Explicit belief, Justification terms and Quasi-consistent evidence

MSc Thesis (*Afstudeerscriptie*)

written by

Joris A. Galema

under the supervision of **Dr. Alexandru Baltag** and **Dr. Malvin Gattinger**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:
August 26, 2025

Members of the Thesis Committee:
Dr. Marianna Girlando
Dr. Mina Young Pedersen
Dr. Aybüke Özgun (chair)

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Logical omniscience is widely discussed within formal epistemology. An agent is said to be logically omniscient if their knowledge and belief is closed under logical consequence. Such a closure may be an ideal to strive for but, in the interest of modeling realistic agents, is something we may want to avoid. Standard epistemic logics that use *possible worlds* often have this problem due to inherent properties of such possible worlds semantics.
However, such models may also be interpreted differently. As opposed to these beliefs being actual *explicit* beliefs, one can interpret them as potential *implicit* beliefs.

When making a distinction between explicit and implicit belief one can take explicit beliefs as fundamental and derive implicit beliefs from these. This can be taken a step further by not merely taking explicit beliefs as fundamental but by considering *explicit evidence* as fundamental. From these we can then derive explicit belief and in turn implicit belief.

Such explicit evidence can come in the form of formulas, but this can lack expressivity that we may want with the intent of modeling non-omniscient agents. Instead, such explicit evidence may come in the form of justification terms to overcome these limitations.

In this thesis we take inspiration from an existing explicit evidence model and enrich it with justification terms that are able to encode derivations. In this manner we obtain more fine-grained models that are able to provide justifications on how agents' obtained gained beliefs and update their beliefs on newly acquired evidence.

Furthermore, we provide an axiomatization for *the logic of quasi-consistent belief (QCB)* and show completeness.

Acknowledgments

Before delving into this work that has become my thesis for the *Master of Logic* (in short ”MoL"), I want to take a moment and devote a section to the people, the places and the moments that have helped me throughout the process of writing this thesis. Without you all of this would not have been possible.

First of all, a thanks to my supervisors Malvin and Alexandru. I have learned so much from working together with the both of you. I fondly look back on those meetings I had with Malvin where we were trying to get through a certain proof and every alley that we took to try to solve it came with other complications. Even though we did not always get through the proof, I do remember laughing a lot and having a nice time brainstorming approaches to tackle the problems. I have to thank Alexandru for the ideas and input on his side. Many ideas that the thesis explores originates from suggestions and interesting leads that Alexandru gave. These topics still entertain my mind daily and for that I am very thankful. With that I would also like the members of Thesis committee who read my thesis and questioned me during my defense. I really enjoyed the defense and the questions and conversations that occurred during.

Apart from supervisors and teachers and all the experts at the ILLC who study topics that interest me so much, there is another very important component of the MoL: the fellow students you meet throughout the program.
It has been so delightful to meet so many people interested in the same thing, but each from a whole different perspective. The countless conversations (which were usually in terms of hypotheticals) put a smile on my face when I think back on them. To my closest friends here: Frank, Laura, Antoine, Horia, Radu, Lamia, Ioli. You have made the experience of doing the MoL so much fun and just for being able to meet you guys I know that the MoL has been a good decision for me.

Last, but certainly not least, I want to thank the lovely lovely staff of Crea cafe: my (at that time and unfortunately not anymore) amazing colleagues. Given that most of this thesis was written at the Crea staff table, I have to thank you all for having been there with me and for me throughout this experience.
Most of you did not have the slightest idea of what I was actually working on. I remember all the looks of confusion that I got anytime someone looked at my laptop screen. The fact that the reaction was each time the same still makes me chuckle when I think about it.
But not knowing what I was doing did not matter. Sitting there at the staff table whilst trying to solve a proof and being able to talk to whomever was behind the bar made the whole process so pleasant. And whether it was someone playing geoguesser or worlde behind the bar, or it was Andy designing and testing a boat, or it was just talking about whatever has been going on in eachother's lives, there was always something going on at Crea that made me happy to be working there.
And of course, the lots of croissants and coffees that kept me fed and caffeinated also have to be mentioned.

A special thank you to party gang: Sofía, Eva, George, Andreas, Andreas and Luca. You all are evidence that distractions and unwinding every now and then are very necessary components to keeping your head cool when stuck in difficult proofs. I cannot wait until we are reunited to create more versions of Jenga.

Thank you all.

Contents

Introduction

The problem of logical omniscience [Hin79, Sta91] has been relevant in the field of epistemic logic ever since Hintika's introduction of a *possible worlds* interpretation for epistemic logics [Hin62].

We say that an agent is *logically omniscient* whenever their belief and/or knowledge is closed under logical consequence.

Currently, epistemic logics and evidence logics that are most common use a notion of possible-worlds and relations over these worlds to derive what is known for an agent and what is believed.
Informally: knowledge of ϕ here is defined as "in all worlds that I consider possible, ϕ is the case". In a similar way, belief is defined as "in all worlds that I think may be the actual world, ϕ is the case."
Knowledge and belief are given modal operators K and B respectively, which given a corresponding relation over possible worlds are defined as necessity modal operator $\Box$ [BDRV01].

Logical omniscience comes as a result of inherent properties of such possible-world semantics and normal modal logics. Specifically, the rule of Necessitation tells us that, for modal operator $\Box$, whenever ϕ is a logical validity, then $\Box\phi$ is as well. Moreover, the K-axiom tells us that $\Box(\phi \to \psi) \to \Box\phi \to \Box\psi$, establishing modus ponens across possible worlds.
When $\Box$ is to be interpreted as knowledge (K) or belief (B) this results in our agents being logically omniscient as they know/believe all logical validities by necessitation and in turn know/believe anything that follows from their knowledge/beliefs via the K-axiom.

Logical omniscience may be an ideal we wish to strive for, but in the interest of modeling realistic agents is something we may wish to avoid. As it is perfectly realistic for an agent to believe some proposition p without believing all logical consequences of this proposition p. Related to this is the hyperintensionality paradox [Cre75, BN23]. Here it is stated that it is perfectly reasonable for an agent to believe p without believing q, even if p and q are logically equivalent. In a realistic non-omniscient setting we would like our beliefs to be *hyperintensional*, meaning that they allow for situations where we may believe p and not q, even when they are logically equivalent. Standard notions of knowledge and belief using possible worlds do not allow for such situations.

To overcome logical omniscience several solutions have been proposed. For a review see for instance [Sol17]. Among those a solution by Hintika [Hin79] with the inclusion of *impossible worlds*. This allows, among the set of possible worlds, some worlds that are 'impossible' meaning that they need not make all valid formulas true.
But as pointed out by Vardi [Var86], such a method does not necessarily solve logical omniscience, it merely establishes logical omniscience for a different non-standard logic where rules of deduction are different.

There is another way of looking at the problem. We can interpret the type of belief defined on relational models not as *actual* belief, but as *potential* ones. To put it in different words: If an agent were to actively believe the actual world to be one where ϕ is true, and ϕ were to logically entail ψ, then whether or not the agent is aware of this entailment, they will believe the actual world to be one where ψ happens to be true.

These two forms of belief Levesque calls *explicit* and *implicit* belief [Lev84]. Explicit belief is an actual non-omniscient hyperintensional form of belief, and an implicit belief is a kind of potential logically omniscient belief.

This distinction between explicit and implicit belief has lead to several accounts incorporating this distinction [Lev84, FH87, Lor20].

For instance: the logic of general awareness [FH87] considers a standard relational model but adds an *awareness* function which, given world w, tells us what formulas an agent is *aware* of. The agent then has an explicit belief in some formula ϕ iff they are aware of ϕ and they have an implicit belief in ϕ, where an implicit belief is the standard modal notion of belief. Note that this means that explicit beliefs are defined in terms of implicit beliefs. As a result explicit beliefs cannot be inconsistent, as there are no possible worlds that are inconsistent. Again, it is an ideal to have consistent beliefs, but things like the Lottery/Preface paradox [Nel00] show us that this on itself may not be entirely realistic as well.

Instead of defining explicit belief in terms of implicit belief, we can also turn this around. We can take explicit beliefs as fundamental and derive possible worlds and, in turn, implicit beliefs from these. As Vardi puts it:

> "First, this approach leaves the notion of possible world as a primitive notion, and it does not give us any intuition about the nature of these worlds. ... Secondly, since the above approach does not elaborate on on the issue of the nature of the possible worlds, it also leaves open the question of where one gets the set W of possible worlds in the first place." [Var86]

One approach that takes this avenue of defining implicit beliefs in terms of explicit beliefs is the *logic of doxastic attitudes* [Lor20]. Here we start with a set of explicit beliefs called a *belief base*. From here we compute doxastic alternatives (i.e. possible worlds) by considering all other (belief base, valuation) pairs that satisfy all of our explicit beliefs. If some formula ϕ is true in all of these then we have an implicit belief in ϕ.

Apart from tackling the conceptual problem of logical omniscience, this method also tackles complexity problems of model checking for epistemic logics.

A similar idea is proposed by Gogoladze [Gog16], but enhanced to incorperate *explicit evidence*. Here we start with a set of explicit evidence from which explicit belief is computed. This computation of explicit belief from explicit evidence is based on a notion of *quasi-consistency* which says that a set of formulas may be inconsistent, it can just not allow for explicit contradictions.
The approach is aimed at modeling non-omniscient agents that may become aware of inconsistencies in their evidence and update their beliefs accordingly. However, certain peculiar examples (highlighted in section 1.3) suggest that some way of encoding derivations that the agents has made may be useful to track where contradictions originate from.

Our main new contribution in this thesis is the use of justification terms [AF19, AFS24] to allow for a way to track derivations agents have made. Justification logic is a field of logic originating from the logic of proofs [Art95, Art01] which in turn originates from Gödel's work on providing *provability semantics* for *intuitionistic logic* [Göd33, Göd38].
Justification logic was introduced into epistemic logic by taking modal formula $\Box\phi$ which, in a provability interpretation, denotes "there is a proof for ϕ", and replacing it with an explicit version $t : \phi$ denoting that t is a proof/justification for ϕ.
Such proofs can then be built up by, for instance, *application*: if we have a proof $t_1 : (\phi \to \psi)$ and a proof $t_2 : \phi$ then we can construct proof $t_1 \cdot t_2 : \psi$. This can be seen as an explicit variant of the K-axiom.

This thesis concerns itself with combining these frameworks. The idea is to add evidence in the form of justification terms into a belief base framework, resulting in an *evidence base*.

Implicit beliefs here are derived from explicit beliefs, which themselves are derived from explicit evidence using a similar computation as in [Gog16].
We show that this addition of justification terms results in more fine-grained models able to model agents becoming aware of certain contradictions, and able to express why an agent believes what they believe i.e. what their justification for their belief is.

The layout of the thesis is as follows:

- In chapter 1 we cover related work and recommended prior reading in more detail. This consists of a more detailed explanation on epistemic logic and the problem of logical omniscience (section 1.1), an explanation of explicit vs. implicit belief and epistemic logic via belief base (section 1.2), a detailed explanation on Gogoladze's Explicit Evidence models where we cover motivating examples for this thesis (section 1.3), and a section on justification logic and justification terms (section 1.4).

- In chapter 2 we cover the proposal of the current thesis: adding justification terms to this explicit belief calculation. We will first cover the justification terms we will be considering in section 2.1, after which we define our notion of evidence base and *maximal quasi-consistent (Qmax) evidence* in section 2.2. In sections 2.3 to 2.5 we discuss notions of belief using evidence bases and qmax evidence. We finish the chapter with a conclusion in section 2.6. Here we also discuss difficulties introducing notions of belief into the formal language. Which motivates changes made in the next chapter.

- In chapter 3 we discuss a formal logic describing explicit and implicit belief. First, in section 3.1, we recall difficulties with formally describing defined notions of explicit and implicit belief and propose the solution of *certification*.
  This certification requires us to make some changes to the mechanisms described in chapter 2. With that we propose the logic of quasi-consistent belief (QCB) in section 3.2 and show completeness w.r.t. finite evidential base models.
  The before mentioned changes we come back to in chapter 4 as they show signs of undesirable behavior regarding the goals set out in this thesis.

- Finally, we conclude and discuss future work in chapter 4.

Chapter 1

Background

This chapter is devoted to supplying some context for the main work of the thesis. It serves as a reference for the main contributions of this thesis.

We will start the chapter in section 1.1 with a brief overview of epistemic logic, most common practices of it, and the problem of logical omniscience.

Then we consider approaches to tackle this problem of logical omniscience using a notion of explicit belief and implicit belief. Moreover, we consider an approach which uses belief bases as opposed to possible worlds, thus using explicit belief as fundamental. This is given in section 1.2

In section 1.3, we cover an approach which shares a lot of similarity to the belief base approach, but instead of having explicit belief as fundamental, it uses explicit evidence as fundamental. From these we can compute explicit beliefs and in turn also implicit belief. In this computation of belief we show peculiar examples which raise questions on the flexibility and expressability of these models using explicit evidence. These examples serve as the motivation for the formal framework we develop in later chapters.

To finish the chapter, section 1.4 gives a brief overview of justification logic and justification terms. These serve as the main component to solve the issues raised in section 1.3.

1.1 Epistemic logic and the problem of logical omniscience

The study of epistemic logic concerns itself with the question of how to formally define and reason about *knowledge* and *belief* [FHMV04, RSW25, GG06]. The study of epistemic logic dates back all the way to Aristotle but it was in Hintikkas work [Hin62] on Knowledge and belief that epistemic logic was given an interpretation based on *possible worlds*. To this day, the most common formal accounts of epistemic logic originate in one way or the other from Hintikka's work on possible worlds for knowledge and belief.

A big topic of conversation withing the field of epistemic logic however is the problem *logical omniscience* [Hin79, Sta91]. This states that, whenever an agent knows/believes some formula ϕ or set of formulas Φ and we have that $\phi \vdash \psi$ ($\Phi \vdash \psi$) then the agent knows/believes ψ. i.e. the agents' set of knowledge/beliefs is closed under deduction.
This is of course an ideal situation which we would want to strive for. But in the interest of modeling *realistic* agents it is something we want to avoid.

In this section we will cover some basic aspects of epistemic logics using possible worlds semantics, and will discuss te problem of logical omniscience and how it emerges in formal settings based on possible worlds.

1.1.1 Epistemic logic and possible worlds

As noted before, Hintikas work on knowledge and belief [Hin62] initiated the possible-worlds formalism for epistemic logic, which has been the basis of many epistemic logics ever since.

In such an approach we interpret knowledge and belief as modal operators K and B respectively. Then for some formula ϕ we have that $K\phi$ denotes that an agent *knows* ϕ and $B\phi$ denotes that an agent *believes* ϕ.

Such modal formulas are then to be interpreted on *relational models* commonly referred to as *Kripke models*[1]

> **Definition 1.1.1.** *(Kripke model)*
>
> A Kripke model is a tuple $M = (W, R, V)$ where W is a non-empty set of *possible worlds*, $R \subseteq W \times W$ is an *accessibility relation*, and $V : Prop \rightarrow \mathcal{P}(W)$ is a *valuation function*.

The relation R acts as an accessibility relation between worlds in W specifying what we consider to be *possible*.
Intuitively, if we consider knowledge, then wRv and wRu tells us that, from world w, we cannot tell whether the *actual world* is v or u. i.e. we cannot distinguish world v from world u.
If we consider belief then wRv and wRu tells us that we *believe* u or v (or any other world that we "see" from w) to be the actual world. Throughout the thesis we will often call such u and v *doxastic alternatives*.
The valuation function then tells us what propositions in $Prop$ are true in which worlds. For now, let us consider R to be an accessibility relation modeling an agents' knowledge. Let's say p is true in some v s.t. wRv but is false in some u s.t. wRu, then we do not *know* whether p is true or false since we cannot distinguish between some worlds where they have different truth values i.e. we consider it possible that p is true and we consider it possible that p is false.
Now what about when q is true in both/all such worlds. Then we *do* know q as it is true in all worlds that we consider possible to be the actual world.

This notion of truth for knowledge and/or belief of/in ϕ corresponds to truth of modal operator $\Box\phi$ which is the following:

> **Definition 1.1.2.** *(Possible worlds semantics for $\Box$)*
>
> For Kripke model $M = (W, R, V)$ we have that:
>
> $$M, w \models \Box\phi \text{ iff } M, v \models \phi \text{ whenever } wRv$$
>
> If R is to represent knowledge we write K for $\Box$. if R is to represent belief then we write B for $\Box$.

Notable is that Knowledge and Belief have the same semantics. The only difference is what relation over the set of worlds is used. The differences between knowledge and belief then come from the conditions set on this relation.
For instance, it is commonly assumed that knowledge is *veridical* i.e. if you know something

[1] For more details on kripke models and modal logic as a whole we refer to [BDRV01].

then it is true i.e. $K\phi \rightarrow \phi$. This condition corresponds to reflexivity of R and is characterized by modal **T** axiom.
This veridicality of course need not be the case for belief. However, commonly in its place is *consistency* which says beliefs are consistent i.e. $\neg B\bot$. This condition then corresponds to seriality of R which is characterized by modal **D** axiom.
These are just two examples of properties commonly discussed concerning knowledge and belief.

A popular topic of discussion is that of positive and negative introspection which states:

(+int) If you know/believe ϕ then you know/believe that you know/believe ϕ
i.e. $K\phi \rightarrow KK\phi$ and $B\phi \rightarrow BB\phi$

(-int) If you do not know/believe ϕ then you know/believe that you do not know/believe ϕ
i.e. $\neg K\phi \rightarrow K\neg K\phi$ and $\neg B\phi \rightarrow B\neg B\phi$

Positive introspection corresponds to transitivity of R (modal axiom **4**), whilst negative introspection corresponds to R being euclidean (modal axiom **5**).

