_2}$ or $I_H \neq I_{(H-A_1)*A_2}$. Note that we leave completely open how the relations between the objects and concepts involved in this type of change are transformed. Specific kinds of hierarchy-reorganisation, such as part-kind transformation, can then be defined by imposing further constraints on the relations in the contraction and in the revision operation.

As an example of a hierarchy-reorganisation, take H to be such that:

$$\mathcal{C}_H = \{[animal], [fish], [mammal], [whale], [orka]\},$$
$$\mathcal{O}_H = I_H = P_H = R_H = \varnothing$$
$$K_H = \{\langle whale, fish\rangle, \langle orka, fish\rangle, \langle orka, whale\rangle, \langle orka, animal\rangle,$$
$$\langle whale, animal\rangle, \langle mammal, animal\rangle, \langle fish, animal\rangle\}.$$

Let A_1 be the part of a conceptual structure $\{\langle whale, fish\rangle, \langle orka, whale\rangle, \langle orka, fish\rangle\}$ and A_2 be composed by
$\mathcal{K}_{A_2} = \{\langle whale, mammal\rangle, \langle orka, whale\rangle\}$, $\mathcal{C}_{A_2} = \{[whale], [orka], [mammal]\}$, $\mathcal{O}_{A_2} = \{\varnothing\}$, and $P_{A_2} = R_{A_2} = I_{A_2} = \varnothing$.

The output of the hierarchy-reorganization $H - A_1 * A_2$ is then equal to the structure H' (Figure 6.6) where:

$$\mathcal{C}_{\mathcal{H}'} = \{[mammal], [whale], [orka], [fish], [animal]\},$$
$$\mathcal{O}'_H = R'_H = P'_H = I'_H = \varnothing$$
$$K'_H = \{\langle whale, mammal\rangle, \langle orka, mammal\rangle, \langle orka, whale\rangle, \langle orka, animal\rangle,$$
$$\langle whale, animal\rangle, \langle mammal, animal\rangle, \langle fish, animal\rangle\}.$$

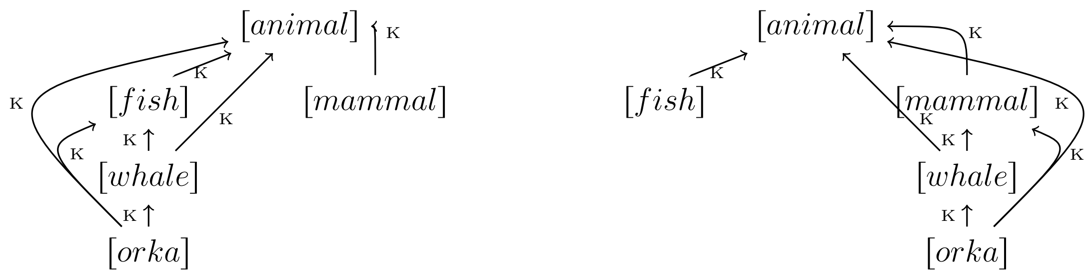

Figure 6.6: The input (on the left) and the output (on the right) of the hierarchy-reorganization example.

Tree-switching. The last change described by Thagard is tree-switching, i.e., the change of the organizing principle of the whole hierarchy. This change implies thus re-interpreting any kind-relation and part-relation. An example of this kind of change is the Darwinian revolution, a revolution that involved the re-interpretation of kind-relations of biological entities as historical kinship and not as they were before as morphological similarities. This is the most radical change that can happen in science for Thagard, up to the point that it is sufficient but not necessary for having a conceptual revolution. Only certain scientific revolutions that are particularly radical exemplify tree-switching.

Since tree-switching is not really about changing the structure of a conceptual system, focusing instead on the external interpretation of the conceptual system, it would be at least unclear how to frame this kind of change in our framework. Using an epistemological metaphor, modeling tree-switching in our framework would be like implementing a gestalt-operation in traditional belief revision that changes the meaning of the logical consequence between beliefs. We therefore do not treat this kind of change in the present work, focusing only on the first eight changes that affect the internal-structure of conceptual system, confident that we do not loose too much in generality, since as Thagard himself acknowledges many scientific revolutions do not even exemplify tree-switching.

6.6 Concluding remarks

Let us recap the main steps of the present work. Starting from Thagard's model of scientific conceptual change, we saw his taxonomy of nine degrees of conceptual change and his claim that belief revision theories can only account for the first two of them. We then presented our system of conceptual revision, i.e., a belief-revision-like system for conceptual structures. We showed how our conceptual revision and contraction operations satisfy several rationality postulates analogous to the AGM ones. We then demonstrated how our system, working at the conceptual level of abstraction, is able to mirror eight out of nine kinds of conceptual changes described by Thagard.

More generally, our framework shows how belief revision theories can be mapped to the conceptual level in order to obtain a logical interpretation of radical conceptual change. The present work is only a first step towards a better understanding of the relationships between belief change and conceptual change. Several directions of future work naturally present themselves. Interesting ways of extending our framework include investigating specific preference orderings and alternative ways of changing them, reconstructing case studies from the history of science as series of conceptual revision and contractions, working with expanding domains to model conceptual combination, adding the possibility of revising conceptual structure with complex information (such as negative one, for instance) to further model logical relationships between elements of a conceptual structure, having a way of comparing differing conceptual structures in order to model Thagard's explanatory coherence notion, and also augmenting our conceptual structures in order to mimic more elaborate approaches to theory-change (e.g., [86, 12, 6, 112, 81]). These extensions would allow to model even Thagard's most radical type of conceptual change, i.e., tree-switching. It would also be interesting to merge conceptual structures with (structured) belief sets, in order to have a revision system capable of revising beliefs and concepts at the same time. Such a conceptual-plus-belief-revision system would be able to model (some of) the interesting connections between conceptual change and belief change, thereby offering a more fine-grained logical reconstruction of scientific change.

Chapter 7

Conclusion

The main theme of this dissertation is to discuss and provide modeling solutions that address the over-idealisation tradition in the literature on belief ascription and belief dynamics, and the resulting problem of logical omniscience. The problem of logical omniscience has various sources, hence various modeling solutions have been developed. In the limited scope of the current work, I have focused on the aspects of the problem that are related to the topics of hyperintensionality, deductive closure, hence fragmentation, and inconsistency-tolerance. The hyperintensional belief base revision (BBR) models developed in Chapter 2 set the foundation for the discussion of these issues. BBR models are dynamic (i.e., model changing) belief revision models based on structured and ordered (sets of) information states. Switching from the traditional Kripke models for belief to a state-based (or support-based) semantics is responsible for many of the advantages of the BBR models. First, based on information states, we can naturally represent incomplete and inconsistent information collections. Exploiting the mereological structure of states, we have a natural account of information growth. The choice of information states as the primitive elements of representation allows us to shift the representational focus of doxastic states from belief sets to collections of information. With changes of belief characterised as artifacts of changes on the level of information, we are able to draw a distinction between the two levels. The motivation for this distinction is best spelled out by Barwise:

> Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference. [14, p.762]

While inferences on the level of beliefs, as well as belief dynamics, are subject to various rationality criteria in the form of logical rules and axioms, the propositional logic of BBR (the belief-free portion of the logic) is still stronger than the modal logic of BBR (i.e., more validities of classical logic are maintained by the belief-free portion). The use of information states and the distinction between the level of information and level of beliefs prove useful also for proposing an account of fragmented belief based on fragments of information, and for defining consistency-tolerant epistemic modalities.

The background logic for these models is the propositional HYPE-logic developed by Leitgeb [91]. Propositional HYPE-logic is non-classical as various classical validities fail, such as excluded middle, the law of non-contradiction, explosion, disjunctive syllogism, and (general) contraposition [91, p.333]. With the addition of a non-normal belief modality B, the logic of BBR is also non-monotonic, hence substructural.

After proposing the BBR models along with a sound axiomatization, various aspects of the models are explored extensively in the following chapters. In Chapter 3 hyperintensionality of belief revision based on BBR models is characterised as a result of the reasoning contexts set by the information states. Based on state-based semantics, the designations of propositions are sets of states, rather than sets of possible worlds. By virtue of this shift, intentional equivalences may come apart, as propositions which designate the same set of worlds may designate distinct sets of states. I presented examples of hyperintensional semantics from the literature, which reconstruct subject-matters as partitions (and divisions) of the logical space (represented as a set of possible worlds), and which explicitly expand the Kripke semantics for belief with topics. These aim to explicate the hyperintensionality of belief attributions and belief revision, in terms of sensitivity to subject-matters and topics of sentences in certain contexts, respectively. I proposed an evaluation of the reasoning contexts set by information states in the same lines, as adding a further dimension to the meaning of sentences besides their classical truth-functional meaning (i.e., their intensions). The primary point that puts apart the possible-worlds-based approaches from our state-based approach lies in the description of logical entailment relations suggested by these approaches. The former suggest two-layered entailment relations which strengthen truth-functional entailment with topicality or subject-matter filters. The latter redefines logical entailment based on support relations. Finally, I presented a discussion of how characterisation of hyperintensionality based on possible-worlds models and state-based models affect the interaction of logical connectives and modalities with topics and with reasoning contexts. The proposed future research in this area is explicitly representing the reasoning contexts and their dynamics. The proposal is motivated by the success of other frameworks in explicating hyperintensionality which do so.

While Chapter 3 is a survey of the hyperintensionality literature, with a comparative assessment of two different modeling approaches based on a number of issues, what is presented is not a comprehensive comparison. Future work that puts forth a comparison of the two approaches in terms of naturalness and flexibility could also benefit the wider literature. On the other hand, the so-called *hyperintensional revolution* is not welcomed by all. In [21] Berto responds to Williamson's accusation of hyperintensional models as *overfitting* raised in [169]. Berto defends hyperintensionality through the idea that a primary task of semantics is to capture competent speakers' intuitions about meaning, entailment, equivalence, and the like, the semantic theories should hence start from ordinary language use, and patterns of reasoning [21, p.5]. I do agree with Berto here, as the task pursued in this dissertation is to challenge the over-idealisation tradition based on insights from ordinary reasoning patterns.