The difference between knowledge and belief thus amounts to frame conditions[2] for accessability relation R. Commonly, knowledge is desribed via an **S5** (reflexive, transitive, euclidean) accesibility relation whilst belief is desribed through a **KD45** (serial, transitive, euclidean) accesibility relation.
Of course, there are different accounts and objections to knowledge and belief having these properties.

Worth noting however, (as it is relevant for next section), is the following:

Fact 1.1.3. Any normal modal logic (which includes **S5** and **KD45**) has the following:

- Necessitation: From $\vdash \phi$ infer $\vdash \Box\phi$
- K-axiom: $\Box(\phi \rightarrow \psi) \rightarrow \Box\phi \rightarrow \Box\psi$[3]

This covers the basics of modal epistemic logic. Of course, there is a lot more work on modal epistemic logic including multi-agent extensions [Leh84, HM90, FHMV04] , dynamic extensions [BMS16, BR16] and adaptions for belief revision [vB07, BS08].

However, we may take a step further into Generalization. As opposed to relational Kripke models we can consider neighborhood models as a generalization of Kripke models.

> **Definition 1.1.4.** *(Neighborhood model)*
>
> A neighborhood model is a tuple $M = (W, E, V)$ consisting of non-empty set of worlds W, neighborhood function $E : W \rightarrow \mathcal{P}(\mathcal{P}(W))$ and valuation function $V : Prop \rightarrow \mathcal{P}(W)$

Instead of having relation between worlds we now have a *neighborhood function* which assigns to each world a family of subsets of W. In a way such a function more or less *highlights* worlds where certain propositions may be true or false.

In an epistemic sense, such a function can then be interpreted as an *evidence* function where each $e \in E(w)$ is interpreted as a piece of evidence that highlights all worlds that are consistent with the evidence. These pieces of evidence need not be consistent with

[2] Again, for more details on such frame conditions and corresponding modal logics we refer to [BDRV01].
[3] K- axiom knows many forms, all equivalent to eachother.

eachother of course. Such neighborhood models are the basis of evidence logics [vBP11]. Here they are called evidence models and have conditions of their own.

Evidence based belief is then defined in terms of *maximally consistent sets of evidence*. Intuitively, an agent has an evidence based belief in a formula ϕ when all their maximally consistent sets of evidence are pieces of evidence for ϕ.
Such evidence models can be enhanced further into toplogical evidence models [BBÖS16, BBG19].

We finish this section with an example that will motivate the next section.

Example 1.1.5. Suppose we have some Kripke model $M = (W, R, V)$ s.t. $M, w \vDash B\phi$ i.e. an agent believes ϕ. Then any $v \in W$ s.t. wRv has that $M, v \vDash \phi$.
Now consider some formula ψ such that ϕ logically entails ψ. Any $v \in W$ s.t. $v \vDash \phi$ then has to have that $v \vDash \psi$. Then all $v \in W$ s.t. wRv also have that $v \vDash \psi$. I.e. we have $M, w \vDash B\psi$.

This example shows us that these notions of belief are closed under deduction.

1.1.2 Logical Omniscience

We saw in example 1.1.5 that the notion of belief defined on relational models is closed under deduction. In an ideal world this would be nice. However, in a realistic scenario such a closure under deduction may not be entirely realistic. As Hintika points out as well:

"it is clearly inadmissible to infer "he knows that q" from "he knows that p" solely on the basis that q follows logically from p, for the person in question may fail to see that p entails q, particularly if p and q are relatively complicated statements"[Hin62].

Moreover, things like lottery/preface paradox [Nel00] show us that our beliefs may not only be not closed under deduction, but may be completely inconsistent by themselves.

This closure under deduction for beliefs is commonly referred to as *logical omniscience* [Hin79, Sta91]. An agent is said to be *logically omniscient* iff:

* Whenever an agent knows/believes some formulas Φ and ψ logically follows from Φ ($\Phi \vdash \psi$) then the agent knows/believes ψ.

As a result of this we have that any omniscient agent knows/believes:

- all tautologies/theorems since these can be derived from an empty set of assumptions
- all logical consequences of other known/believed formulas

Recall fact 1.1.3 telling us that any normal modal logic has Necessitation and K-axiom. Notice how the first of above consequences relates to Necessitation of belief. As we have that, any ϕ s.t. $\vdash \phi$ has $\vdash B\phi$. Moreover, the second consequence relates to K-axiom as well. Let us say we have some belief in ϕ i.e. $B\phi$. Now take any ψ s.t. ψ logically follows from ϕ i.e. $\phi \vdash \psi$ i.e. $\vdash \phi \rightarrow \psi$.
By necessitation of B we get that $B(\phi \rightarrow \psi)$ and since we have $B\phi$ we get, by K-axiom, $B\psi$.

Indeed, the K-axiom ensures modus ponens across accessible worlds and thus leads to closure under deduction across possible worlds. Which in turn, given the considered modal semantics of belief, leads to logical omniscience. So much so that it has been called "omniscience" axiom as well [vB19].

1.2 Explicit belief and Belief base

In this section we will consider some possible approaches to tackle logical omniscience in epistemic logic. All the approaches we will consider are reliant on the distinction between *explicit* and *implicit* belief. The first approaches will still rely on possible-worlds semantics. Afterwards we will take an alternative approach which uses *belief bases*. Belief base semantics will be similar to our proposal so we will cover them in a bit more detail.

1.2.1 Explicit and Implicit belief

As Levesque points out

> ".. all formalizations of belief based on a possible-world semantics suffer from the fact that at any given point, the set of sentences considered to be believed is closed under logical consequence." [Lev84]

Levesque suggests that the type of belief that such relational models typically formalize is not one of *actively held* beliefs, but rather one of *potential* ones. If an agent were to actively believe the actual world to be one where ϕ is true, and ϕ were to logically entail ψ, then whether or not the agent is aware of this entailment, they will believe the actual world to be one where ψ happens to be true.
The agents' belief in ϕ Levesque calls an *explicit* belief, whereas the agents' belief in ψ is called an *implicit* belief.
With this distinction between explicit and implicit belief Levesque aims to tackle the problem of logical omniscience.

Fagin and Halpern build further on this intuition that beliefs may be divided into explicit non-omniscient ones and implicit omniscient ones. They suggest that a notion of *awareness* is relevant towards finding out which formulas are explicitly believed and what formulas are implicitly believed [FH87]. With this they propose a *Logic of General Awareness*. In essence this approach takes a possible-worlds relational model $M = (W, R, V)$ and enriches it with a syntactic *awareness* function $A : W \to \mathcal{L}$.
An implicit belief $L\phi$ then corresponds to the familiar definition of ϕ being true in all worlds that the agent considers possibile i.e.

$$M, w \vDash L\phi \text{ iff } M, v \vDash \phi \text{ whenever } wRv$$

An explicit belief $B\phi$ then corresponds to an implicit belief in ϕ together with being aware of ϕ i.e.

$$M, w \vDash B\phi \text{ iff } \phi \in A(w) \text{ and } M, w \vDash L\phi$$

It follows that this notion of explicit belief is not necessarily closed under deduction. We may for instance have that $\phi \to \psi$ and we are unaware of a formula ψ whilst having an explicit belief in ϕ, causing us to not explicitly believe ψ.

As Fagin and Halpern note however, this notion of explicit belief is not allowed to be inconsistent. The reason for this is that the notion of explicit belief relies on what is implicitly believed and this relation R that is used to find what is implicitily believed is assumed to be KD45, i.e. *transitive*, *serial* and *euclidean*. The fact that R is serial means that for any $w \in W$ we have that there is some $v \in W$ such that wRv.
Say we want to model an agent having an inconsistent explicit belief like $p \wedge \neg p$. This would require $L(p \wedge \neg p)$ which would mean that any v s.t. wRv has $v \vDash p \wedge \neg p$. But this can never be the case. So inconsistent explicit beliefs are impossible due to implicit beliefs not being able to be inconsistent.

In the models we have considered so far the set of possible worlds and the relation for doxastic alternatives are fundamental in the model, essentially telling us that our implicit beliefs are more fundamental than our explicit beliefs. Would we not want to model it the other way around? Would we not want that our explicit beliefs, the actual held beliefs, are fundamental and tell us what we implicitly believe?

1.2.2 Belief bases

As Lorini notes "The concept of explicit belief is tightly connected with the concept of belief base" [Lor20]. In this paper, Lorini proposes the *Logic of Doxastic Attitudes* LDA which formalizes a notion of explicit and implicit belief based on belief bases as opposed to relational semantics. These belief bases consist of all formulas that are explicitly believed by an agent.
In this way we take explicit beliefs to be fundamental as opposed to implicit beliefs and doxastic alternatives.

The formalization in [Lor20] uses a definition of (multi-agent) Belief base[4]

Definition 1.2.1. *(Multi-agent belief base)*

A (multi-agent) belief base is a tuple $B = (B_1, ..., B_n, V)$ where each B_i is a set of formulas and V is the set of *true atomic propositions*

Each B_i consists of formulas α that are either formulas of propositional logic (p, $\neg\alpha$, $\alpha \wedge \beta$) or are formulas of the form $\triangle_i\alpha$ denoting that agent i explicitly believes α. Since belief bases consists of the explicit beliefs of agents, truth conditions for explicit belief in α simply come down to checking if α is in the belief base:

Definition 1.2.2. *(LDA-Explicit belief)*

For multi-agent belief base $B = (B_1, ..., B_n, V)$ we have:

$$B \models \triangle_i\alpha \text{ iff } \alpha \in B_i$$

Since there is no need for these B_i to be closed under any form of logical consequence we have that explicit beliefs are not closed under deduction. Moreover, these belief bases are not required to be consistent so we may have inconsistent explicit beliefs. This will result in implicit beliefs in everything as we will see later on.

Now, to define implicit belief we first need some notion of *doxastic alternative*. Which is defined as follows:

Definition 1.2.3. *(LDA doxastic alternatives)*

For any multi-agent belief bases $B = (B_1, ..., B_n, V)$ and $B' = (B'_1, ..., B'_n, V')$, we have:

$$B\mathcal{R}_iB' \text{ iff } B' \models \alpha \text{ for all } \alpha \in B_i$$

When $B\mathcal{R}B'$ we say that B' is a *doxastic alternative* at B.

Intuitively, we consider doxastic alternatives as those multi-agent belief bases that make all of our explicit beliefs true. In a way we can see such multi-agent belief bases as this formal setting's version of *possible world*.

[4] In this thesis we only consider the single-agent case, but here we consider Lorini's multi-agent version.

Implicit belief ($\Box_i\phi$) then follows the same intuition of "true in all doxastic alternatives" i.e.

> **Definition 1.2.4.** *(LDA-Implicit belief)*
>
> For multi-agent belief base B and set of multi-agent belief bases Ctx we have:
>
> $$(B, Ctx) \models \Box_i\phi \text{ iff } (B', Ctx) \models \phi \text{ whenever } B\mathcal{R}_iB' \text{ with } B' \in Ctx$$

This set Ctx is called the *context* and is meant to express the agents' common belief. If the multi-agent belief bases correspond to a notion of possible worlds here, then the context corresponds to the set of possible worlds. Such a (B, Ctx) pair is called a multi-agent belief model (MAB).
Note that, when Ctx is all belief bases then, intuitively, we have implicit belief in ϕ if it follows from our explicit beliefs i.e. logical closure of explicit beliefs. If Ctx is a proper subset then implicit belief corresponds to potential beliefs given explicit beliefs and common belief among agents. In other words: implicit belief corresponds to potential beliefs.

As noted before, explicit beliefs need not be closed under deduction and need not be consistent at all. Consider the following example:

Example 1.2.5. Consider an MAB model (B, Ctx) where $B_1 = \{p, \neg p\}$. Agent 1 explicitly believes both p and $\neg p$ (but does not explicitly believe the contradiction $p \wedge \neg p$).
No belief base $B' = (B'_1, ...B'_n, V')$ can have that $B' \models p$ and $B' \models \neg p$. Therefore, there is no B' such that $B\mathcal{R}_1B'$.
Now consider any formula ϕ. Implicitly, agent 1 will believe this formula since there is no Belief base B' such that $B\mathcal{R}_1B'$. So we have that $(B, Ctx) \models \Box_1\phi$. Implicitly agent 1 believes anything and everything.

Explicitly believing two contradictory things leads to implicitly believing everything. When we think again of implicit belief as potential beliefs i.e. everything that follows from our explicit beliefs, we see that this phenomenon corresponds to applying *ex falso* ($\bot \to \phi$) to our set of inconsistent explicit beliefs.

The reason we gain this implicit belief in everything is because there is no doxastic alternative at B. Which means that $\mathcal{R}_i$ is not serial. We can impose a consistency restriction on each belief bases which will result in $\mathcal{R}_i$ being serial. As an effect we will have that implicit belief is consistent. This would ensure that **D** axiom holds for $\Box_i$.

Lorini continues this line of thought with the possibility of extra conditions on $\mathcal{R}$ to ensure positive and negative introspection. If we add the condition that $B_i \subseteq B'_i$ whenever $B\mathcal{R}_iB'$ then we ensure transitivity of $\mathcal{R}$ i.e. positive introspection of implicit belief. Furthermore, we can add the requirement that $B_i \supseteq B'_i$ whenever $B\mathcal{R}_iB'$ to ensure that $\mathcal{R}_i$ is euclidean i.e. negative introspection.

The formal framework can then be extended further. For instance to express explicit and implicit common belief [LR22], to model belief revision [LS21] and modeling reinforcement learning algorithms [ELML+25].

1.3 Explicit Evidence models

In this section we will consider a logic that aims to incorporate the possibility of inconsistent beliefs and non-omniscient agents into evidence models. It bears a similarity as previous approach using belief base, as in that implicit beliefs are computed from explicit beliefs. Fundamental in both is a set of formulas which eventually are used to compute doxastic

alternatives, i.e. possible worlds in order to derive implicit beliefs.
Where it differs is that there is an extra layer of computation added which computes explicit belief from *explicit evidence*. This logic is called the *logic of explicit evidence* [Gog16]. The motivation for this logic is to deal with logical omniscience in a way that we may have contradictory beliefs, but we are able to fix these beliefs later when we have *become aware of the contradiction* in our belief.

Gogoladze [Gog16] makes a first attempt at doing so via *explicit belief models*. Here we consider *explicit beliefs* $\mathcal{B}$ as subsets of our language (similar to what we have seen in section 1.2.2) and it is required that $\mathcal{B}$ is *quasi-consistent* meaning that $\bot \notin \mathcal{B}$ i.e. we do not directly believe a contradiction.
This quasi-consistency condition is more relaxed than a *consistency* condition, allowing an agent to have inconsistent beliefs, for instance by having $p \in \mathcal{B}$ and $\neg p \in \mathcal{B}$, it is just not allowed for an agent to directly believe a contradiction.
However, Gogoladze then raises the question of how we could fix these inconsistent beliefs.

The solution [Gog16] proposes is the use of *explicit evidence models* which use a notion of *explicit evidence* as fundamental. Such explicit evidence is then used to compute explicit belief.
In these models we can express that an agent becomes aware of a contradiction in their evidence and computes their beliefs accordingly. In that way, we may have inconsitent beliefs and figure out that this is inconsistent, adding evidence and thus recomputing explicit beliefs at every step. Previously we would only be able to add or remove beliefs, now we can add evidence which may cause us to lose beliefs.

This thesis will take a lot of inspiration from the way explicit beliefs are derived from explicit evidence in [Gog16]. Hence we will delve more in depth into how such explicit evidence models work and will provide some defintions, explanations and examples.

To start we will define the language used in [Gog16].

Definition 1.3.1. *(Language: $\mathcal{L}_{EE}$)*

Given countable set of atomic propositions $Prop$, formulas $\phi \in \mathcal{L}_{EE}$ are inductively defined as follows:[a]

$$\phi := p \mid \bot \mid \phi \to \phi \mid B^e\phi \mid B^i\phi \mid K^e\phi \mid K^i\phi$$

Here p ranges over propositional formulas $Prop$, $\bot$ marks a contradiction and is never true, $\to$ denotes material implication, $B^e\phi$ and $B^i\phi$ denote explicit and implicit belief resp. Similarly $K^e\phi$ and $K^i\phi$ denote (resp.) explicit and implicit knowledge.

Moreover, we define shorthands for tautology $\top$, negation $\neg$, conjunction $\wedge$ and disjunction $\vee$ as follows:

$$\begin{aligned}\neg\phi :=& \phi \to \bot\\ \top :=& \neg\bot\\ \phi \vee \psi :=& (\neg\phi) \to \psi\\ \phi \wedge \psi :=& \neg(\neg\phi \vee \neg\psi)\end{aligned}$$

[a] Originally the language was defined using negation and conjunction. However, since $\bot$ is an important part of the logic and its way of dealing with contradictions we will use it as a primitive and define other connectives in terms of $\bot$ and $\to$.

The language $\mathcal{L}_{EE}$ is evaluated on *Explicit evidence models* which are defined as follows:

Definition 1.3.2. (*Explicit Evidence model*)

An explicit evidence model is a tuple $M = (W, W_0, E_s, E_h, \|.\|)$ where:

- W is a set of worlds.
- $W_0 \subseteq W$ is a set of worlds representing background beliefs (bias).
- $E_s \subseteq \mathcal{L}_{EE}$ is a set of formulas representing *soft* evidence
- $E_h \subseteq E_s$ is set of formulas representing *hard* evidence.
- $\|.\| : Prop \to \mathcal{P}(W)$ is a valuation function

We just have one extra requirement which is that $\top \in E_h$ from which it follows that $\top \in E_s$.

This requirement that $\top \in E_h$ will make sure that there is always one particular thing that the agent will always explicitly know and believe, that being that the world is consistent (or if you consider how the shorthand is defined, that false is indeed false).

One may then also expect that we should have a requirement for $\bot \notin E_s$ since evidence for a contradiction is rather unintuitive. This, however, is not the case. As it turns out, the possibility of $\bot \in E_s$ is essential for the logic to behave in the way that it intends to and achieves to.