In Chapter 4 the discussion shifts from reasoning contexts in the sense of degrees of content resolution to reasoning fragments. Fragmented belief states are defined as shifting belief states of agents based on their focus of attention, conversational context, etc., where each fragment is deductively closed in itself, logically independent from the others, and they are activated one at a time to guide agents' actions. I focused on various defenses of fragmentation in the literature, before pointing out the natural transition from fragmented information states, which are the building blocks of the BBR models, to fragmented belief states. I have shaped this chapter to address the requirements of deductive closure and consistency, as part of the problem of logical omniscience, rather than hyperintensionality. There are, however, obvious connections and even overlaps between the literature on hyperintensionality and fragmentation.

Both hyperintensionality and fragmentation of belief states are proposed as modeling solutions for the problem of logical omniscience. Recall that the problem of logical omniscience indeed refers to various issues of logical and deductive closure, summarised as closure conditions C1-C8 Figure 4.1. There are various fragmentation approaches that address in particular C3, C6, C7, which we can categorise as pure fragmentation approaches. Examples include Montague-Scott neighborhood structures applied to belief ascription by Vardi in [165], and local reasoning models by Fagin and Halpern described in [45], where beliefs of agents are modelled as sets of propositions (where a proposition P is a set of possible worlds) assigned to each possible state of an agent, however, for each fragment of the belief state of an agent, there is a different set of propositions. These approaches, however, do not solve the logical omniscience problem on their own, since the designations of propositions and the primitive elements of formalisation of the belief states are still possible worlds. For instance, the local reasoning models validate knowledge of valid formulas and closure under valid implication. On the other hand, Levesque's work on distinguishing implicit and explicit beliefs, by introducing basically a state-semantics for explicit beliefs and keeping a possible worlds evaluation for implicit beliefs, addresses both the hyperintensionality aspect and the fragmentation aspect of the problem of logical omniscience. Levesque's approach avoids, only in terms of explicit beliefs, closure under believed implication and non-consistency in virtue of an agent having inconsistent information, and closure under valid implication and belief of valid formulas in virtue of *lack of awareness*, i.e., partialness of the explicit belief states [45, p.46].[1]

Another set of approaches suggests what I call a partition-sensitive fragmentation. I stated Yalcin's question-sensitive and resolution-sensitive models [172, 173], and Leitgeb's partition-based reasoning [90]. These approaches explain fragmentation based on what I called in Chapter 3 the reasoning contexts. Here, fragmentation occurs as a result of different ways of partitioning the logical space. Leitgeb holds that this affects which sentences are *entertainable* at a time. They leave no room for entertaining inconsistent beliefs

[1]Vardi notes that while Levesque's logic avoids the logical closure of beliefs under classical logic, it follows from Levesque's results that beliefs are closed under relevance logics [165, p.294]. The diagnosis concerning *lack of awareness* is due to Fagin and Halpern, as well as the diagnosis that deductive closure of beliefs is avoided only when the agents actually have inconsistent information [45, p.47]. They then introduce awareness logic as a response to some problems of Levesque's logic [45, p.48], achieving similar results by introducing an awareness filter on a possible worlds modeling for explicit beliefs. The awareness operator assigns to each possible state, a set of primitive propositions that the agent is aware at that state, and the agents cannot have explicit knowledge about formulas of which they are not aware. Note the similarity of the awareness filter and the topicality filters.

within the same resolution state, moreover, Yalcin does not allow distinct partitions with the same level of resolutions, or where one resolution strictly refines the other. According to the fragmentation view defended in this dissertation, however, each fragment of an information state has the same content resolution, in other words, the reasoning contexts which bring about hyperintensionality are distinct from the conversational contexts that determine which fragment is activated. Reasoning contexts may only shift with new information (as new subject-matter or topics are added), while fragments of a belief state exist simultaneously and the shifts between them occur as the focus situation of the agent shifts. This is also the view I take towards fragmentation via the BBR models. In this view, fragmentation may have motivations other than shifting levels of resolution or thresholds (as in [90]) such as memory access, resource bounded reasoning, conversational contexts, etc.

The BBR models, then, distinguish between the reasoning contexts set by information states that is responsible for hyperintensional sensitivity of the models and the conversational context that is responsible for fragmentation. The flexibility of the BBR models proves advantages in terms of combining modeling solutions for both concerns. There are other examples in the literature that achieve this, for instance, Jago proposes incorporating a model of local reasoning in the style of [45] within the TMS belief ascription models based on truth-maker semantics he introduces in [78, p.15]. Accordingly, each local fragment is modelled on the TMS account. The BBR models lay the foundations for such an incorporation of a full theory of fragmented beliefs, where the belief states are fragmented based on different focal issues, as well as how these issues relate to their parts and other issues. Such a fragmentation approach can indeed be applied to group reasoning settings (replacing the society of minds with real societies), where members of a group might structure an otherwise equivalent body of information and might reach different conclusions based on different connections they make with the focal issue. Focal issues can be understood as gravitational points that fix a conversational context, while different networks agents build around these points might still lead to different conclusions. Applying the BBR models for fragmented minds, as well as for group reasoning then remains an exciting research direction.

In Chapter 5 the concept of information bases of Chapter 2 is given an alternative formalisation in an information-based semantics for intuitionistic logic. While the models in this chapter are formally interesting in themselves, they also allow us to elaborate on the notion of *information sources*, identifying *trusted sources* as sole sources of consistent belief in virtue of the consistency-sensitive epistemic modalities. Exiting venues of future research remains, in particular formalising a more interesting, non-persistent belief modality, and a more fine-grained mechanism to order and choose trusted sources of information. The insights for the future of this work come from the BBR models, evidence models [160, 157], and informational semantics [124].

Finally, in Chapter 6 I have presented a belief-revision-like system for conceptual revision. This chapter contributes to the emerging literature that connects belief revision and scientific change in a more robust way. I also argued that concept possession can be understood as another dimension of reasoning, which can be aligned with reasoning contexts set by information states, in a way that the concept repertoire of an agent partly determines the reasoning context she is in. Future research that follow from this work is then in two different directions. First, we can apply the conceptual models to improve the understanding of inconsistent and patch-work concepts. Second, by incorporating conceptual structures within a belief revision system, we can improve our understanding of how concept possession interacts with hyperintensional belief revision.

Bibliography

[1] Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530.

[2] Aloni, M. (2005). A formal treatment of the pragmatics of questions and attitudes. *Linguistics and Philosophy*, 28:505–539.

[3] Aloni, M., Beaver, D., Clark, B., and Van Rooij, R. (2007a). 6: The dynamics of topic and focus. In *Questions in dynamic semantics*, pages 123–145. Brill.

[4] Aloni, M., Butler, A., and Dekker, P. (2007b). *Questions in dynamic semantics*, volume 17. Brill.

[5] Aloni, M. and Égré, P. (2010). Alternative questions and knowledge attributions. *The Philosophical Quarterly*, 60(238):1–27.

[6] Andersen, H., Barker, P., Chen, X., and Tweney, R. D. (2006). The cognitive structure of scientific revolutions. *Aestimatio: Sources and Studies in the History of Science*, 3:153–162.

[7] Anderson, A. R. and Belnap, N. D. (1962). Tautological entailments. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 13(1/2):9–24.

[8] Andreas, H. (2011). A structuralist theory of belief revision. *Journal of Logic, Language and Information*, 20:205–232.

[9] Angell, R. B. (1989). Deducibility, entailment and analytic containment. In *Directions in relevant logic*, pages 119–143. Springer.

[10] Balbiani, P., Fernández-Duque, D., and Lorini, E. (2019). The dynamics of epistemic attitudes in resource-bounded agents. *Studia Logica*, 107:457–488.

[11] Baltag, A., Fiutek, V., and Smets, S. (2013). DDL as an internalization of dynamic belief revision. In *Krister Segerberg on Logic of Actions*, pages 253–280. Springer.

[12] Balzer, W., Moulines, C. U., and Sneed, J. D. (2012). *An architectonic for science: The structuralist program*, volume 186. Springer Science & Business Media.

[13] Barwise, J. (1989). *The Situation in Logic*, volume 4. Center for the Study of Language (CSLI).

[14] Barwise, J. (2016). Three views of common knowledge. In *Readings in Formal Epistemology: Sourcebook*, pages 759–772. Springer.

[15] Barwise, J. and Perry, J. (1981). Situations and attitudes. *The Journal of Philosophy*, 78(11):668–691.

[16] Belnap, N. D. (1960). Entailment and relevance1. *The Journal of Symbolic Logic*, 25(2):144–146.

[17] Berto, F. (2015). A modality called negation. *Mind*, 124(495):761–793.

[18] Berto, F. (2018). Aboutness in imagination. *Philosophical Studies*, 175(8):1871–1886.

[19] Berto, F. (2019). Simple hyperintensional belief revision. *Erkenntnis*, 84(3):559–575.

[20] Berto, F. (2022). *Topics of thought: The logic of knowledge, belief, imagination*. Oxford University Press.

[21] Berto, F. (2024). Hyperintensionality and overfitting. *Synthese*, 203(4):117.

[22] Berto, F. and Hawke, P. (2021). Knowability relative to information. *Mind*, 130(517):1–33.

[23] Berto, F. and Jago, M. (2019). *Impossible worlds*. Oxford University Press, Oxford.

[24] Berto, F., Sedlár, I., and Blicha, M. (2018). The theory of topic-sensitive intentional modals. *The logica yearbook*, pages 31–56.

[25] Beth, E. (1967). Semantic construction of intuitionistic logic. mededelingen der koninklijke nederandse akademie van wetenschappen, afd. letterkunde, 19 (11): 357–388, 1956. 25 errett bishop. foundations of constructive analysis. *McGraw-Hill, New York*, 20:21.