To make this use of $\bot$ in explicit soft evidence clear we will cover *maximal closed quasi consistent evidence* next.

Definition 1.3.3. *(Q-max evidence)*

A set $e \subseteq E_s$ of soft evidence is called *maximal closed quasi consistent* (Q-max) w.r.t. E_s iff

1. $\bot \notin e$
2. if $\{\phi \to \psi, \phi\} \subseteq e$ and $\psi \in E_s$, then $\psi \in e$
3. $E_h \subseteq e$
4. if there is some $e' \subseteq E_s$ that satisfies conditions 1,2,3 and $e \subseteq e'$ then $e = e'$

Let us delve a bit deeper into this definition, as it is of high importance for this particular logic and it will serve as the inspiration for the explicit belief compuatation of the formal system set up in current work.

Our first requirement states that $\bot \notin e$, this [Gog16] calls *quasi-consistency*, which informally means that we are filtering out any contradiction from E_s with this computation.

Next we have that e is closed under modes ponens, but restricted to E_s. This restriction to E_s is essential as this is the key factor that allows us to still have inconsistent explicit beliefs which we are not aware of yet. Condition 3 tells us that all our hard evidence is also soft evidence which will result in explicit knowledge implying explicit belief. These two conditions together [Gog16] calls e being *closed*.

Lastly e should be maximal with respect to these properties.

Now suppose, for example, that we have some set of soft evidence E_s such that $\bot \notin E_s$. If we want to take some maximal subset of E_s such that it is closed (conditions 2 and 3) and quasi-consistent (condition 1), we see that we will always end up with E_s since it does not contain $\bot$ and it is the largest subset of E_s.

Fact 1.3.4. Whenever $\bot \notin E_s$, we have that E_s is the only Q-max evidence.

Then what if we do have $\bot \in E_s$? Starting with an easy example, let us say that $E_s = \{\top, \bot\}$ with $E_h = \{\top\}$. Condition 2 does not apply as we have nothing of the form $\phi \to \psi \in E_s$. Condition 1 will however apply and tells us that E_s is not q-max, and that we have to remove this element $\bot$. This leaves us with one maximal (w.r.t. the conditions) subset of e, that being $\{\top\} = E_h = E_s \setminus \{\bot\}$. The qmax computation has filtered out $\bot$ from E_s.
$E_s \setminus \{\bot\}$ being the only qmax set does not only apply this example. It is not hard to show that, whenever $\bot \in E_s$ but there is no $\phi \to \bot \in E_s$ with $\phi \in E_s$ as well, we will have that $E_s \setminus \{\bot\}$ is the only qmax evidence.

Getting somewhat more involved, what happens then if we have $\bot \in E_s$ and we *do* have some $\phi \to \bot \in E_s$ and $\phi \in E_s$? In this case we need to remove $\bot$ but we still have the condition that $\{\phi \to \bot, \phi\} \subseteq e$ and $\bot \in E_s$ implies that $\bot \in e$. This means that we have to remove $\phi \to \bot$ or ϕ as well.
This leaves us with two qmax evidence sets e_1 and e_2 where $\phi \in e_1$, $\phi \notin e_2$ and $\phi \to \bot \notin e_1$,$\phi \to \bot \in e_2$. We have essentially split up our evidence into two different subsets, one where we take ϕ to be true, one where we take ϕ to be false.

This is essentially the idea of how to get to our explicit beliefs. We are splitting up our set of evidence into subsets that are maximal in a way that we *think* they are consistent. Then if some formula occurs in all such qmax sets we can safely believe it since there is no reason not to.
So our explicit beliefs are defined as those formulae on which all our Q-max evidence agree on, i.e. the intersection of all our qmax sets.

Definition 1.3.5. *(Explicit belief)*

For a set of soft evidence E_s, the set of explicit beliefs $\mathcal{B}$ is as follows:

$$\mathcal{B} = \bigcap \{e \subseteq E_s \mid e \text{ is Q-max}\}$$

For explicit evidence model $M = (W, W_0, E_s, E_h, ||.||)$ we then have:

$$M, w \vDash B^e \phi \text{ iff } \phi \in \mathcal{B}$$

The author then goes on with defining implicit belief in terms of explicit belief. Similar to definition 1.2.4 we consider all possible worlds that satisfy all our explicit beliefs. If all of these worlds satisfy some formula ϕ then we have an implicit belief in ϕ.

Definition 1.3.6. *(Implicit Belief)*

For explicit evidence model $M = (W, W_0, E_h, E_s, ||.||)$ we have:

$$M, w \vDash B^i \phi \text{ iff } M, v \vDash \phi \text{ for } v \in W_0 \text{ s.t. } M, v \vDash \bigwedge \mathcal{B}$$

Again, the use of possible world semantics for implicit beliefs ensures logical omnisciency for implicit belief. Corresponding to the intuition that implicit beliefs are to be interpreted as potential beliefs which we may still uncover from our explicit beliefs.

Now, again similar as before in the logic LDA, our explicit beliefs are allowed to be inconsistent, which again results in implicit belief in everything.

Example 1.3.7. Suppose $E_s = \{\top, p, p \to \bot\}$. Then, since $\bot \notin E_s$, we have that E_s is itself the only QMax set.
Then $p \in \mathcal{B}$ and $p \to \bot \in \mathcal{B}$ which means we have inconsistent explicit beliefs.
Moreover, since $p \wedge (p \to \bot)$ is not satisfiable, there can be no possible world that satisfies all $\phi \in \mathcal{B}$. This results in all possible worlds that satisfy all $\phi \in \mathcal{B}$, which is none, to satisfy any arbitrary formula ψ.
I.e. we implicitly believe all formula ψ.

Notice the similarity of this example with previous example 1.2.5. As we expect, since the semantics of both are so similar (when we have that E_s is quasi-consistent) we get similar results.

Where the similarities end is in the addition of Qmax calculation. Which is jumpstarted by adding piece of evidence $\bot$ which signals that we have become aware of the contradiction in our evidence.

Example 1.3.8. Consider again example 1.3.7. Now let's say we become aware of the contradiction, i.e. we add $\bot$ to E_S.
$E_s' = \{\top, p, p \to \bot, p\}$. If we want to take qmax sets we again, but now we see that, since $\bot \in E_s'$, we will have to remove from things. We get qmax sets $\{\top, p\}$ and $\{\top, p \to \bot\}$. Taking the intersection of these results in $\mathcal{B} = \{\top\}$. We have lost our inconsistent explicit beliefs.

As we see here, adding piece of evidence $\bot$ results in the loss of our explicit beliefs. We have made our beliefs consistent again by realizing that our evidence contradicts.

However, let us now take this example a bit further.

Example 1.3.9. Let's say we are in the scenario of example 1.3.8. We have some evidence for $p \to \bot$ and p and we have realized that our evidence is contradictory. We have $E_s' = \{\top, p, p \to \bot, p\}$ resulting in $\mathcal{B} = \{\top\}$.
No suppose we gain some new pieces of evidence, q and $q \to \bot$. Then $E_s'' = \{\top, p, p \to \bot, \bot, q, q \to \bot\}$. Again we consider qmax sets to derive explicit beliefs. These are $\{\top, p, q\}, \{\top, p, q \to \bot\}, \{\top, p \to \bot, q\}$ and $\{\top, p \to \bot, q \to \bot\}$.
Then again we have that $\mathcal{B} = \{\top\}$.

Again, we have no inconsistent beliefs. But we have just added q and $q \to \bot$ to our evidence, without specifying that we are aware of this contradiction. If propositions p and q are completely unrelated it may be reasonable to still have inconsistent belief for q. This is something not expressible by the explicit evidence models as they are.

Another interesting example to consider is the following:

Example 1.3.10. Suppose we have some set of explicit (soft) evidence $E_s = \{\top, p, p \to q\}$ with $E_h = \{\top\}$. Now from this we realize that (after applying modes ponens) we have a belief in q
$E_s' = \{\top, p, p \to q, q\}$.
Now we gain some other piece of (soft) evidence that says $p \to \bot$. We apply MP and become aware of a contradiction p and $p \to \bot$: $E_s'' = \{\top, p, p \to q, q, p \to \bot, \bot\}$.
Since $\bot \in E_s$ we will have a different set of explicit beliefs than our soft evidence. This will be $\mathcal{B} = \{\top, p \to q, q\}$.

Now we have lost our belief in p but we keep our belief in q which we have acquired from our evidence for p which we have now disregarded. Should we still believe q?

We leave aside these examples now, but later in chapter 2 we provide a way to prevent this from happening.

1.4 Justification logic

Justification logic [Art08, AFS24, AF19] has roots in both formal epistemology and mathematical logic. It originates from the logic of proofs [Art95, Art01] which in turn originates from Gödels first contribution[5] to developing a proof based semantics for intuitionistic logic [Göd33, Göd38].

1.4.1 The Logic of Proofs

Gödel considered the modal logic **S4**[6] as a calculus describing provability [Göd33]. the modal operator $\square$ here gets interpreted as "there exists a proof for ..."

As Brouwer puts it, the notion of truth in intuitionistic logic of a formula ϕ corresponds to "there exists a proof for ϕ" [TVD14]. Heyting and Kolmogorov later extended this description with what is commonly refered to as Brouwer-Heyting-Kolmogorov (BHK) semantics [Hey31, Kol32].

In Gödel's work on provability logic [Göd33] a translation from IPC (intuitionistic Propositional Calculus) to S4 formulas is given. This translation is done by adding "provable" operator $\square$ before formulas as follows [7]:

$$
\begin{aligned}
tr(p) &= \square p \\
tr(\neg\phi) &= \square\neg tr(\phi) \\
tr(\phi \wedge \psi) &= tr(\phi) \wedge tr(\psi) \\
tr(\phi \vee \psi) &= tr(\phi) \vee tr(\psi) \\
tr(\phi \rightarrow \psi) &= \square(tr(\phi) \rightarrow tr(\psi))
\end{aligned}
$$

Gödel then shows that any IPC formula ϕ provable in IPC has that $tr(\phi)$ is provable in S4. Later McKinsky and Tarsky proved the converse as well [MT48] meaning we have:

$$IPC \vdash \phi \text{ iff } S4 \vdash tr(\phi)$$

The reading of $\square\phi$ now contains an existential quantification as it only expresses that there is *some proof* of ϕ as opposed to showing a specific proof. As Artemov notes as well "... difficulties with reading S4-modality $\square\phi$ as $\exists x$. Proof(x, y) are caused by the non-constructive character of the existential quantifier." [Art01]

Consider for instance the K-axiom:

$$\square(\phi \rightarrow \psi) \rightarrow \square\phi \rightarrow \square\psi$$

This, in Gödel's account, is interpreted as: if there is a proof that ϕ implies ψ and there is a proof that ϕ then there is a proof that ψ. But then, what is this proof of ψ?

Artemov later extended this[8] by making these proofs *explicit* [Art95, Art01]. In *the logic of proofs* (LP) [Art01] we, instead of having a modal operator $\square$, consider an operator $t : \phi$, where t is some term, called a *proof term*, encoding the proof for ϕ. In this way the proof is made explicit in the form of a proof term.

[5] One of the earliest work on provability logic was by Orlov [OS24], who suggested prefixing formulas with a provability operator.

[6] **S4** is transitive and reflexive. For more on the modal logic **S4** see for instance [BDRV01].

[7] This translation is also for for a translation from ASP to S4[ON04].

[8] As a matter of fact godel did as well in [Göd38], however lectures not published until after Artemov.

Artemov continues in his work to show that any valid **S4** formula ϕ has some *realization* r such that ϕ^r is a valid **LP** formula. Here a realization means an assignment of proof terms to modal operators $\Box$ occurring in ϕ.

This gives us the following:

$$IPC \vdash \phi \text{ iff } S4 \vdash tr(\phi) \text{ iff } LP \vdash tr(\phi)^r$$

(where $tr(\phi)^r$ is some *realization* of $tr(\phi)$)

The K-axiom then gets transformed into the following axiom, called *application*:

$$t_1 : (\phi \rightarrow \psi) \rightarrow t_2 : \phi \rightarrow (t_1 \cdot t_2) : \psi$$

As noted before, the K-axiom ensures closure under modes ponens across accessible worlds. However, since we are now dealing with explicit proof terms we circumvent this closure in a particular way.
To put it into different terms: we do not have that $t : (\phi \rightarrow \psi) \rightarrow t : \phi \rightarrow t : \psi$. We need an explicit derivation where we *combine* the terms that we have.

This extends to the *hyperintensionality paradox* [Cre75, BN23]. Here it is stated that it is perfectly reasonable for an agent to have a belief in some proposition p without having a belief in proposition q, even if p and q are logically equivalent.
Possible world semantics with modal operator $\Box$ however does not allow for this since we have $\phi \leftrightarrow \psi \vdash \Box\phi \leftrightarrow \Box\psi$.

In contrast, the use of explicit proof terms circumvents this since we do not have that $\phi \leftrightarrow \psi \vdash t : \phi \leftrightarrow t : \psi$. But let us say we have $\vdash \phi \leftrightarrow \psi$. Then we should be able to build a proof t of $\phi \leftrightarrow \psi$ i.e. $t : (\phi \leftrightarrow \psi)$. Now let's say we have a proof $t_1 : \phi$, then we should be able to construct proof $f(t, t_1) : \psi$ (i.e. $t_1 : \phi \rightarrow f(t, t_1) : \psi$). Similarly for when we have a proof $t_2 : \psi$.
Thus the logic of proofs, and more generally such proof terms, allow for the flexibility to still derive this equivalence explicitly.

1.4.2 The Logic of Justifications

The logic of proofs originally intended to connect IPC to classical proofs, but it was subsequently introduced into epistemic logic in the form of *justification logic* [Art08]. Here, the proof terms are interpreted as *justifcation terms* i.e. $t : \phi$ denotes t is a *justication* for ϕ.
In the original paper on justication logic it was used to tackle Gettier cases [Get63, Het16], contributing to the *justified true belief vs knowledge* debate.

In its simplest form justification logic comes as *basic justification logic* J_0. It is axiomatized by classical propositional axioms with rule modes ponens, application axiom (as above) and *sum axioms* which are the following: $t_1 : \phi \rightarrow (t_1 + t_2) : \phi$, $t_2 : \phi \rightarrow (t_1 + t_2) : \phi$.
Intuitively these sum axioms tell us that, whenever we have a justification for some formula ϕ we can add other justifications to it and they will still count as justifications for ϕ[9].

The logic of justification J is then J_0 with the *axiom internalization rule*, which is an explicit variant of the necessitation rule from normal modal logics, but restricted to axioms[10].

As we now have an explicit variant of K-axiom (i.e. application), and an explicit variant of Necessitation rule (i.e. axiom internalization), we can see that the logic of justification J really acts as an explicit counterpart to the smallest normal modal logic K.

[9] In our formal system discussed in chapters 2 and 3 we will not be considering sum terms. In chapter 2 we will however consider other terms called *weakening terms* which have similarities but lead to different behaviour.

[10] There are more nuances to this but these are not relevant for the current thesis. For more information see Artemovs original work[Art08]

Artemov continues this line of thought and shows correspondance results for several modal logics including **T** (adding axiom $t : \phi \to \phi$) **D** (adding axiom $\neg t : \bot$), **4** (adding axiom $t : \phi \to !t : (t : \phi)$ and **5** (adding axiom $(\neg t : \phi) \to ?t : (\neg t : \phi)$[11]).

For instance: **KD45** corresponds to **JD45** which consists of

$$J + (\neg t : \bot) + ((t : \phi) \to !t : (t : \phi)) + ((\neg t : \phi) \to ?t : (\neg t : \phi))$$

Similarly: **S5** corresponds to **JT45** which consists of the same but *factivity* axiom $t : \phi \to \phi$ instead of $\neg t : \bot$.

The semantics of jutification models are further defined using Fitting models[Fit05].

> **Definition 1.4.1.** *(Fitting models)*
>
> A Fitting model is a tuple $M = (W, R, E, V)$ where (W, R, V) is a standard kripke model and E is an evidence function which assigns to each (term, formula) pair a set of worlds in W.

Intuitively, if $w \in E(t, \phi)$, then t is *admissible* evidence for ϕ at world w. Which may not necessarily mean it is true evidence, but if it were then it would tell us that ϕ is the case.

Then belief is defined as follows:

> **Definition 1.4.2.** *(Belief on Fitting models)*
>
> For Fitting model $M = (W, R, E, V)$ we have:
>
> $$M, w \models t : \phi \text{ iff } w \in E(t, \phi) \text{ and } M, v \models \phi \text{ whenever } wRv$$

Note now the similarity between belief $t : \phi$ in such fitting models and explicit belief $B\phi$ in the awareness models we have seen in section 1.2.1. In awareness models we have an explicit belief in ϕ if $\phi \in A(w)$ and any doxastic alternative makes ϕ true. Here we have it it, with a term t if t is admissable for ϕ which implies awareness of ϕ and we again have that all doxastic alternatives satisfy ϕ.

Again we have our beliefs defined in terms of possible worlds, but this time with explicit derivations as fundamental as well.

[11] Originally suggested by [Pac06] and [Rub06] to have a S5 realization of logic of proofs.

Chapter 2

From evidence base to implicit beliefs

In this chapter we will describe a new mechanism to compute beliefs from explicit evidence. Our explicit evidence will come in the form of sets of *justification terms*. As discussed in section 1.4, these justification terms can be seen as encodings of derivations that the agent has constructed in order to arrive at certain conclusions. In section 2.1 we will cover what kinds of justification terms we use and what formulas they are *admissible* for.

Constructing such a term may mean that the agent gains a belief in whatever the conclusion of this term is. It may also mean that the agent discovers inconsistencies in their evidence, which in turn will cause them to lose some beliefs. This updating of beliefs relies heavily on a notion of *quasi-consistent evidence* which is similar to the one discussed in section 1.3. We will discuss the computation of *maximal quasi consistent (Qmax)* sets of evidence in section 2.2. Moreover, this section covers what kinds of evidence sets we will be considering, such a set we will call an *evidence base*.

In section 2.3, we will define how to compute explicit beliefs from such an evidence base. Moreover, we will provide examples and show that this notion of explicit belief, due to the addition of justification terms, overcomes the limitations discussed in section 1.3 which we illustrated with examples 1.3.9 and 1.3.10.

Section 2.4 will then cover what our implicit beliefs are for some evidence base. As opposed to all previous approaches mentioned in chapter 1 to model implicit belief, in this chapter we will do so without using possible-worlds. Instead, it relies on a notion of evidence closure which could intuitively be thought of as all the conclusions an omniscient reasoner could come to given the agents evidence. Later in chapter 3 we will consider possible worlds again.

Recall section 1.2.2 where the form of implicit belief defined over doxastic alternatives naturally corresponds to logical closure of belief. This is the idea behind our first definition of implicit belief, which we call *implicit belief*.
There are however more notions of (implicit) belief we can define in this formal setting. We define two more notions of belief: one being *evidence based implicit belief* which does not only consider the logical closure of explicit beliefs, but considers the logical closure of all qmax sets of evidence. The other being *potential explicit belief* which first considers the closure of a set of evidence and then computes explicit belief.

The notions of implicit belief and evidence based implicit belief can be seen as the implicit beliefs of a non-omniscient agent, which may be contradictory. Potential explicit belief could then be seen as the explicit beliefs of an omniscient reasoner, which cannot be contradictory and could be seen as all beliefs that a non-omniscient agent may eventually arrive at, thus corresponding more closely to *potential belief*.

2.1 Justification terms and admissibility

Here we will cover the kinds of justification terms we will be using throughout. For more information on justification logic and justification terms see section 1.4.
Moreover, we will define an admissibility relation $\gg$ adapted from [BRS14] which connects our terms to formulas. This relation will tell us what a term actually tells us, i.e. what a term is *admissible* for.

For now, we will leave the language somewhat simple, only considering the minimum to describe our belief computation. In chapter 3 we will cover a logic where we will further specify the language. All we need for our language right now is that it should be able to express formulas of the form: $\bot$ (contradiction), $\phi \to \psi$ (implication), and $t : \phi$ (justification).

Definition 2.1.1. *(Language)*

Given countable set of atomic formulas $Prop$, we define terms $t \in \mathcal{L}^t$ and formulae $\phi \in \mathcal{L}$ by mutual recursion as follows:

$$\phi := p \mid \bot \mid \phi \to \phi \mid t : \phi$$

$$t ::= c_\phi \mid t \cdot t \mid t \sqcup t \mid !t$$

We have propositional formulas ($p \in Prop$), contradiction ($\bot$) which is never true, material implication ($\to$) and justifications ($t : \phi$) which intuitively says that t is a justification for a belief in ϕ.

Regarding the justification terms we are considering, c_ϕ is our *atomic term*. It can be seen as a direct piece of evidence for ϕ, i.e. a witnessing of ϕ or *certificate* of ϕ [BRS14]. Note however that such a term need not mean that ϕ is actually true.

The other terms $t_1 \cdot t_2$, $t_1 \sqcup t_2$ and $!t$ denote terms that are obtained by combining other terms. Here $t_1 \cdot t_2$ denotes *application* of two terms, $t_1 \sqcup t_2$ denotes *weakening* of two terms and $!t$ denotes the *internalization* of a term. What this means will become apparent after discussing admissibility of terms, which we will do next.

Definition 2.1.2. *(Admissibility)*

Given $\mathcal{L}$ and $\mathcal{L}^t$ as defined in definition 2.1.1, we call *Admissibility* the smallest relation $\gg \subseteq \mathcal{L}^t \times \mathcal{L}$ such that:

- $c_\phi \gg \phi$
- If $t_1 \gg (\phi \to \psi)$ and $t_2 \gg \phi$ then $t_1 \cdot t_2 \gg \psi$
- If $t_1 \gg \phi$ and $t_2 \gg \phi$ then $t_1 \sqcup t_2 \gg \phi$
- If $t \gg \phi$ then $!t \gg (t : \phi)$

Application '$\cdot$' is left associative so we may write $t_1 \cdot t_2 \cdot ... \cdot t_n$ for $((t_1 \cdot t_2) \cdot ...) \cdot t_n$.
Implication $\to$ is right associative so we may write $\phi_1 \to \phi_2 \to ... \to \phi_n$ for $\phi_1 \to (\phi_2 \to (... \to \phi_n))$.
Weakening $\sqcup$ has associativity so we may omit brackets for weakening terms.

For ease of notation we occasionally write t_ϕ for a term t s.t. $t \gg \phi$.

Moreover, we write $t^{\gg}$ whenever there is some $\phi \in \mathcal{L}$ s.t. $t \gg \phi$ signaling that t is well formed. We may then refer to such admissible terms as 'arguments'[a]
Similarly, we write $t^{\not\gg}$ whenever there is no $\phi \in \mathcal{L}$ such that $t \gg \phi$ signaling that t is not a well formed argument.

We write $\mathcal{L}^{t^{\gg}}$ for the set of terms that are admissible for some formula, i.e.

$$\mathcal{L}^{t^{\gg}} := \{t \in \mathcal{L}^t \mid t^{\gg}\}$$

[a] We do not yet call these justifications as, what a justification in our system is will become apparent later on.

As noted in section 1.4, terms obtained by using *application* actually correspond to the rule (MP). An agent combining two terms via application can be seen as an agent taking their evidence and reasoning with modus ponens.

A novel addition to justification logic/justification terms proposed in this thesis is *weakening terms*. These bear some similarity to *sum* terms discussed in section 1.4. However, they serve a different purpose. Whereas sum is meant to express that we can put any justification terms together and this *summed* justification term will still tell us what both terms tell us , our weakening terms require both terms to tell us the same thing, i.e. both terms have to be admissible for the same formula.

The motivation for weakening terms is that there may be multiple ways of combining terms to reach a common conclusion. We may include a derivation d of conclusion ϕ into another derivation d' of ϕ by weakening to construct term $d \sqcup d'$. Even though it is possible that these paths of derivation d and d' which show a derivation of conclusion ϕ may be mutually exclusive (for instance one may contain term c_p whilst the other contains term $c_{p \to \bot}$) they at least tell us the same conclusion. We illustrate this with an example.

Example 2.1.3. We may have some term $c_{p \to q} \cdot c_p$ and term $c_{(\neg p \to q)} \cdot c_{\neg p}$. These can not both be truthful derivations at the same time. But they both tell us that q is the case so their weakening $(c_{p \to q} \cdot c_p) \sqcup (c_{(\neg p \to q)} \cdot c_{\neg p})$ could be considered trustworthy. Either p is true or not, in both cases q is true.

Lastly, we have *internalization* terms $!t$. Briefly mentioned in section 1.4 these terms are meant to express positive introspection.
Whenever we have some term t for ϕ we may construct term $!t$ for $t : \phi$ which is an argument that tells us that we believe ϕ due to argument t (which we will later call a justification). Of course, we may believe ϕ due to multiple arguments. We can have $t_1 \gg \phi$ and $t_2 \gg \phi$ which could both be trusted arguments. In this case $!t_1 \gg t_1 : \phi$ and $!t_2 \gg t_2 : \phi$ and also $!(t_1 \sqcup t_2) \gg (t_1 \sqcup t_2) : \phi$.

To finish this section we will define some functions which we will be using throughout the work:

Definition 2.1.4. *(subterm functions)*

$subterm : \mathcal{L}^t \to \mathcal{P}(\mathcal{L}^t)$:

$$subterm(t) = \{t\} \cup \begin{cases} \emptyset & \text{if } t = c_\alpha \\ subterm(t_1) \cup subterm(t_2) & \text{if } t = t_1 \cdot t_2 \text{ or } t = t_1 \sqcup t_2 \\ subterm(t') & \text{if } t = !t' \end{cases}$$

$sub_{\sqcup} : \mathcal{L}^t \to \mathcal{P}(\mathcal{L}^t)$:

$$sub_{\sqcup}(t) = \{t\} \cup \begin{cases} sub_{\sqcup}(t_1) \cup sub_{\sqcup}(t_2) & \text{if } t = t_1 \sqcup t_2 \\ \emptyset & \text{otherwise} \end{cases}$$

$sub_{Atm} : \mathcal{L}^t \to \mathcal{P}(\mathcal{L}^t)$

$$sub_{Atm}(t) = \begin{cases} \{t\} & \text{if } t = c_\phi \\ sub_{Atm}(t') & \text{if } t = !t' \\ sub_{Atm}(t_1) \cup sub_{Atm}(t_2) & \text{otherwise} \end{cases}$$

The function $subterm$ returns a set of all subterms of a given term, $sub_{\sqcup}$ returns all the weakened terms in a term t, note that this is just t when t is not a weakening. sub_{Atm} returns all atomic terms of a given term.

2.1.1 Justification terms and (topological) evidence models

For readers more familiar with evidence model [vBP11], and topological evidence models [BBÖS16], we want to make a comparison here between the type of justification terms that we are considering and the notion of *evidence* in (topological) evidence models. This transitions nicely into the next section where we will be considering such justification terms as pieces of evidence.

We will be considering *truth sets* here. Given a set of possible worlds W we define the truth set of formula ϕ, denoted as $\|\phi\|$, as $\|\phi\| := \{w \in W \mid w \vDash \phi\}$.

As noted before, evidence models use an evidence function which assigns sets of worlds to a world. Such sets of worlds are interpreted as pieces of evidence. In evidence models we make a distinction between *'direct'* pieces of evidence and *'combined'* pieces of evidence. Combined evidence can then be seen as taking two pieces of evidence (direct or combined) and creating a piece of evidence which entails both of them, i.e. the intersection of their truth sets.
In topological evidence models our set of direct evidence is the *sub-basis* of a topology on our set of possible worlds. Combined evidence is then obtained by combining all our evidence, which constitutes the *basis* of a topology on the set of possible worlds. To go from a basis to a topology we then close this basis under arbitrary unions, which are called *open sets*.

Now consider our justification terms. As noted before, our atomic terms can be seen as *witnesses* for a certain formula ϕ, i.e. direct pieces of evidence. Then we have terms obtained by application $t_1 \cdot t_2$. Now, even this term itself may only be admissible for formula ϕ, its subterms must be admissible for $\psi \to \phi$ and ψ for some ψ. So an argument $t_1 \cdot t_2$ for ϕ using arguments t_1 for $\psi \to \phi$ and argument t_2 for ψ can be seen as combining pieces of evidence $\|\psi \to \phi\|$ and $\|\psi\|$ i.e. $\|\psi \to \phi\| \cap \|\psi\| = \|\psi \wedge \phi\|$.
As noted before, weakening terms can be seen as derivations that tell us that we may have several several derivations for a formula ϕ that may be mutually inconsistent. In which case they can not all be truthfull at the same time. But we do know that, if at least one of them turns out to be truthful then ϕ should be the case. Consider the example above (example 2.1.3) where we have term $c_{p \to q} \cdot c_p$ and term $c_{\neg p \to q} \cdot c_p$. If we were to consider these as truth sets (as above) we have $\|p \to q\| \cap \|p\| = \|p \wedge q\|$ and $\|\neg p \to q\| \cap \|\neg p\| = \|(\neg p) \wedge q\|$. These two truth sets are mutually inconsistent since their intersection is empty. But if we were to take their union we get $\|p \wedge q\| \cup \|\neg p \wedge q\| = \|q\|$. Intuitively this is what weakening terms tell us as well.
Terms composed using weakening then relate to the union of (combined) pieces of evidence.

So weakening terms relate in this manner to open sets in a topology[1].

Given this (somewhat informal) argument of how these justification terms show similarities to notions of evidence as truth sets in evidence models, we do like to note that the use of terms adds a lot of expressivity that truth sets do not offer. For instance, $\|p \vee \neg p\|$ and $\|p \rightarrow (p \vee \neg p)\|$ are the same truth sets. But, due to the hyperintensional nature of justfication terms, a term for $p \rightarrow (p \vee \neg p)$ is a different term than one for $p \vee \neg p$.
Adding to that, having a term for p and one for $\neg p$ does not automatically mean we have constructed a term for $p \vee \neg p$. We will need a term for $p \rightarrow (\neg p \vee p)$ or one for $\neg p \rightarrow (p \vee \neg p)$ and apply it to another. In case we have both, we can get two different terms for $p \vee \neg p$ which we can weaken to a term that tells us that whether p is the case or $\neg p$ is the case, either way $p \vee \neg p$ is the case.

2.2 Evidence base and Qmax evidence

Having discussed what our *justifcation terms* are, we can move on to the new sets of evidence we will be considering. From now on we will use "evidence" and "term" interchangeably as we will build our sets of explicit evidence out of justification terms discussed in previous section. With that we will impose some restrictions on what these sets of evidence may look like.
For instance, we would not want our pieces of derived evidence to come completely out of nowhere. If we have gotten some conclusion by modus ponens (application) using two pieces of evidence, then we better have those two pieces of evidence as well. Moreover, we would not want to allow ill-formed terms in our evidence.

We would like to consider certain sets of evidence with such desired properties. We will call these *Evidence bases*. In this section we will cover such evidence bases.

Moreover, we will discuss what *maximal quasi-consistent (Qmax)* evidence is which serves as the foundation of our beliefs. Such quasi-consistent evidence was already covered in section 1.3 when discussing the formal framework in [Gog16]. But since our evidence now consists of justification terms as opposed to formulas in order to overcome limitations of the previous approach to explicit belief, we will have to redefine what Qmax evidence is.

2.2.1 Evidence bases

We will start by defining *quasi subterm closure*, a property our evidence bases should have.

> **Definition 2.2.1.** *(Subterm quasi-closed evidence)*
>
> We say that a set of evidence $E \subseteq \mathcal{L}^t$ is *subterm quasi closed* iff
>
> - $\{t_1, t_2\} \subseteq E$ whenever $t_1 \cdot t_2 \in E$.
> - $\{t_1, t_2\} \cap E \neq \emptyset$ whenever $t_1 \sqcup t_2 \in E$.
> - $t \in E$ whenever $!t \in E$

Notice how a weakening in our set of evidence does not require all subterms to be in E. This is the reason we call it *quasi closed* as opposed to *closed*. This will become more important when considering qmax evidence and its relation to explicit belief.

[1] Of course, there are more nuances to this, but we leave it somewhat informal here as it is not the goal of this thesis to relate these two concepts.

With this definition we define an *Evidence base*:

Definition 2.2.2. *(Evidence base)*

An *Evidence base* is a set $E \subseteq \mathcal{L}^t$ of *explicit evidence* such that:

- E is quasi subterm closed (definition 2.2.1)
- E consists only of admissible evidence: (each $t \in E$ has that $t^{\gg}$)

We call the set of all evidence bases **EB**.

Since **EB** is closed under arbitrary unions we have that there is a largest evidence base which is exactly the set of all admissible terms:

$$\bigcup \mathbf{EB} = \mathcal{L}^{t^{\gg}}$$

In the interest of modeling realistic agents we generally want these evidence bases to be finite. But we put no such restrictions on evidence bases at the moment.

Throughout we may be interested in the formulas for which we have some argument i.e. the formulas for which we have a term admissible for it:

Definition 2.2.3. For set of evidence $E \subseteq \mathcal{L}^t$ we denote the set of formulas of E as:

$$\Phi(E) = \{\phi \in \mathcal{L} \mid \text{ there is } t_\phi \in E\}$$

An ineteresting kind of evidence base is one that we will call *"collected"*. Such evidence bases are intuitively *closed under weakening up to redundancy*.

Definition 2.2.4. *(Collected evidence base)*

We say an evidence base $E \in \mathbf{EB}$ is *collected* when

$t_1 \sqcup t_2 \in E$ whenever $\{t_1, t_2\} \subseteq E$, $t_1 \sqcup t_2{}^{\gg}$ and $sub_{\sqcup}(t_1) \cap sub_{\sqcup}(t_2) = \emptyset$.

We call the set of collected evidence bases **CEB**.

We denote $Coll(E)$ as the smallest collected evidence base containing E, which we call the collected closure of E.

We add the 'up to redundancy' condition ("$sub_{\sqcup}(t_1) \cap sub_{\sqcup}(t_2) = \emptyset$") to avoid making all collected evidence bases infinite

Proposition 2.2.5. If E is an evidence base then:

1. if $\{t_1, ..., t_n\} \subseteq E$ and for each t_i we have $t_i \neq t_j$, $t_i \gg \phi$ and $sub_{\sqcup}(t_i) = \{t_i\}$, then $t_1 \sqcup ... \sqcup t_n \in Coll(E)$.
2. if E is finite then $Coll(E)$ is finite.

Proof.

1. Suppose $\{t_1, ..., t_n\} \subseteq E$ such that each t_i has that $t_i \gg \phi$, t_i is not a weakening term, and $t_i \neq t_j$ if $i \neq j$.

   Suppose $n = 1$, then $\{t_1, ..., t_n\} = \{t_1\}$ so $t_1 \in E$ whenever $\{t_1, ..., t_n\} \subseteq E \subseteq$

$Coll(E)$.

Induction Hypothesis: For $\{t_1, ..., t_k\} \subseteq E$ we have that $t_1 \sqcup ... \sqcup t_k \in Coll(E)$.

Suppose $n = k + 1$. Then $\{t_1, ..., t_k, t_{k+1}\} \subseteq E \subseteq Coll(E)$ with each t_i a unique non-weakening term admissible for ϕ.
By I.H. we have that $t_1 \sqcup ... \sqcup t_k \in Coll(E)$. Moreover, since each t_i is not a weakening and is unique we have $sub_\sqcup(t_1 \sqcup ... \sqcup t_k) \cap sub_\sqcup(t_{k+1}) = \emptyset$. Since $Coll(E)$ is closed under weakening and all t_i are admissible for ϕ and are in E we have that $t_1 \sqcup ... \sqcup t_k \sqcup t_{k+1} \in Coll(E)$.