[26] Bílková, M., Majer, O., and Peliš, M. (2016). Epistemic logics for sceptical agents. *Journal of Logic and Computation*, 26(6):1815–1841.

[27] Bılková, M., Majer, O., Peliš, M., and Restall, G. (2010). Relevant agents. *Advances in modal logic*, 8:22–38.

[28] Borgoni, C. (2021). Rationality in fragmented belief systems. In *The fragmented mind*, pages 137–155. Oxford University Press Oxford.

[29] Bozdag, S. (2022). A semantics for hyperintensional belief revision based on information bases. *Studia Logica*, 110(3):679–716.

[30] Bozdag, S. and De Benedetto, M. (2022). Taking up thagard's challenge: a formal model of conceptual revision. *Journal of Philosophical Logic*, 51(4):791–824.

[31] Carey, S. (1987). *Conceptual Change in Childhood*. Bradford books. MIT Press.

[32] Carnap, R. (1947). *Meaning and necessity; a study in semantics and modal logic.* University of Chicago Press.

[33] Cherniak, C. (1990). *Minimal rationality*. Mit Press.

[34] Ciardelli, I. (2022). *Inquisitive Logic: consequence and inference in the realm of questions*. Springer Nature.

[35] Ciardelli, I., Groenendijk, J., and Roelofsen, F. (2019). *Inquisitive Semantics*, volume 6. Oxford University Press.

[36] Cresswell, M. J. (1975). Hyperintensional logic. *Studia Logica: An International Journal for Symbolic Logic*, 34(1):25–38.

[37] Cresswell, M. J. (2004). Possibility semantics for intuitionistic logic. *The Australasian Journal of Logic*, 2.

[38] Cresto, E. (2008). A model for structural changes of belief. *Studia Logica*, 88:431–451.

[39] Cross, C. and Roelofsen, F. (2024). Questions. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2024 edition.

[40] Deigan, M. (2020). A plea for inexact truthmaking. *Linguistics and Philosophy*, 43(5):515–536.

[41] Doyle, J. (1991). Rational belief revision. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 163–174.

[42] Dunn, J. M. (1993). Star and perp: Two treatments of negation. *Philosophical perspectives*, 7:331–357.

[43] Dunn, J. M. and Restall, G. (2002). Relevance logic. In *Handbook of philosophical logic*, pages 1–128. Springer.

[44] Elga, A. and Rayo, A. (2021). Fragmentation and information access. In *The fragmented mind*, pages 37–53. Oxford University Press Oxford.

[45] Fagin, R. and Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76.

[46] Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. (2004). *Reasoning about knowledge*. MIT press.

[47] Fermé, E. and Hansson, S. O. (2018). *Belief change: introduction and overview.* Springer.

[48] Fermé, E. L. (2001). Five faces of recovery. In *Frontiers in belief revision*, pages 247–259. Springer.

[49] Fine, K. (1986). Analytic implication. *Notre Dame Journal of Formal Logic*, 27(2):169–179.

[50] Fine, K. (2017). Truthmaker semantics. *A Companion to the Philosophy of Language*, pages 556–577.

[51] Fine, K. (2020a). Yablo on subject-matter. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, pages 129–171.

[52] Fine, K. (2020b). Yablo on subject-matter. *Philosophical Studies*, 177:129–171.

[53] Fine, K. and Jago, M. (2019). Logic for exact entailment. *The Review of Symbolic Logic*, 12(3):536–556.

[54] Foley, R. (2009). Beliefs, degrees of belief, and the lockean thesis. In *Degrees of belief*, pages 37–47. Springer.

[55] Friedman, M. (2001). *Dynamics of reason.* Csli Publications Stanford.

[56] Gabbay, D. M. (1985). *Theoretical foundations for non-monotonic reasoning in expert systems.* Springer.

[57] Gamerschlag, T., Gerland, D., Osswald, R., and Petersen, W. (2014). Frames and concept types. *Studies in linguistics and philosophy*, 94.

[58] Girard, P. and Rott, H. (2014). Belief revision and dynamic logic. *Johan Van Benthem on logic and information dynamics*, pages 203–233.

[59] Groenendijk, J. (1999). The logic of interrogation: Classical version. In *Semantics and Linguistic Theory*, pages 109–126.

[60] Groenendijk, J. and Roelofsen, F. (2009). Inquisitive semantics and pragmatics. In *Meaning, content and argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics and Rhetoric (SPR-09)*, pages 41–72.

[61] Groenendijk, J. A. G. and Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers.* PhD thesis, Univ. Amsterdam.

[62] Grove, A. (1988). Two Modellings for Theory Change. *Journal of Philosophical Logic*, 17:157–170.

[63] Gärdenfors, P. (1978). Conditionals and changes of belief. *Acta Philosophica Fennica*, 30:381–404.

[64] Gärdenfors, P. (1982). Rules for rational changes of belief. *Philosophical Essays dedicated to Lennart Aqvist on his fiftieth birthday*, 34:88–101.

[65] Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states.* The MIT Press.