2. Suppose E is finite. Since $Coll(E)$ is the smallest collected evidence base containing E we have that :

$$Coll(E) = E \cup \{t_1 \sqcup t_2 \mid \{t \in sub_\sqcup(t_1 \sqcup t_2) \mid t \neq t' \sqcup t''\} \subseteq E, t_1 \sqcup t_2{}^{\gg}, sub_\sqcup(t_1) \cap sub_\sqcup(t_2) = \emptyset\}$$

Since E is finite we have that $\{t_1 \sqcup t_2 \mid \{t \in sub_\sqcup(t_1 \sqcup t_2) \mid t \neq t' \sqcup t''\} \subseteq E, t_1 \sqcup t_2{}^{\gg}, sub_\sqcup(t_1) \cap sub_\sqcup(t_2) = \emptyset\}$ is finite as well, since if it were infinite it would contain some weakening term $t_1 \sqcup t_2$ s.t. $sub_\sqcup(t_1) \cap sub_\sqcup(t_2) \neq \emptyset$.
So $Coll(E)$ is finite when E is.

■

The reason we give such evidence bases the name "collected" is because we can visualize an agent having a collected evidence base as an agent having collected all their arguments for a certain formula ϕ and having put it all on a big pile for ϕ. This pile can then be seen as the largest weakening of arguments for ϕ (that are not weakenings themselves).

Moreover, we may consider *well organized* evidence bases which enhance *collected* evidence bases as follows:

Definition 2.2.6. *(Well organized evidence base)*

An evidence base $E \in \mathbf{EB}$ is *well organized* whenever:

- E is collected (definition 2.2.4)
- For each $t_1, t_2, t_3 \in E$ such that:
  - $t_1 \in substerm(t_2)$
  - $t_1 \in sub_\sqcup(t_3)$
  - $t_2 \notin sub_\sqcup(t_3)$

  we have that $t_2[t_1 := t_3] \in E$

  where $t_2[t_1 := t_3]$ is the term that is obtained by replacing all occurrences of t_1 with t_3.

We call the class of well organized evidence bases **WOEB**.

We denote $Org(E)$ as the smallest well organized evidence base containing E.

The idea of well organized evidence bases is that, whenever we have some derivation that uses argument t for ϕ we also have access to any derivation that uses a weakening of t instead of t, given that we have access to this weakening. Having access to this weakening is guaranteed by the collected property.
In turn, for any derivation that the agent has made with any argument for ϕ, the agent should also have access to the derivation using the whole collection of evidence for ϕ, which would correspond to using the largest weakened term for ϕ in the derivation.
In a way, these kind of agents can be thought of as agents that do not merely *combine evidence*, but *reason with formulas*.

As a final note: we have that $Org(E)$ contains all derivations t that are obtained by substituting subderivations of t with suitable weakenings. Moreover, $Org(E)$ is finite if E is finite.

Proposition 2.2.7. If E is an evidence base then:

1. $Org(E)$ contains any derivation $t[t_1 := t_2]$ whenever $t_1 \in subterm(t)$, $t_1 \in sub_{\sqcup}(t_2)$ and $t \notin sub_{\sqcup}(t_2)$.

2. $Org(E)$ is finite when E is finite.

Proof.

1. Follows from definition.

2. Suppose E is finite, then $Coll(E)$ is finite. Then any term added by $Org(Coll(E)) = Org(E)$ has to be a substitution $t_2[t_1 := t_3]$ with $t_1 \in subterm(t_2)$, $t_1 \in sub_{\sqcup}(t_3)$ and $t_2 \notin sub_{\sqcup}(t_3)$. Since $t_2 \notin sub_{\sqcup}(t_3)$ we have that $Org(E)$ is not infinite.

■

2.2.2 Qmax evidence

Our evidence base is used to compute explicit belief and eventually also implicit belief. We may want to simply consider the arguments in our evidence base and say that every formula that we have evidence for is explicitly believed. But we may have explicit evidence for p and $\neg p$ in which case it is reasonable to not believe either of these. We may only want to explicitly believe what explicit evidence tells us when we actually *trust* that evidence. On the other hand, we are trying to prevent our agent(s) from being omniscient. An agent may have not yet figured out that they have evidence that contradicts.

We will take inspiration from [Gog16] and use a modified *qmax calculation* to account for the justification terms that we have added.
This qmax calculation uses a notion of closure with respect to some set of evidence E, which intuitively means that, given some subset e of evidence base E, we want to make all the derivations that we can from evidence in e, but restricted to E.

However, An important aspect that the use of justification terms offers us is the ability to access *subderivations*. This allows us to remove terms based on their subterms. We may for instance want to disregard a derivation that is built up from some arguments that we do not trust. This was not possible in the old qmax calculation as seen in example 1.3.10.

We will therefore want our qmax sets to be quasi closed under subterms (definition 2.2.1) as well. As this will ensure that subderivations are contained in a set of evidence whenever the derivation is. Except for weakened terms as we do not require trust in all weakening terms to trust the weakening term.

A final thing we note before we define what such a closure is. As opposed to Gogoladze [Gog16] we will not be considering *knowledge* and so we will not consider sets of *hard evidence* that should be present in all closures. However, we do assume a logic Λ that our agent reasons with, for which its axioms will always be part of a closure when it is in E. Such axioms could be seen as our variant of hard evidence[2]. This logic Λ we leave unspecified for now but later in chapter 3 we will define such logics.

This brings us to the following definition of closure w.r.t. E

[2]However, it could be interesting to explore accounts in which these axioms are not necessarily contained in such a closure. one could consider versions where an agent may consider evidence against their axioms as plausible, and then update their logic Λ when they deem suitable, this we leave as possible future work.

Definition 2.2.8. *(E-closure)*

For evidence base $E \in \mathbf{EB}$, we say that $e \subseteq E$ is closed w.r.t. E (E-closed) iff

- If $t_1 \cdot t_2 \in E$ then $(\{t_1, t_2\} \subseteq e$ iff $t_1 \cdot t_2 \in e)$.
- If $t_1 \sqcup t_2 \in E$ then $(\{t_1, t_2\} \cap e \neq \emptyset$ iff $t_1 \sqcup t_2 \in e)$
- If $!t \in E$ then $(t \in e$ iff $!t \in e)$.
- If $c_\phi \in E$ and ϕ is a Λ-axiom then $c_\phi \in e$

We denote $Clos^E(e)$ set of smallest E-closures containing e. When $Clos^E(e) = \{e'\}$ we write $Clos^E(e)$ for e'.

Note that any E-closure is quasi-closed under subterms.

Fact 2.2.9. Suppose $e \subseteq E$ is closed w.r.t. E. Then e is quasi-closed under subterms.

Further note that, due to subterm quasi-closure of weakening terms not requiring both the subterms to be present, we have that the E-closure of some set of terms need not consist of a single set of terms.
However, we will mostly be interested in quasi-subterm closed subsets, for which $Clos^E(e)$ consists of one unique set.

Fact 2.2.10. For evidence base $E \in \mathbf{EB}$ If $e \subseteq E$ is quasi-closed under subterms then $Clos^E(e)$ consists of a single set

Moreover, when this case, the E-closure of some set e corresponds to *closure under composition w.r.t. E*. This is a definition we will be referring to throughout.

Definition 2.2.11. *(closure under composition)*

We say a set of evidence $e \subseteq \mathcal{L}^t$ is *closed under composition w.r.t E* iff

- if $\{t_1, t_2\} \subseteq e$ and $t_1 \cdot t_2 \in E$ then $t_1 \cdot t_2 \in e$
- If $\{t_1, t_2\} \cap e \neq \emptyset$ and $t_1 \sqcup t_2 \in E$ then $t_1 \sqcup t_2 \in e$.
- If $t \in e$ and $!t \in E$ then $!t \in e$.

If a set of evidence is closed w.r.t. $\mathcal{L}^{t^{\gg}}$ then we say it is closed under composition.

Notice that a set being E-closed implies that a set is closed under composition w.r.t. E

Fact 2.2.12. If e is E-closed then e is closed under composition w.r.t. E

Qmax belief calculation further relies on a notion of quasi-consistent evidence (it is what the Q stands for). Quasi consistent evidence is allowed to contain inconsistent evidence; for instance c_p and $c_{p\to\bot}$. It is however not allowed that we would be able to construct an argument for $\bot$ w.r.t. E from such a set of evidence. This would mean that, if $\{c_p, c_{p\to\bot}\} \subseteq e$ for quasi-consistent subset $e \subseteq E$, then $c_{p\to\bot} \cdot c_p \notin E$.

To put it more formally:

Definition 2.2.13. *(Quasi-consistent evidence)*

For evidence base $E \in \mathbf{EB}$ we call a set $e \subseteq E$ *quasi consistent* w.r.t. E iff there is no term $t \in Clos^E(e)$ s.t. $t \gg \bot$.

We say that a set of evidence E is *quasi-consistent* if E is quasi-consistent w.r.t. itself.

Again, as in [Gog16], we will want to look at maximal such subsets, which brings us to our definition of *maximally quasi-consistent (qmax)* sets.

Definition 2.2.14. *(maximally quasi-consistent set (Qmax))*

For evidence base $E \in \mathbf{EB}$ we call a set of evidence $e \subseteq E$ *maximally quasi-consistent* whenever e is quasis-consistent w.r.t. E but no proper extension $e' \subseteq E$ of e is quasi-consistent w.r.t. E.

Given evidence base $E \in \mathbf{EB}$ We denote the set of all Qmax evidence w.r.t. E as:

$$Qmax(E) = \{e \subseteq E \mid e \text{ is Qmax w.r.t. } E\}$$

Due to maximality of qmax sets we have that $e \in Qmax(E)$ implies that $e = clos^E(e)$.

Fact 2.2.15. For $e \in Qmax(E)$ we have that $e = Clos^E(e)$.

Note that each qmax set e of evidence base E only consists of admissible evidence since $e \subseteq E$. Moreover, fact 2.2.9 tells us that e is quasi closed under subterms, meaning that it is itself an evidence base as well.

Fact 2.2.16. Given $E \in \mathbf{EB}$ we have that each $e \in Qmax(E)$ has $e \in \mathbf{EB}$.

Since each $e \in Qmax(E)$ has that $e = Clos^E(e)$ we have the following as well:

Proposition 2.2.17. For evidence base $E \in \mathbf{EB}$ we have that:

$e \in Qmax(E)$ iff

1. $e \subseteq E$.
2. e is quasi-consistent.
3. e is closed under composition w.r.t. E.
4. e is quasi-closed under subterms.
5. e is maximal w.r.t. these properties.

Proof.

($\rightarrow$)
1. $e \subseteq E$ by definition.
2. e is quasi-consistent by definition.
3. $e = Clos^E(e)$ so by fact 2.2.12 we have that e is closed under composition w.r.t. E.
4. $e = Clos^E(e)$ so by fact 2.2.9 we have that e is quasi-closed under subterms.
5. e is maximal w.r.t. these properties.
   Suppose $e' \subseteq E$ is quasi-consistent w.r.t. E, closed under composition w.r.t E. Then e' is quasi-consistent and E-closed. $e \subseteq e' = Clos^E(e') = Clos^E(e) = e$, so $e' = e$.

(←) Suppose e has properties 1–5. Then $e \subseteq E$, e is $E-$closed since it is both closed under composition w.r.t. E and quasi-closed under subterms, e is quasi-consistent, and maximal w .r.t. these properties. So then $e \in Qmax(E)$.

■

Let us now go over the behaviour of qmax evidence in a bit more detail:

Let us say we have some evidence base E and suppose that there is no evidence admissible for $\perp$ i.e. we have not derived any contradiction in our evidence i.e. E is quasi-consistent. Then, due to maximality of qmax sets, we have that E is actually the only piece of qmax evidence. This brings us the following fact:

Fact 2.2.18. Whenever E is quasi consistent i.e. there is no $t \in E$ s.t. $t \gg \perp$ we have that $Qmax(E) = \{E\}$.

Suppose we *do* have some evidence admissible for $\perp$ in our explicit evidence E. But this piece of evidence is some basic piece of evidence $c_\perp$. Also suppose this is the only piece of evidence admissible for $\perp$ that we have. Since we have not derived this contradiction using any other piece of evidence we have that $c_\perp$ is the only reason that E is not quasi-consistent. If we remove $c_\perp$ we have our only qmax set again i.e. $Qmax(E) = \{E \setminus \{c_\perp\}\}$. This makes sense as obtaining some evidence that is blatantly false can be disregarded immediately.

Now consider a more interesting scenario: $E = \{c_p, c_{p\to\perp}\}$. Again, we have no term admissible for $\perp$ in E, so E is quasi-consistent so it is the only Qmax set.
However, let's say we become aware of our contradictory evidence by combining our evidence. Now we get $E' = \{c_p, c_{p\to\perp}, c_{p\to\perp} \cdot c_p\}$, In which case we will get two Qmax sets: one with c_p and one with $c_{p\to\perp}$.

What we see here is that the addition of $c_{p\to\perp} \cdot c_p$ has made us aware that term c_p and $c_{p\to\perp}$ are *in conflict*. This notion of conflict will prove usefull throughout so we will make it a definition:

Definition 2.2.19. For any evidence base E we say that a term $t \in E$ conflicts with $e \subseteq E$ w.r.t. E whenever:

- if $t = c_\phi$ then there is $t'_\perp \in Clos^E(e \cup \{c_\phi\})$.
- if $t = t_1 \cdot t_2$ then $\{t_1, t_2\} \not\subseteq e$.
- if $t = t_1 \sqcup t_2$ then $\{t_1, t_2\} \cap e = \emptyset$.
- if $t = !t'$ then $t' \notin e$.

We may write $t \lightning_E e$ whenever term t conflicts with e w.r.t. E. We may write $t \not\lightning_E e$ whenever term t does not conflict with e w.r.t. E.

We say that a set of evidence $e \subseteq E$ is *in conflict w.r.t* E if there is some $t \in e$ such that $t \lightning_E e$. Which we may write as $\lightning_E e$. Similarly, if e is not in conflict w.r.t. E then we write $\not\lightning_E e$.

Let us focus in on this definition. The first case, where t is an atomic term, tells us that t conflicts with a set e w.r.t. E whenever there is some contradiction that we have become aware of ($t'_\perp \in E$), and adding c_ϕ to e will result in, by E-closure, t' being in e, which can not be the case if e happens to be quasi-consistent. Notice that this condition is equivalent to stating that $e \cup \{c_\phi\}$ is not quasi-consistent w.r.t. E.

In essence this tells us that e contains all the necessary components apart from c_ϕ to derive contradiction t'. The negation of this condition can be seen a the building block of the maximality of qmax sets, as we see in proposition 2.2.20, we will have $c_\phi \notin e$ iff $c_\phi \lightning_E e$. Turning that around, we have $c_\phi \in e$ iff $c_\phi \not\lightning_E e$ meaning that $c_\phi \in e$ iff there is no $t_\bot \in Clos^E(e \cup \{c_\phi\})$ i.e. as long as adding c_ϕ does not lead to adding a contradiction term by composition, it has to be a part of e.

The composite cases in definition 2.2.19 then tell us that a term t conflicts with e if its components conflict with e. Let us say that $e = \{c_p\}$ and $E = \{c_p, c_{p\to\bot}, c_{p\to\bot} \cdot c_p\}$. Then $c_{p\to\bot} \cdot c_p$ conflicts with e since $c_{p\to\bot} \notin e$ which (if e is qmax) is the case because $c_{p\to\bot}$ conflicts with e.

Note that this also means that any $e \subseteq E$ has that e is quasi-closed under subterms if it is not in conflict w.r.t. E.

The intuition behind this conflict is that whenever we leave out some term from a qmax set then there has been a reason for that; namely that is in conflict with the qmax set.

Proposition 2.2.20. For any $e \in Qmax(E)$ and any $t \in E$ we have that $t \notin e$ iff $t \lightning_E e$.

Proof. Suppose $e \in Qmax(E)$.

(←) Suppose t conflicts with e w.r.t. E.

- Suppose $t = c_\phi$. Then there is term $t'_\bot \in E$ s.t. $t' \in Clos^E(e \cup \{c_\phi\})$. Suppose, for contradiction, that $c_\phi \in e$. Then $t' \in Clos^E(e) = e$. This can however not be the case since $t' \gg \bot$. So it cannot be the case that $c_\phi \in e$.
- Suppose $t = t_1 \cdot t_2$. Then we have that $\{t_1, t_2\} \not\subseteq e$. Then $t \notin e$ since e is quasi closed under subterms.
- Suppose $t = t_1 \sqcup t_2$. Then we have that $\{t_1, t_2\} \cap e = \emptyset$ and thus $t \notin e$ due to quasi closure under subterms.
- Suppose $t = !t'$, then $t' \notin e$ and so, due to quasi closure under subterms $t \notin e$.

So $t \notin e$.

(→) Suppose, for contraposition, that $t \in E$ does not conflict with $e \in Qmax(E)$.

- Suppose $t = c_\phi$. Then there is no $t' \in E$ s.t. $t' \in Clos^E(e \cup \{c_\phi\})$. That means that $e \cup \{c_\phi\}$ is quasi-consistent. Since e is maximally quasi consistent w.r.t. E and $e \cup \{c_\phi\} \subseteq E$ we have that $e = e \cup \{c_\phi\}$. So $c_\phi \in e$.
- Suppose $t = t_1 \cdot t_2$ then $\{t_1, t_2\} \subseteq e$. Since e is qmax and $t \in E$ we have that $t \in e$.
- Suppose $t = t_1 \sqcup t_2$ then $\{t_1, t_2\} \cap e \neq \emptyset$. If either $t_1 \in e$ or $t_2 \in e$ we have that $t_1 \sqcup t_2 \in e$.
- Suppose $t = !t'$ then $t' \in e$ and so, because e is qmax and $t \in E$, we have that $t \in e$.

By contraposition we have that $t \notin e$ implies that $t \lightning_E e$.

So $t \notin e$ iff t conflicts with e w.r.t E.

■

Now, this extends to supersets of evidence as well. If t conflicts with e w.r.t. E we can add elements to both e and E and we will still have that the conflict is present.

Fact 2.2.21. If t conflicts with e w.r.t. E, $e \subseteq e'$ and $E \subseteq E'$ then t conflicts with e' w.r.t. E'.

Furthermore, we can extend any set that is not in conflict w.r.t. E to a qmax set of E:

Definition 2.2.22. *(qmax extension)*

For a set of evidence $e \subseteq E$ that is not in conflict w.r.t E, we construct set $e' \subseteq E$ such that $e' \in Qmax(E)$ as follows:

1. Make an enumeration of $c_\phi \in E \setminus e$
2. Let

$$e_0 := e$$
$$e_{i+1} := \begin{cases} e_i \cup \{c^i_\phi\} & \text{if } c^i_\phi \not\lightning_E e_i \\ e_i & \text{otherwise} \end{cases}$$
$$e' := Clos^E(e_n)$$

   where n is the size of $\{c_\phi \mid c_\phi \in E \setminus e\}$.

We call such a set e' a *qmax extension of e w.r.t. E*.

Note that this qmax extension need not be unique. For instance $\emptyset$ can be extended to all qmax sets of E for any evidence base E.

We may then want to look at all qmax extensions of a certain (non-conflicting) set.

Definition 2.2.23. *(set of qmax extensions)*
Given evidence base $E \in \mathbf{EB}$ and $e \subseteq E$ s.t. $\not\lightning_E e$ we denote the set of all qmax extension of e w.r.t. e as:

$$qmax^{\uparrow E}(e) = \{e' \subseteq E \mid e' \text{ is a qmax extension of } e \text{ w.r.t. } E\}$$

Proposition 2.2.24. For any $e \subseteq E$ that is not in conflict w.r.t E we have that $qmax^{\uparrow E}(e) \subseteq Qmax(E)$.

Proof. Suppose $e \subseteq E$ is not in conflict w.r.t. E. Then there is no $t \in e$ s.t. $t \lightning_E e$. Take any $e' \in qmax^{\uparrow E}(e)$.

- $e' \subseteq E$ since we are only adding atomic terms from E and then closing it w.r.t. E.
- e' is quasi-consistent w.r.t. E since, at any stage e_i, we are only adding atomic term c_ϕ to e_i if $e_i \cup \{c_\phi\}$ is quasi-consistent w.r.t. E.
- maximal since we go over all atomic terms and then close w.r.t. E.

■

Before we end this section we would like to define a sense of *awareness of contradictions* as this will remain relevant for explicit and implcit beliefs.

Definition 2.2.25. We say an agent is aware of a contradiction in $e \subseteq E$ if $Clos^{\mathcal{L}^{t^\gg}}(e)$ contains some $t_\perp$ s.t. $t \in E$. i.e. there is some derivation built up from e that shows that e is inconsistent.
We say an agent is aware of all contradictions when, for any set of evidence $e \subseteq E$ that is not quasi-consistent w.r.t. $\mathcal{L}^{t^\gg}$, the agent is aware of the contradiction.

2.3 Explicit beliefs

Qmax evidence will form the basis of explicit beliefs, and later also implicit beliefs. Our qmax evidence tells us what consistent subsets of our evidence are up to our awareness. It then makes sense that whenever all these consistent sets agree on some piece of evidence, that we trust this piece of evidence and therefore believe whatever it tells us:

Definition 2.3.1. *(Explicit beliefs)*

For evidence base $E \in \mathbf{EB}$ we define the set of explicit beliefs $\mathcal{B}^e_E \subseteq \mathcal{L}$ as the following:

$$\mathcal{B}^e_E := \Phi(\bigcap Qmax(E))$$

with $\Phi(.)$ as defined in definition 2.2.3.

In other words, we have $\phi \in \mathcal{B}^e_E$ iff there is some term t admissible for ϕ such that, for each $e \in Qmax(E)$, we have $t \in e$. Such a term t we will refer to as a *trusted term* or *justification*.

An agent believes a formula ϕ if they have some argument t occurring in all their qmax sets of evidence. However, consider the following example:

Example 2.3.2.
Suppose E is an evidence base as follows:

$$E = \{c_p, c_{p\to\bot}, c_{p\to q}, c_{(p\to\bot)\to q}, c_{p\to q} \cdot c_p, c_{(p\to\bot)\to q} \cdot c_{p\to\bot}, c_{p\to\bot} \cdot c_p\}$$

We have an argument for p (c_p), an argument for $\neg p$ ($c_{p\to\bot}$) and arguments that tell us that, whichever of the two, q is the case ($c_{p\to q}, c_{(p\to\bot)\to q}$).
Moreover, we have applied all suitable terms to each other.
If we now consider the qmax sets of E we have the following:

$$\begin{aligned} Qmax(E) = \{&\{c_p, c_{p\to q}, c_{(p\to\bot)\to q}, c_{p\to q} \cdot c_p\} \\ &\{c_{p\to\bot}, c_{p\to q}, c_{(p\to\bot)\to q}, c_{(p\to\bot)\to q} \cdot c_{p\to\bot}\}\end{aligned}$$

Then $\cap Qmax(E) = \{c_{p\to q}, c_{(p\to\bot)\to q}\}$ which means that $q \notin \mathcal{B}^e_E$.

In the above example we see that, even though we have an argument for q in all of our qmax sets, we do not have a belief in q since we do not have the same argument for q in each qmax set, i.e. we do not have a justification for q. We may however want to have an explicit belief in q in such situations since we do have an argument for q in each qmax set.

This is exactly where weakening terms enter the picture[3]. Recall example 2.1.3 where we had a similar situation (one argument for p, one for $\neg p$, for each an argument to obtain an argument for q). Here we noted that we can add a weakening term which informally tells us “whichever of these arguments is actually true, in both cases ϕ is the case". And this is exactly what we need in this scenario.
Let's say we have two qmax sets, one containing t_ϕ and one containing t'_ϕ. If we were to figure out that t and t' are both admissible for ϕ we would add $t \sqcup t'$ to our evidence base. Due to qmax evidence being closed under composition w.r.t. E (definition 2.2.11), we will have that in both qmax sets this term $t \sqcup t'$ appears.

[3]Later in chapter 3 we will see another kind of evidence base and qmax computation for which we do not require weakening terms. Specifically because having a justification for ϕ whenever there is an argument for ϕ in each qmax set is already inherent in these new kind of evidence bases and qmax sets due to other properties.

To continue on the previous example:

Example 2.3.3. Continuing on example 2.3.2 where we have:

$$E = \{c_p, c_{p\to\bot}, c_{p\to q}, c_{(p\to\bot)\to q}, c_{p\to q}\cdot c_p, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, c_{p\to\bot}\cdot c_p\}$$

Suppose we now add $t_q := (c_{p\to q}\cdot c_p)\sqcup(c_{(p\to\bot)\to q}\cdot c_{p\to\bot})$:

$$E' = E\cup\{(c_{p\to q}\cdot c_p)\sqcup(c_{(p\to\bot)\to q}\cdot c_{p\to\bot})\}$$

If we now consider the qmax sets of E':

$$\begin{aligned}Qmax(E') = \{&\{c_p, c_{p\to q}, c_{(p\to\bot)\to q}, c_{p\to q}\cdot c_p, t_q\}\\ &\{c_{p\to\bot}, c_{p\to q}, c_{(p\to\bot)\to q}, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, t_q\}\end{aligned}$$

We see that $(c_{p\to q}\cdot c_p)\sqcup(c_{(p\to\bot)\to q}\cdot c_{p\to\bot})$ appears in each qmax set resulting in $q\in\mathcal{B}^e_E$.

So adding this weakening of all arguments for q has resulted in an explicit belief in q since we have a justification for q in this case. If we then consider collected evidence bases **CEB** (definition 2.2.4) we get the following:

We can then consider collected evidence bases again (which are closed under weakening up to redundancy), and we have the following:

Proposition 2.3.4. If $E\in$ **CEB** is collected we have that

$$\Phi(\bigcap Qmax(E)) = \bigcap\{\Phi(e)\mid e\in Qmax(E)\}$$

i.e. there is some justification t for ϕ iff each qmax set e has some argument t' for ϕ.

Proof. Suppose $E\in$ **CEB** is a collected evidence base.

($\subseteq$) Suppose $\phi\in\Phi(\bigcap Qmax(E))$. Then there is $t\in\cap Qmax(E)$ s.t. $t\gg\phi$. Then any $e\in Qmax(E)$ has that $t\in e$. Then any $e\in Qmax(E)$ has some $t\in e$ such that $t\gg\phi$ so $\phi\in\bigcap\{\Phi(e)\mid e\in Qmax(E)$.

($\supseteq$) Suppose each $e\in Qmax(E)$ has some term $t_\phi\in e$. Since each qmax set is quasi-closed under subterms we have that each $e\in Qmax(E)$ has some term $t_\phi\in e$ s.t. $sub_\sqcup(t)=\{t\}$ i.e. t is not a weakening term. Then all such t are also contained in E and since E is closed under weakening (up to redundancy) we have that the weakening t' of all such t is contained in E. Then by closure under composition w.r.t. E, each qmax set e will have $t'\in e$. And thus there is $t_\phi\in\cap(Qmax(E))$.

■

Let us continue on example 2.3.3. This time by adding some evidence against q and becoming aware of the contradiction with current evidence.

Example 2.3.5. Continuing on example 2.3.3 where we have:

$$\begin{aligned}E' &= E\cup\{t_q\}\\ &= \{c_p, c_{p\to\bot}, c_{p\to q}, c_{(p\to\bot)\to q}, c_{p\to q}\cdot c_p, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, c_{p\to\bot}\cdot c_p, (c_{p\to q}\cdot c_p)\sqcup(c_{(p\to\bot)\to q}\cdot c_{p\to\bot})\}\end{aligned}$$

Now let's say we obtain $c_{q\to\bot}$ and we apply it to t_q. $E'' = E' \cup \{c_{q\to\bot}, c_{q\to\bot}\cdot t_q\}$.

$$\begin{aligned}Qmax(E'') = \{&\{c_{p\to q}, c_p, c_{p\to q}\cdot c_p, c_{(p\to\bot)\to q}, t_q\},\\ &\{c_{p\to q}, c_{(p\to\bot)\to q}, c_{(q\to\bot)}\}\\ &\{c_p, c_{(p\to\bot)\to q}, c_{q\to\bot}\}\\ &\{c_{p\to\bot}, c_{p\to q}, c_{q\to\bot}\}\\ &\{c_{(p\to\bot)\to q}, c_{p\to\bot}, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, c_{p\to q}, t_q\}\}\end{aligned}$$

In this case there is no term occuring in each qmax evidence, i.e. no justifications, so we have no explicit beliefs.

Now suppose, instead of applying $c_{q\to\bot}$ to t_q we were to first apply it to $c_{p\to q}\cdot c_p$ and then to $c_{(p\to\bot)\to q}\cdot c_{p\to\bot}$:

1. $E''' = E' \cup \{c_{q\to\bot}, c_{q\to\bot}\cdot(c_{p\to q}\cdot c_p)\}$

   Then $$\begin{aligned}Qmax(E''') = \{&\{c_{p\to q}, c_p, c_{p\to q}\cdot c_p, c_{(p\to\bot)\to q}, t_q\},\\ &\{c_p, c_{(p\to\bot)\to q}, c_{q\to\bot}\}\\ &\{c_{(p\to\bot)\to q}, c_{p\to\bot}, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, c_{p\to q}, t_q, c_{q\to\bot}\}\}\end{aligned}$$

   In which case we only have an explicit belief that $\neg p \to q$.

2. $E'''' = E''' \cup \{c_{q\to\bot}\cdot(c_{(p\to\bot)\to q}\cdot c_{p\to\bot})\}$

   Then $$\begin{aligned}Qmax(E'''') = \{&\{c_{p\to q}, c_p, c_{p\to q}\cdot c_p, c_{(p\to\bot)\to q}, t_q\},\\ &\{c_{p\to q}, c_{(p\to\bot)\to q}, c_{(q\to\bot)}\}\\ &\{c_p, c_{(p\to\bot)\to q}, c_{q\to\bot}\}\\ &\{c_{p\to\bot}, c_{p\to q}, c_{q\to\bot}\}\\ &\{c_{(p\to\bot)\to q}, c_{p\to\bot}, c_{(p\to\bot)\to q}\cdot c_{p\to\bot}, c_{p\to q}, t_q\}\} = Qmax(E'')\end{aligned}$$

   In which case we have the same explicit beliefs as when we have applied $c_{q\to\bot}$ to t_q, our weakest evidence for q.

In this example we see that there are several ways of combining evidence to arrive at the same qmax sets and in turn the same explicit belief. One way is by applying our argument for $p \to \bot$ to the weakening of all our arguments for q, the other is by applying our argument for $q \to \bot$ to each argument for q.

This motivates the following proposition:

Proposition 2.3.6. Whenever we have some weakening term $t \in E$ s.t. $t \gg \phi$ and we have some term $t' \in E$ s.t. $t' \gg \phi \to \bot$ then:

$$Qmax(E \cup \{t'\cdot t\}) = Qmax(E \cup \{t'\cdot t'' \mid t'' \in sub_{\sqcup}(t)\})$$

Proof.

(⊆) Take any $e \in Qmax(E \cup \{t'\cdot t\})$

- e is quasi-consistent by definition.
- $e \subseteq E$ since $(t'\cdot t)_\bot \notin e$. So $e \subseteq E \cup \{t'\cdot t'' \mid t'' \in sub_{\sqcup}(t)\}$.
- e is closed under composition w.r.t. $E \cup \{t'\cdot t'' \mid t'' \in sub_{\sqcup}(t)\}$:
  * Suppose $\{t_1, t_2\} \subseteq e$ and $t_1\cdot t_2 \in E \cup \{t'\cdot t'' \mid t'' \in sub_{\sqcup}(t)\}$.
    If $t_1\cdot t_2 \in E \subseteq E \cup \{t'\cdot t\}$ then $t_1\cdot t_2 \in e$ since $e \in Qmax(E \cup \{t'\cdot t\})$.
    If $t_1\cdot t_2 \in \{t'\cdot t'' \mid t'' \in sub_{\sqcup}(t)\}$ then $t_1 = t'$ and $t_2 \in sub_{\sqcup}(t)$. But $t_2 \in e$

and $t_2 \in sub_\sqcup(t)$ would imply that $t \in e$ in which case $t' \cdot t \in e$ which cannot be the case. So $t_1 \cdot t_2 \notin \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}$.

  * Suppose $\{t_1, t_2\} \cap e \neq \emptyset$ and $t_1 \sqcup t_2 \in E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}$/ Then $t_1 \sqcup t_2 \in E$ so then $t_1 \sqcup t_2 \in e$ since $e \in Qmax(E \cup \{t' \cdot t\})$.
  * Suppose $t_1 \in e$ and $!t_1 \in E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}$. Then $!t_1 \in E$ and similar as previous case we have $!t_1 \in e$.

- e is quasi-closed under subterms by definition.
- e is maximal:
  Suppose $e \subseteq e' \in Qmax(E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\})$. Suppose, for contraposition that $s \notin e$. Suppose for induction, that $s = c_\phi \notin e$. Then $c_\phi \lightning_{E \cup \{t' \cdot t\}} e$ so there is $t''_\perp \in Clos^{E \cup \{t' \cdot t\}}(e \cup \{c_\phi\})$.
  If $t'' \in E$ then $c_\phi \lightning_E e$ so $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e'$.
  If $t'' = t' \cdot t$ then $c_\phi \in subterm(t') \cup subterm(t)$. i.e. $c_\phi \in subterm(t')$ or $c_\phi \in subterm(t)$. If $c_\phi \in subterm(t')$ then $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e$. If $c_\phi \in subterm(t)$ then there is $t''' \in sub_\sqcup(t)$ s.t. $c_\phi \in subterm(t''')$ so then $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e$. So then $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e'$.
  So we have that $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e'$ and thus $c_\phi \notin e'$.
  Composition cases are trivial via I.H.