[66] Hansson, B. (2010a). A note on theory change and belief revision. In *Belief Revision meets Philosophy of Science*, pages 155–162. Springer.

[67] Hansson, S. O. (2010b). Changing the scientific corpus. In *Belief revision meets philosophy of science*, pages 43–58. Springer.

[68] Hansson, S. O. (2012). *A textbook of belief dynamics: theory change and database updating.* Applied Logic Series. Springer, Dordrecht.

[69] Hawke, P. (2018). Theories of aboutness. *Australasian Journal of Philosophy*, 96(4):697–723.

[70] Hawke, P. and Özgün, A. (2023). Truthmaker semantics for epistemic logic. In *Kit Fine on Truthmakers, Relevance, and Non-classical Logic*, pages 295–335. Springer.

[71] Hawke, P., Özgün, A., and Berto, F. (2020). The fundamental problem of logical omniscience. *Journal of Philosophical Logic*, 49(4):727–766.

[72] Hintikka, J. (1983). Situations, possible worlds, and attitudes. *Synthese*, pages 153–162.

[73] Hintikka, K. J. J. (1962). *Knowledge and belief: An introduction to the logic of the two notions.* Cornell University Press.

[74] Horn, L. R. and Wansing, H. (2022). Negation. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Winter 2022 edition.

[75] Hulstijn, J. (1997). Structured information states: Raising and resolving issues. In *Proceedings of MunDial*, volume 97, pages 99–117. University of Munich Munich, Germany.

[76] Jäger, G. (2007). 5: Only updates. on the dynamics of the focus particle only. In *Questions in Dynamic Semantics*, pages 103–122. Brill.

[77] Jago, M. (2014). *The impossible: an essay on hyperintensionality.* Oxford University Press, Oxford.

[78] Jago, M. (2024). Belief in truthmaker semantics. *Synthese*, 204(4):123.

[79] Kindermann, D. (2021). On the availability of presuppositions in conversation. In *The fragmented mind*, pages 199–223. Oxford University Press Oxford.

[80] Kindermann, D. and Onofri, A. (2021). The fragmented mind: An introduction. In *The fragmented mind*, pages 1–33. Oxford University Press Oxford.

[81] Kornmesser, S. and Schurz, G. (2018). Analyzing theories in the frame model. *Erkenntnis*, pages 1–34.

[82] Krämer, S. (2022). Mighty belief revision. *Journal of Philosophical Logic*, 51(5):1175–1213.

[83] Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives*, volume 36. Oxford University Press.

[84] Kripke, S. A. (1965). Semantical analysis of intuitionistic logic i. In *Studies in Logic and the Foundations of Mathematics*, volume 40, pages 92–130. Elsevier.

[85] Kripke, S. A. (1980). Naming and necessity: Lectures given to the princeton university philosophy colloquium. In *Semantics of natural language*, pages 253–355. Springer.

[86] Kuhn, T. S. (1997). *The structure of scientific revolutions*, volume 962. University of Chicago press Chicago.

[87] Kyburg, H. E. (1961). *Probability and the logic of rational belief.* Middletown: Wesleyan University Press.

[88] Kyburg Jr, H. E. (1970). Conjunctivitis. In *Induction, acceptance and rational belief*, pages 55–82. Springer.

[89] Lakatos, I. (1970). Falsification and the methodology of scientific research programmes' in i. lakatos and a. musgrave (eds.) criticism and the growth of knowledge. In *Proceedings of the International Colloquium in the Philosophy of Science*, volume 4, page 196.

[90] Leitgeb, H. (2017). *The Stability of Belief: How Rational Belief Coheres with Probability.* Oxford University Press.

[91] Leitgeb, H. (2019). HYPE: A System of Hyperintensional Logic (with an Application to Semantic Paradoxes). *Journal of Philosophical Logic*, 48:305–405.

[92] Leitgeb, H. (2021). Exact truthmaking as inexact truthmaking by minimal totality facts. *Logic in High Definition: Trends in Logical Semantics*, pages 67–75.

[93] Leitgeb, H. and Segerberg, K. (2007). Dynamic doxastic logic: why, how, and where to? *Synthese*, 155:167–190.

[94] Levesque, H. J. (1984). A logic of implicit and explicit belief. In *AAAI*, volume 84, pages 198–202.

[95] Levi, I. (1967). Information and inference. *Synthese*, 17(4):369–391.

[96] Lewis, C. I. (1943). The modes of meaning. *Philosophy and phenomenological research*, 4(2):236–250.

[97] Lewis, D. (1970). General semantics. *Synthese*, 22(1):18–67.

[98] Lewis, D. (1982). Logic for equivocators. *Noûs*, pages 431–441.

[99] Lewis, D. (1988). Relevant implication. *Theoria*, 54:161–174.

[100] Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74:549–567.

[101] Lewis, D. K. (1973). *Counterfactuals*. Harvard University Press.

[102] Lindström, S. and Rabinowicz, W. (1989). On probabilistic representation of non-probabilistic belief revision. *Journal of Philosophical Logic*, pages 69–101.