($\supseteq$) Take any $e \in Qmax(E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\})$

- e is quasi-consistent by definition.
- $e \subseteq E$ since $(t' \cdot t'')_\perp \notin e$ for $t'' \in sub_\sqcup(t)$. So $e \subseteq E \cup \{t' \cdot t\}\}$.
- e is closed under composition w.r.t. $E \cup \{t' \cdot t\}$:
  * Suppose $\{t_1, t_2\} \subseteq e$ and $t_1 \cdot t_2 \in E \cup \{t' \cdot t\}$.
    If $t_1 \cdot t_2 \in E \subseteq E \cup \{t' \cdot t\}$ then $t_1 \cdot t_2 \in e$ since $e \in Qmax(E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\})$.
    If $t_1 \cdot t_2 \in \{t' \cdot t\}$ then $t_1 = t'$ and $t_2 = t$. But $t \in e$ would then imply that $t' \cdot t \in e$ which cannot be the case. So $t_1 \cdot t_2 \notin \{t' \cdot t\}$.
  * Suppose $\{t_1, t_2\} \cap e \neq \emptyset$ and $t_1 \sqcup t_2 \in E \cup \{t' \cdot t\}$/ Then $t_1 \sqcup t_2 \in E$ so then $t_1 \sqcup t_2 \in e$ since $e \in Qmax(E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\})$.
  * Suppose $t_1 \in e$ and $!t_1 \in E \cup \{t' \cdot t\}$. Then $!t_1 \in E$ and similar as previous case we have $!t_1 \in e$.
- e is quasi-closed under subterms by definition.
- e is maximal:
  Suppose $e \subseteq e' \in Qmax(E \cup \{t' \cdot t\})$. Suppose, for contraposition that $s \notin e$. Suppose for induction, that $s = c_\phi \notin e$. Then $c_\phi \lightning_{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}} e$ so there is $t''_\perp \in Clos^{E \cup \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}}(e \cup \{c_\phi\})$.
  If $t'' \in E$ then $c_\phi \lightning_E e$ so $c_\phi \lightning_{E \cup \{t' \cdot t\}} e'$.
  If $t'' \in \{t' \cdot t'' \mid t'' \in sub_\sqcup(t)\}$ then $c_\phi \in subterm(t') \cup subterm(t''')$ for some $t''' \in sub_\sqcup(t)$. i.e. $c_\phi \in subterm(t')$ or $c_\phi \in subterm(t''')$. If $c_\phi \in subterm(t')$ then $c_\phi \lightning_{E \cup \{t' \cdot t\}} e$. If $c_\phi \in subterm(t''')$ for some $t''' \in sub_\sqcup(t)$. Then $c_\phi \in subterm(t)$ so then $c_\phi \lightning_{E \cup \{t' \cdot t\}} e$. So then $c_\phi \lightning_{E \cup \{t' \cdot t\}} e'$.
  So we have that $c_\phi \lightning_{E \cup \{t' \cdot t\}} e'$ and thus $c_\phi \notin e'$.
  Composition cases are trivial via I.H.

■

This is useful in well organized evidence bases as deriving a contradiction with some term for ϕ means we have the derivation with the largest weakening for ϕ as well. So then deriving a contradiction with *some* term for ϕ is equivalent to deriving the contradiction for *all* terms admissible for ϕ.

Now suppose, we become aware of all contradictions in our evidence (definition 2.2.25). We then have that our set of explicit beliefs is consistent.

Proposition 2.3.7. Whenever we are aware of all contradictions in our evidence, our set of explicit beliefs is consistent.

Proof. Suppose, for contraposition, that our set of explicit beliefs is inconsistent. Then there is set of terms $e \subseteq \cap Qmax(E) \subseteq E$ such that $Clos^{\mathcal{L}^{t^{\gg}}}(e)$ contains a term for $\bot$. Since $e \subseteq \cap Qmax(E)$ we have that e is quasi-consistent w.r.t. E. So then it cannot be the case that there is $t_\bot \in Clos^E(e)$. So we are unaware of some contradiction in our evidence. ∎

On might then expect the converse to be the case as well i.e. whenever we have consistent explicit beliefs we are aware of all contradictions in our evidence. This is not the case however. Consider the following:

Example 2.3.8. Suppose $E = \{c_q, c_{p\to\bot}, c_{q\to p}, c_{q\to\bot}, c_{q\to p} \cdot c_q, c_{q\to\bot} \cdot c_q\}$. Here we have some unaware contradiction ($\{c_q, c_{q\to p}, c_{p\to\bot}, c_{q\to p} \cdot c_q\}$). When we consider our qmax sets however, we get the following:

$$Qmax(E) = \{\{c_q, c_{p\to\bot}, c_{q\to p}, c_{q\to p} \cdot c_q\}, \{c_{q\to\bot}, c_{p\to\bot}, c_{q\to p}\}\}$$

in which case $\mathcal{B}^e_E = \{p \to \bot, q \to p\}$ which is not inconsistent.

We are already losing trust in evidence for p for another reason, that being that we have built up the argument for p from another argument that we do not trust.
It is somewhat similar to disproving someones argument by saying that their assumptions are just false. If you then add the contradiction of p and $p \to \bot$ to your evidence it would be like saying "even if your assumptions were true, you would run into a problem with this".

To finish this section, we would like to compare this notion of qmax evidence sets with its inspiration [Gog16].

Recall example 1.3.9 in which we had become aware of a contradiction but could not express which contradiction we had become aware of. Let us go through this example again, but now in our framework with the addition of justification terms.

Example 2.3.9. Suppose we have evidence base $E = \{c_p, c_{p\to\bot}, c_q, c_{q\to\bot}\} \in$ **EB**. Since there is no $t \in E$ s.t. $t >> \bot$ we have that E is quasi-consistent and thus $Qmax(E) = \{E\}$ meaning that $\cap Qmax(E) = E$.

Suppose now we figure out that p and $p \to \bot$ is a contradiction. We get $E' = E \cup \{(c_{p\to\bot} \cdot c_p)\}$. Then:

$$Qmax(E') = \{\{c_{p\to\bot}, c_{q\to\bot}, c_q\}, \{c_p, c_{q\to\bot}, c_q\}\}$$

Since we now have $c_{p\to\bot} \cdot c_p$ in our evidence we have that we cannot have both subterms in the same qmax set due to quasi-consistency and closure under application. Then we get one qmax set with c_p in it and one with $c_{p\to\bot}$ in it. So we stop trusting these pieces of evidence.

Now let's say we also figure out that q and $q \to \bot$ is a contradiction. We do the same as before and get $E'' = E \cup \{(c_{p\to\bot} \cdot c_p)\} \cup \{(c_{q\to\bot} \cdot c_q)\} \cup \{(c_{p\to\bot} \cdot c_p) \sqcup (c_{q\to\bot} \cdot c_q)\}$. When we calculate our qmax sets we get the following:

$$Qmax(E'') = \{\{c_{p\to\bot}, c_{q\to\bot}\}, \{c_p, c_q\}, \{c_{p\to\bot}, c_q\}, \{c_p, c_{q\to\bot}\}\}$$

Now we see that there is no formula such that each Qmax evidence contains some evidence for it.

As we see, the addition justification terms allows us to express becoming aware of *certain* contradictions in our evidence, without necessarily becoming aware of all of them.

Moreover, consider example 1.3.10 where we have gained a belief in q by applying modus ponens to $p \rightarrow q$ and p. But we lose belief in p due to evidence $p \rightarrow \bot$ without losing the belief in q.

Let us consider such a scenario again in our framework:

Example 2.3.10. Suppose we have some set of evidence $E = \{c_p, c_{p\rightarrow q}\}$. Now from this we realize that we have evidence for q if we combine our evidence: $E' = \{c_p, c_{p\rightarrow q}, (c_{p\rightarrow q} \cdot c_p)\}$. Now we would have an explicit belief in p, $p \rightarrow q$ and q.

Now we gain some other piece of evidence $c_{p\rightarrow\bot}$. We apply MP (application) and become aware of a contradiction $\{c_p, c_{p\rightarrow\bot}\}$ so $E'' = \{c_p, c_{p\rightarrow q}, (c_{p\rightarrow q} \cdot c_p), c_{p\rightarrow\bot}, (c_{p\rightarrow\bot} \cdot c_p)\}$.

Then:

$$Qmax(E'') = \{\{c_p, c_{p\rightarrow q}, (c_{p\rightarrow q} \cdot c_p)\}, \{c_{p\rightarrow q}, c_{p\rightarrow\bot}\}\}$$

So we have one explicit belief: $p \rightarrow q$ and we no longer believe q due to losing our trust in the evidence we used to derive q.

We see that the addition of justification terms to this qmax calculation has added the expressivity that we set out to add.

2.4 Forms of Implicit beliefs

As noted before in section 1.2, implicit belief corresponds to all potentially derivable beliefs given explicit belief [Lev84]. Here we use this word 'derivable' which presupposes we have some proof system Λ.
Recall our definition of E-closure (definition 2.2.8) where we already assume our agent uses some logic Λ for their reasoning. We will use this closure w.r.t. the set of all admissible terms to define implicit belief. We will refer to such a full closure of an evidence base e as the $\Lambda-$closure of e.

Definition 2.4.1. *(Λ-closure)*

For evidence base $e \in \mathbf{EB}$ define the Λ-closure of e, denoted as $Clos^{\Lambda}(e)$ as:

$$Clos^{\Lambda}(e) = Clos^{\mathcal{L}^{t\gg}}(e)$$

We say a set e is Λ-closed if it is $\mathcal{L}^{t\gg}$-closed.

As opposed to previous discussed methods for defining implicit beliefs, we will, in this chapter, take a fully syntactic approach to define implicit belief and other forms of belief. Later in section 3.2 we will again consider a possible worlds approach to implicit belief.

Note that any Λ-closed set is always infinite. Even if Λ were only to have one axiom we would have that a term for this axiom exists within the Λ-closed set and then both closure under internalization and closure under weakening will result in this set being infinite.

This closure under weakening is now not only not restricted to some set of evidence, but also allowed to be redundant. If t is in some Λ closed set, then $t \sqcup t$ is also in it, and $t \sqcup t \sqcup ... \sqcup t$ as well.

Moreover, if $t \in E$ and E is Λ closed, then any weakening t' containing t ($t \in sub_{\sqcup}(t')$) will be contained in E. Such weakening terms which contain terms not contained in E we call *arbitrary weakenings*.

Definition 2.4.2. *(Arbitrary weakening)*
Whenever $t \in E$ and some t' is admissible for the same formula as t then $t \sqcup t'$ and $t' \sqcup t$ are arbitrary weakenings.

We say a set of evidence E is closed under arbitrary weakening whenever any arbitrary weakening is in E.

One useful property that these Λ-closures have is the following:

Proposition 2.4.3. For evidence base $E \in \mathbf{EB}$ we have that $t \in Clos^{\Lambda}(E)$ iff $t \in E$ or t is an axiom term or t is obtained by combining elements of E and Λ-axiom terms and arbitrary weakenings.

Proof. Induction on complexity of t:

- Suppose $t = c_{\phi}$.
  $c_{\phi} \in Clos^{\Lambda}(E)$ iff $c_{\phi} \in E$ or ϕ is a Λ-axiom since c_{ϕ} cannot be obtained by composition and $Clos^{\Lambda}(E)$ is the smallest Λ-closure containing E.

- Suppose $t = t_1 \cdot t_2$.
  then t cannot be an axiom term.
  - ($\longleftarrow$) Suppose $t \in E$ or t is obtained by composition and arbitrary weakenings of elements of E and axiom terms. Then t_1 and t_2 are as well. so by I.H. we have that $\{t_1, t_2\} \subseteq Clos^{\Lambda}(E)$ so then $t_1 \cdot t_2 \in Clos^{\Lambda}(E)$
  - ($\longrightarrow$) Suppose $t \in Clos^{\Lambda}(E)$. Since $Clos^{\Lambda}(E)$ is the smallest Λ-closure containing E it has to be the case that $\{t_1, t_2\} \subseteq Clos^{\Lambda}(E)$. Then by I.H. t_1 and t_2 are either axiom terms, contained in E or composed from these and arbitrary weakenings. Then t is composed of such.

- Suppose $t = t_1 \sqcup t_2$.
  $t_1 \sqcup t_2 \in Clos^{\Lambda}(E)$ iff $\{t_1, t_2\} \cap Clos^{\Lambda}(E) \neq \emptyset$ iff either $t_1 \in Clos^{\Lambda}(E)$ or $t_2 \in Clos^{\Lambda}(E)$ iff (I.H.) t_1 or t_2 is contained in E, is an axiom term or composed of these and arbitrary weakenings iff $t_1 \sqcup t_2$ is an arbitrary weakening.

- Suppose $t = !t'$.
  $!t \in Clos^{\Lambda}(E)$ iff $t \in Clos^{\Lambda}(E)$ iff (I.H.) t is an axiom term, contained in E or composed of such and arbitrary weakenings iff $!t$ is composed.

∎

Taking the Λ-closure of some evidence base E will only add terms that are admissible and are composed in a certain way of elements of E and Λ-axiom terms (which are atomic) and arbitrary weakenings. This means that the resulting set of evidence is itself an evidence base as well.

Fact 2.4.4. if E is an evidence base then $Clos^{\Lambda}(E)$ is an evidence base.

This Λ-closure will be the mechanism we use to derive more notions of belief from our evidence base.

2.4.1 Implicit belief

We may now want to define implicit beliefs as those formulae that follow (w.r.t. logic Λ) from our explicit beliefs. This would correspond to composing all terms that we can from all of our justification. This is our first notion of implicit belief that we will define, which we simply call *implicit belief*.

Definition 2.4.5. *(Implicit belief)*

For evidence base $E \in \mathbf{EB}$ we define the set of *implicit beliefs* as follows:

$$\mathcal{B}^i_E := \Phi(Clos^{\Lambda}(\bigcap Qmax(E)))$$

with $\Phi(.)$ as defined in definition 2.2.3.

In other words: we have $\phi \in \mathcal{B}^i_E$ iff there is term t_ϕ such that $t \in Clos^{\Lambda}(\bigcap Qmax(E))$.

Recall proposition 2.4.3 telling us that terms in the Λ-closure of some evidence base $E \in \mathbf{EB}$ will consist of terms in E, axiom terms, and compositions of these and arbitrary weakenings. This means that, whenever we have some justifications which we could compose, together with Λ-axioms, into a term admissible for ϕ we have an implicit belief in ϕ.

Since we are taking the Λ-closure of our set of justifications, we have that all of our explicit beliefs are contained in our set of implicit beliefs:

Proposition 2.4.6. For any evidence base $E \in \mathbf{EB}$ we have that :

$$\mathcal{B}^e_E \subseteq \mathcal{B}^i_E$$

Proof. Suppose $\phi \in \mathcal{B}^e_E = \Phi(\cap Qmax(E))$. Then there is some term $t_\phi \in \cap Qmax(E)$. Since $\cap Qmax(E) \subseteq Clos^{\Lambda}(\cap Qmax(E))$ we have that $t \in Clos^{\Lambda}(\cap Qmax(E)))$ which means that $\phi \in \Phi(Clos^{\Lambda}(\cap Qmax(E))) = \mathcal{B}^i_E$.

■

Note that all Λ-axioms will always be implicitly believed, as they are a part of every Λ-closure. Recall that application works as an explicit variant of the K-axiom. Closure under application then ensures that any Λ-validity is always implicitly believed. Now recall that justification logic uses an axiom internalization rule which acts as an explicit variant of the necessitation rule. Our closure under internalization in Λ-closures then ensures that for any Λ-validity ϕ we implicitly believe $t : \phi$ whenever $t \gg \phi$ with t composed of Λ-axioms. Thus ensuring that implicit belief is closed under an explicit variant of necessitation.

Thus our notion of implicit belief is closed under logical consequence and contains all Λ-validities.

Proposition 2.4.7. If $\phi \rightarrow \psi \in \mathcal{B}^i_E$ and $\phi \in \mathcal{B}^i_E$ then $\psi \in \mathcal{B}^i_E$.

Proof. Suppose $\{\phi \rightarrow \psi, \phi\} \subseteq \mathcal{B}^i_E$.
Then there is terms $t_{\phi\rightarrow\psi}$ and t_ϕ in $Clos^{\Lambda}(\cap Qmax(E))$. Since $Clos^{\Lambda}(\cap Qmax(E))$ is closed under composition we have that $t_{\phi\rightarrow\psi} \cdot t_\phi \in Clos^{\Lambda}(\cap Qmax(E))$ so then $\psi \in \mathcal{B}^i_E$

■

Proposition 2.4.8. If $\vdash_\Lambda \phi$ then $\phi \in \mathcal{B}^i_E$ and there is some term t such that $t : \phi \in \mathcal{B}^i_E$.

Proof. Suppose $\vdash_\Lambda \phi$. Then ϕ can be derived from Λ-axioms. Thus there exists a term t that consists of Λ-axiom terms. By proposition 2.4.3 we have that t in any Λ-closure. So $\phi \in \mathcal{B}^i_E$ for any $E \in \mathbf{EB}$. Moreover, since Λ-closures are closed under internalization we have that $!t_{t:\phi} \in Clos^\Lambda(\cap Qmax(E))$ for any $E \in \mathbf{EB}$. So $t : \phi \in \mathcal{B}^i_W$ for any $E \in \mathbf{EB}$.

■

2.4.2 Evidence based implicit belief

In previous section we defined *implicit belief* as all beliefs that logically follow from our explicit beliefs. But in this case we may be disregarding conclusions that are not only implicit in our set of justifications, but are also implicit among all qmax sets. More specifically, we can get the following:

Example 2.4.9. Suppose $E = \{c_p, c_r, c_{p\to r\to\bot}, c_{p\to r\to\bot} \cdot c_p, c_{p\to r\to\bot} \cdot c_p \cdot c_r\}$. We have evidence for p, evidence for r but also evidence that p and r can not both be true. Moreover we are aware of this contradiction in our evidence.

We will get 3 qmax sets:

$$Qmax(E) = \{\{c_p, c_r\}, \{c_p, c_{p\to r\to\bot}, c_{p\to r\to\bot} \cdot c_p\}, \{c_r, c_{p\to r\to\bot}\}\}$$

Then $\bigcap Qmax(E) = \emptyset$. And thus our implicit beliefs consist only of Λ-validities.

Suppose we have that $\vdash_\Lambda p \longrightarrow (p \vee r)$ and $\vdash_\Lambda r \longrightarrow (p \vee r)$. Then there is term t_1 admissible for $p \longrightarrow (p \vee r)$ and t_2 admissible for $r \longrightarrow (p \vee r)$ which are both built up from Λ-axiom terms, meaning that they occur in any $\Lambda-$closure.

If we take the Λ-closure of $\bigcap Qmax(E) = \emptyset$ then it would not contain any term t such that $t \gg (p \vee r)$ since this is not a Λ-validity.

However, looking at the qmax sets that we have, we see that, if we were to Λ-close these, they would all contain a term for $(p \vee r)$. Moreover, since Λ-closure is closed under weakening, there is a common element among the Λ-closures of the qmax sets.

Above example shows that there are some beliefs that are entailed by all our qmax sets, i.e. the Λ closure of each qmax set contains a term for ϕ, but these formulas are not implicitly believed.