[103] Lindström, S. and Rabinowicz, W. (1999). Belief change for introspective agents. *Spinning Ideas: Electronic Essays Dedicated to Peter Gardenfors on his Fiftieth Birthday.*

[104] Lindström, S. and Rabinowicz, W. (2005). Epistemic entrenchment with incomparabilities and relational belief revision. In *The Logic of Theory Change: Workshop, Konstanz, FRG, October 13–15, 1989 Proceedings*, pages 93–126. Springer.

[105] Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry.* Princeton university press.

[106] Lorini, E. (2020). Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282:103233.

[107] Mai, J. (2023). Realizing impossibilities. *Synthese*, 203(1):8.

[108] Makinson, D. (1987). On the status of the postulate of recovery in the logic of theory change. *Journal of Philosophical logic*, pages 383–394.

[109] Makinson, D. (2005). *Bridges from classical to nonmonotonic logic*. King's College.

[110] Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 25(6):205–207.

[111] Martin, E. and Osherson, D. (1998). Belief revision in the service of scientific discovery. *Mathematical social sciences*, 36(1):57–68.

[112] Masterton, G., Zenker, F., and Gärdenfors, P. (2017). Using conceptual spaces to exhibit conceptual continuity through scientific theory change. *European Journal for Philosophy of Science*, 7:127–150.

[113] Minsky, M. (1975). A framework for representing knowledge. *The psychology of computer vision*, pages 211–277.

[114] Nelson, D. (1949). Constructible falsity. *The Journal of Symbolic Logic*, 14(1):16–26.

[115] Odintsov, S. and Wansing, H. (2021). Routley star and hyperintensionality. *Journal of Philosophical Logic*, 50(1):33–56.

[116] Olsson, E. J. (2007). Lindström and rabinowicz on relational belief revision. *Hommagea Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.*

[117] Olsson, E. J. and Enqvist, S. (2010). *Belief revision meets philosophy of science*, volume 21. Springer Science & Business Media.

[118] Park, W. (2010). Belief revision vs. conceptual change in mathematics. In *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, pages 121–134. Springer.

[119] Perry, J. (1986). From worlds to situations. *Journal of Philosophical Logic*, 15:83–107.

[120] Perry, J. and Barwise, J. (1983). *Situations and attitudes*, volume 181. MIT Press Cambridge, MA.

[121] Peter, G., Hans, R., et al. (1995). Belief revision. *Handbook of Logicin Artificial Intelligence and Logic Programming*, 4:35–132.

[122] Plaza, J. (2007). Logics of public communications. *Synthese*, 158:165–179.

[123] Popper, K. (2014). *Conjectures and refutations: The growth of scientific knowledge.* Routledge.

[124] Punčochář, V. (2017). Algebras of information states. *Journal of Logic and Computation*, 27(5):1643–1675.

[125] Punčochář, V. (2019). Substructural inquisitive logics. *The Review of Symbolic Logic*, 12(2):296–330.

[126] Punčochář, V., Bílková, M., Bozdag, S., and Ferguson, T. M. (2024). Consistency-sensitive epistemic modalities in information-based semantics. *Studia Logica*, pages 1–39.

[127] Punčochář, V. and Sedlár, I. (2022). Routley star in information-based semantics. In *Indrzejczak, A., Zawidzki, M. (eds.) Proceedings of 10th International Conference on Non-classical Logics: Theory and Applications. Electronic Proceedings in Theoretical Computer Science*, pages 285–297.

[128] Punčochář, V. and Tedder, A. (2021). Disjunction and negation in information based semantics. In *Logic, Language, Information, and Computation: 27th International Workshop, WoLLIC 2021, Virtual Event, October 5–8, 2021, Proceedings 27*, pages 355–371. Springer.

[129] Punčochář, V. (2017). Knowledge Is a Diamond. In Kennedy, J. and de Queiroz, R. J., editors, *Logic, Language, Information, and Computation*, pages 304–320, Berlin, Heidelberg. Springer.

[130] Ribeiro, M. M. and Wassermann, R. (2007). Base revision in description logics-preliminary results. In *Proceedings of the International Workshop on Ontology Dynamics (IWOD-07)*, volume 6982.

[131] Ribeiro, M. M. and Wassermann, R. (2009). Agm revision in description logics. *Proceedings of ARCOE*.

[132] Rott, H. (2001). *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Clarendon.

[133] Rott, H. (2009). Shifting Priorities: Simple Representations for Twenty-Seven Iterated Theory Change Operators. In Makinson, D., Malinowski, J., and Wansing, H., editors, *Towards Mathematical Philosophy: Papers from the Studia Logica conference Trends in Logic IV*, pages 269–296. Springer Netherlands, Dordrecht.

[134] Rott, H. and Pagnucco, M. (1999). Severe withdrawal (and recovery). *Journal of Philosophical Logic*, 28:501–547.

[135] Schurz, G. (2010). Abductive belief revision in science. In *Belief revision meets philosophy of science*, pages 77–104. Springer.

[136] Sedlár, I. (2021). Hyperintensional logics for everyone. *Synthese*, 198(2):933–956.

[137] Segerberg, K. (1991). The basic dynamic doxastic logic of agm. In *The Goldbaltt Variations*, volume 1, pages 76–107. Uppsala Prints and Preprints in Philosophy.

[138] Segerberg, K. (1995). Belief revision from the point of view of doxastic logic. *Logic Journal of the IGPL*, 3(4):535–553.

[139] Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change, and statistics: Proceedings of the Irvine Conference on Probability and Causation*, pages 105–134. Springer.

[140] Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.

[141] Stalnaker, R. (1986). Possible worlds and situations. *Journal of Philosophical Logic*, 15:109–123.

[142] Stalnaker, R. (1991). The problem of logical omniscience, i. *Synthese*, pages 425–440.

[143] Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 128(1):169–199.

[144] Strößner, C. (2020). Predicate change: a study on the conservativity of conceptual change. *Journal of Philosophical Logic*, 49(6):1159–1183.

[145] Sylvan, R. (1988). Relevant containment logics and certain frame problems of ai. *Logique et Analyse*, pages 11–25.

[146] Thagard, P. (1984). Frames, knowledge, and inference. *Synthese*, 61:233–259.

[147] Thagard, P. (1988). *Computational Philosophy of Science*. MIT Press.

[148] Thagard, P. (1990). The conceptual structure of the chemical revolution. *Philosophy of Science*, 57:183–209.

[149] Thagard, P. (1993). *Conceptual Revolutions*. Princeton University Press.

[150] Thagard, P. (2002). *Coherence in thought and action*. MIT press.

[151] Thagard, P. and Findlay, S. (2010). Changing minds about climate change: Belief revision, coherence, and emotion. In *Belief revision meets philosophy of science*, pages 329–345. Springer.

[152] Thomas, K. (1977). Objectivity, value judgment, and theory choice. *Thomas Kuhn, The Essential Tension: Selected Studies in Scientific Tradition and Change*, pages 320–339.

[153] Väänänen, J. (2007). *Dependence logic: A new approach to independence friendly logic*, volume 70. Cambridge University Press.

[154] Van Benthem, J. (1995). Logic and the flow of information. In *Studies in Logic and the Foundations of Mathematics*, volume 134, pages 693–724. Elsevier.

[155] Van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17:129–155.

[156] Van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge University Press.

[157] Van Benthem, J., Duque, D. F., and Pacuit, E. (2012). Evidence logic: A new look at neighborhood structures. In *Advances in Modal Logic AiML 9*, pages 97–118. London College Publications.

[158] Van Benthem, J., Fernández-Duque, D., and Pacuit, E. (2014). Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic*, 165(1):106–133.

[159] Van Benthem, J. and Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17:157–182.

[160] Van Benthem, J. and Pacuit, E. (2011). Dynamic Logics of Evidence-Based Beliefs. *Studia Logica*, 99:61–92.

[161] Van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthese*, 147(2):229–275.

[162] Van Ditmarsch, H., Van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*, volume 337. Springer Science & Business Media.

[163] Van Fraassen, B. C. (1969). Facts and tautological entailments. *The Journal of Philosophy*, 66(15):477–487.

[164] Van Linder, B., Van der Hoek, W., and Meyer, J. J. C. (1995). Actions that make you change your mind. In Wachsmuth, I., Rollinger, C.-R., and Brauer, W., editors, *KI-95: Advances in Artificial Intelligence*, pages 185–196, Berlin, Heidelberg. Springer.

[165] Vardi, M. Y. (1986). On epistemic logic and logical omniscience. In *Theoretical aspects of reasoning about knowledge*, pages 293–305. Elsevier.

[166] Wansing, H. (2017). Negation. *The Blackwell guide to philosophical logic*, pages 415–436.

[167] Wansing, H. (2024). One heresy and one orthodoxy: On dialetheism, dimathematism, and the non-normativity of logic. *Erkenntnis*, 89(1):181–205.

[168] Wansing, H. and Belnap, N. (2010). Generalized truth values. a reply to dubois. *Logic Journal of IGPL*, 18(6):921–935.

[169] Williamson, T. (2024). *Overfitting and Heuristics in Philosophy*. Oxford University Press.

[170] Wolter, F. and Zakharyaschev, M. (1999). Multi-dimensional description logics. In *IJCAI*, volume 99, pages 104–109. Citeseer.

[171] Yablo, S. (2014). *Aboutness*. Carl G. Hempel lecture series. Princeton University Press.

[172] Yalcin, S. (2018). Belief as question-sensitive. *Philosophy and Phenomenological Research*, 97(1):23–47.

[173] Yalcin, S. (2021). Fragmented but rational. In *The fragmented mind*, pages 156–179. Oxford University Press Oxford.

[174] Yang, F. and Väänänen, J. (2016). Propositional logics of dependence and independence. *Annals of Pure and Applied Logic*, 167:557–589.

[175] Özgün, A. and Berto, F. (2021). Dynamic hyperintensional belief revision. *The Review of Symbolic Logic*, 14:766–811.