For this reason we define *evidence based implicit belief* as whatever follows from all of our qmax evidence.

Definition 2.4.10. *(Evidence based implicit belief)*

For evidence base $E \in \mathbf{EB}$ we define the set of *evidence based implicit beliefs* as follows:

$$\mathcal{B}^{i(ev)}_E := \Phi(\bigcap\{Clos^\Lambda(e) \mid e \in Qmax(E)\})$$

with $\Phi(.)$ as defined in definition 2.2.3.

In other words: we have $\phi \in \mathcal{B}^{i(ev)}_E$ iff there is term t_ϕ such that, for each $e \in Qmax(E)$, we have $t \in Clos^\Lambda(e)$.

So we have an evidence based implicit belief in ϕ if there is some argument for ϕ that occurs in the Λ-closure of each qmax set. Again recall proposition 2.4.3 telling us that any element of a Λ-closure of evidence base e consists of elements of e, axiom terms, and compositions of these and arbitrary weakenings.
This means we have an evidence based implicit belief in ϕ is there is a argument for ϕ that can be composed of elements of e, axiom terms and arbitrary weakenings for each $e \in Qmax(E)$.

Now of course, we may again argue that we want to have an evidence based implicit belief in ϕ if we can construct an argument for ϕ in each qmax set $e \in Qmax(E)$, where this argument does not have to be the same argument across all qmax set Λ-closures.
But we have that any Λ-closure is closed under arbitrary weakenings. So, recalling proposition 2.3.4, we have that, if there is an argument for ϕ in the Λ-closure of each qmax set, then the weakening of all these terms is in all these Λ-closures as well.

To put it more formally:

Proposition 2.4.11. For evidence base $E \in \mathbf{EB}$ we have the following:

$$\Phi(\bigcap\{Clos^{\Lambda}(e) \mid e \in Qmax(E)\}) = \bigcap\{\Phi(Clos^{\Lambda}(e)) \mid e \in Qmax(E)\}$$

i.e. there is t_ϕ s.t. $t \in Clos^{\Lambda}(e)$ for $e \in Qmax(E)$ iff each $e \in Qmax(E)$ has some $t_\phi \in Clos^{\Lambda}(e)$.

Proof.

($\subseteq$) Suppose there is t_ϕ s.t. $t \in Clos^{\Lambda}(e)$ for each $e \in Qmax(E)$. Then each $e \in Qmax(E)$ has some $t \in Clos^{\Lambda}(e)$ s.t. $t \gg \phi$.

($\supseteq$) Suppose each $e \in Qmax(E)$ has that there is some $t_\phi \in Clos^{\Lambda}(e)$. Since each $Clos^{\Lambda}(e)$ is closed under arbitrary weakening we have that the weakening of all such t is contained in each $Clos^{\Lambda}(e)$. So we have a specific term t_ϕ in $Clos^{\Lambda}(e)$ for $e \in Qmax(e)$.

■

Again, for the same reasons as for implcit belief, we have that evidence based belief is closed under an explicit variant of necessitation. Moreover, evidence based implicit belief is closed under logical consequence.

Moreover, we have that any implicit belief is also an evidence based implicit belief.

Proposition 2.4.12. For evidence base $E \in \mathbf{EB}$ we have:

$$\mathcal{B}_E^i \subseteq \mathcal{B}_E^{i(ev)}$$

Proof. Since $\cap Qmax(E) \subseteq e$ for $e \in Qmax(E)$ we have that $Clos^{\Lambda}(\cap Qmax(E)) \subseteq Clos^{\Lambda}(e)$ for $e \in Qmax(E)$. Then $Clos^{\Lambda}(\cap Qmax(E)) \subseteq \cap\{Clos^{\Lambda}(e) \mid e \in Qmax(E)\}$.

■

From this it also follows that all explicit beliefs are evidence based implicit beliefs since $\mathcal{B}_E^e \subseteq \mathcal{B}_E^i \subseteq \mathcal{B}_E^{i(ev)}$.

2.4.3 Inconsistent explicit belief and the principle of explosion

For both forms of implicit belief we have that an explicit belief implies an (evidence based) implicit belief.
This is also the case when we have inconsistent explicit beliefs. For instance:

Example 2.4.13. Suppose $E = \{c_p, c_{p\to\bot}\}$. Since E is quasi-consistent we have that $Qmax(E) = \{E\}$. If we now take the Λ-closure of E we will have $c_{p\to\bot} \cdot c_p \in Clos^{\Lambda}(E)$. Which means that we have an implicit belief/evidence based implicit belief in $\bot$.

In this example we have some contradictory evidence of which we have not yet become aware that they are contradictory. This results in inconsistent explicit belief and, in turn, implicitly believing a contradiction.
If we then have ex falso for Λ i.e. $\vdash_\Lambda \bot \to \phi$ we will then have an implicit belief/evidence based implicit belief in every formula ϕ.

Proposition 2.4.14. If $\vdash_\Lambda \bot \to \phi$ then inconsistent explicit belief implies an implicit belief in any ϕ.

Proof. Suppose $\vdash_\Lambda \bot \to \phi$. Moreover, suppose $\mathcal{B}^e_E$ is inconsistent, meaning that there is $e_\bot \subseteq \cap Qmax(E)$ s.t. there is $t_\bot \in Clos^{\Lambda}(e_\bot)$.
Then $e_\bot \subseteq e$ for $e \in Qmax(E)$. Then $Clos^{\Lambda}(e_\bot) \subseteq Clos^{\Lambda}(e)$ for $e \in Qmax(E)$ so then $t \in Clos^{\Lambda}(e)$ for $e \in Qmax(E)$. Since $t \gg \bot$, $t \in Clos^{\Lambda}(e)$ for $e \in Qmax(E)$ and $\vdash_\Lambda \bot \to \phi$ we have that there is term $t'_\phi \in Clos^{\Lambda}(e)$ for $e \in Qmax(E)$. i.e. we have an implicit belief in ϕ, whatever ϕ is.

■

Inconsistent explicit belief leads to implicit belief in everything (given that we have ex falso). In the spirit of interpreting implicit belief as potential ones this is rather strange then. As a belief in a contradiction is something any rational agent will never have.

However, putting into a different persepective, these forms of implicit belief we have just discussed can be seen as some Λ-logically omniscient critic, looking at the evidence and derivations you have made so far. And then showing you what follows from what you think is consistent in your evidence. If this contains inconsistencies the critic will show you exactly where these inconsistencies can be found.
In a way it can be interpreted as an outside view on an agent's evidence and the derivations they have made. It could then contain beliefs that a rational agent would never be able to actually explicitly believe.

Considering that we can interpret this form of implicit belief as an omniscient agent showing you where your arguments/justifications run into trouble, one can then wonder what the explicit beliefs of this logically omniscient agent are.

2.5 Potential explicit belief

Previous definitions of implicit belief gave us forms of implicit belief which could be inconsistent. They can contain formulas that a rational agent, given enough time and resources, would never actually come to believe. As noted, it could be interpreted as a Λ-omniscient agent, considering the evidence you have gathered and your thoughts on what makes up maximally consistent sets of evidence, and then showing where you have missed some details and what the result of these missed derivations are.
This could make us consider another form of implicit belief. The explicit beliefs of this

Λ-omniscient agent, the explicit belief where we have unlimited time and resources to combine the evidence we have and then figure out what the maximally consistent subsets are. We will call such belief *potential explicit belief* as it corresponds to explicit beliefs any rational agent is able to obtain given enough time and recources.

Practically, what this means is that we want to first Λ-close our set of evidence and then we calculate our explicit beliefs via qmax calculation.

Definition 2.5.1. *(Potential Explicit Belief)*

For any evidence base $E \in \mathbf{EB}$ we define the set of *potential explicit beliefs* as follows:

$$\mathcal{B}_E^{\mathcal{P}} := \Phi(\bigcap Qmax(Clos^{\Lambda}(E)))$$

i.e. $\phi \in \mathcal{B}_E^{\mathcal{P}}$ iff there is t_ϕ in each $e^* \in Qmax(Clos^{\Lambda}(E))$.

So we have a potential explicit belief in ϕ if there is some argument for ϕ that is in the Λ-closure of each qmax set of $Clos^{\Lambda}(E)$, i.e. there is some justification for ϕ in $Clos^{\Lambda}(E)$. Notice then that the only difference between potential explicit belief and explicit belief is this closure of E in the potential variant.
This is the reason it is called potential explicit belief as it consists of explicit beliefs that we may obtain when we can combine all of our evidence, without any computational limitation. Another way to then look at such a fomr of belief as some sort of revision to the limit. We keep adding new pieces of evidence composed of the evidence we have, Λ-axioms and arbitrary weakenings, and we keep using the qmax belief calculation to revise our beliefs.

Again, considering that $Clos^{\Lambda}(E)$ is closed under arbitrary weakenings, we have that there are two equivalent definitions for this notion of belief.

Proposition 2.5.2. For evidence base $E \in \mathbf{EB}$ we have the following:

$$\Phi(\bigcap Qmax(Clos^{\Lambda}(E))) = \bigcap\{\Phi(e^*) \mid e^* \in Qmax(Clos^{\Lambda}(E))\}$$

i.e. there is t_ϕ s.t. $t_\phi \in e^*$ for $e^* \in Qmax(Clos^{\Lambda}(E))$ iff each $e^* \in Qmax(Clos^{\Lambda}(E))$ has some $t_\phi \in e^*$.

Proof.

($\subseteq$) Suppose there is t_ϕ s.t. $t \in e^*$ for each $e^* \in Qmax(Clos^{\Lambda}(E))$. Then each $e^* \in Qmax(Clos^{\Lambda}(E))$ has some $t \in e^*$ s.t. $t \gg \phi$.

($\supseteq$) Suppose each $e^* \in Qmax(Clos^{\Lambda}(E))$ has that there is some $t_\phi \in e^*$. Since $Clos^{\Lambda}(E)$ is closed under arbitrary weakening we have that the weakening of all such t is contained in $Clos^{\Lambda}(E)$. So, by composition closure of each e^*, we have a specific term t_ϕ in e^* for $e^* \in Qmax(Clos^{\Lambda}(E))$.

∎

Consider again example 2.4.13 where we obtain an implicit belief in $\bot$ (and thus in everything if $\vdash_\Lambda \bot \rightarrow \phi$). Let us now consider this example with this notion of potential explicit belief.

Example 2.5.3. Suppose $E = \{c_p, c_{p\rightarrow\bot}\}$. If we take the Λ-closure of E it will contain $c_{p\rightarrow\bot} \cdot c_p$. If we then look at the qmax sets of this Λ-closure we see that none of these can contain this argument since it is admissible for $\bot$.

We see that, given the same evidence base E, we may have an implicit belief in a contradiction, namely when our explicit beliefs are inconsistent, but this contradiction does not occur in our set of potential explicit beliefs. This is because potential explicit beliefs come from justifications we get from the Λ-closure of our evidence base. These justifications may not be admissible for $\bot$ due to quasi-consistency.

Now consider again example 2.4.13. We have that $Qmax(E) = \{E\}$ so taking the Λ-closure of each qmax set is the same as just taking the Λ-closure of E. If we then do the qmax calculation again we get $Qmax(Clos^{\Lambda}(E))$ which is exactly what we use to calculate potential explicit beliefs for E.

We can make it a bit more involved by looking at an evidence base E that is not quasi-consistent, but that does lead to inconsistent explicit beliefs.

Example 2.5.4. Suppose $E = \{c_p, c_q, c_{p\to\bot}, c_{q\to\bot}, c_{p\to\bot} \cdot c_p\}$. Then

$$Qmax(E) = \{\{c_p, c_q, c_{q\to\bot}\}, \{c_{p\to\bot}, c_q, c_{q\to\bot}\}\}$$

The Λ-closure of both qmax sets contains $c_{q\to\bot} \cdot c_q$ which means we have an implicit belief in $\bot$.
Considering potential explicit beliefs we will first take the Λ-closure of E which includes $c_{q\to\bot} \cdot c_q$. Then considering the qmax sets of $Clos^{\Lambda}(E)$ will result in four qmax sets: $\{c_p, c_q, ...\}, \{c_p, c_{q\to\bot}, ...\}, \{c_{p\to\bot}, c_q, ...\}, \{c_{p\to\bot}, c_{q\to\bot}, ...\}$.
Now computing qmax sets of closure of qmax sets will result in the same sets.

As it turns out, computing qmax sets of Λ-closure of qmax sets is the same effect as Λ-closing our evidence base first and then computing its qmax sets. fig. 2.1 shows a visual representation of this. theorem 1 shows a proof for this.

This phenomenon is in line with the intuition that we can think of potential explicit beliefs as beliefs that are obtained by revising in the limit. It tells us that we can slowly start adding derivations built up from our original set E and axioms and recompute/revise our explicit beliefs accordingly, i.e. by computing maximally consistent subsets of evidence up to our awareness.
If we keep doing this infinitely we will end up with potential explicit belief.

What follows are claims and proofs (proposition 2.5.5, lemmas 2.5.6 and 2.5.7, and theorem 1) to show that taking qmax sets, closing these, and taking qmax sets again, is equivalent to first closing our evidence and then considering its qmax sets.

Proposition 2.5.5. For $e^* \in Qmax(Clos^{\Lambda}(E))$ we have that $e^* \subseteq Clos^{\Lambda}(e)$ for $e \in qmax^{\uparrow E}(e^* \cap E)$.

Proof. Suppose $t \in e^*$. Then $t \in Clos^{\Lambda}(E)$ so we know from proposition 2.4.3 that t is an axiom term, $t \in E$ or t is composed of axiom terms, elements of E and arbitrary weakenings.

Take any $e \in qmax^{\uparrow E}(e^* \cap E)$.

- If t is atomic then it is in E or it is an axiom term.
  - If t is an axiom term then $t \in Clos^{\Lambda}(e)$ by defintion of Λ-closure.
  - If $t \in E$ then $t \in e^* \cap E \subseteq e \subseteq Clos^{\Lambda}(e)$.

(I.H.) For t' of lower complexity than t we have that $t' \in e^*$ implies that $t' \in Clos^{\Lambda}(e)$.

- If $t = t_1 \cdot t_2$ then $\{t_1, t_2\} \subseteq e^*$ since $t \in e^*$ and e^* is qmax. Then by I.H. we have

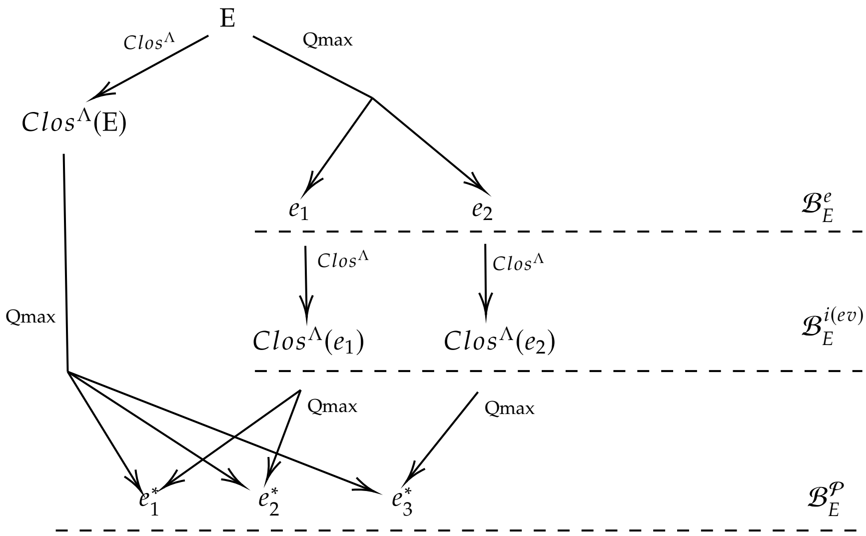

Figure 2.1: Left path closes E first and then considers its qmax sets, right path considers qmax sets of E, then closes these after which it considers the qmax sets of these closures, both paths end up in the same sets.

that $\{t_1, t_2\} \subseteq Clos^{\Lambda}(e)$. Then $t_1 \cdot t_2 \in Clos^{\Lambda}(e)$ since $Clos^{\Lambda}(e)$ is closed under composition.

- If $t = t_1 \sqcup t_2$ then $\{t_1, t_2\} \cap e^* \neq \emptyset$ since $t \in e^*$ and e^* is qmax. Then by I.H. we have that $\{t_1, t_2\} \cap Clos^{\Lambda}(e) \neq \emptyset$. Then $t_1 \sqcup t_2 \in Clos^{\Lambda}(e)$.
- If $t = !t'$ then $t' \in e^*$ since $t \in e^*$ and e^* is qmax. Then by I.H. we have that $t' \in Clos^{\Lambda}(e)$. Then $!t' \in Clos^{\Lambda}(e)$.

So $e^* \subseteq Clos^{\Lambda}(e)$ for $e \in qmax^{\uparrow E}(e^* \cap E)$.

■

Lemma 2.5.6. For any $e \in Qmax(E)$ we have that $Qmax(Clos^{\Lambda}(e)) \subseteq Qmax(Clos^{\Lambda}(E))$.

Proof. Take any $e^* \in Qmax(Clos^{\Lambda}(e))$.

- $e^* \subseteq Clos^{\Lambda}(e) \subseteq Clos^{\Lambda}(E)$ since $e \subseteq E$.
- e^* is already quasi consistent.
- Suppose $\{t_1, t_2\} \subseteq e^*$ and $t_1 \cdot t_2 \in Clos^{\Lambda}(E)$. Since $e^* \subseteq Clos^{\Lambda}(e)$ we have that $\{t_1, t_2\} \subseteq Clos^{\Lambda}(e)$ so then $t_1 \cdot t_2 \in Clos^{\Lambda}(e)$ so then $t_1 \cdot t_2 \in e^*$ because $e^* \in Qmax(Clos^{\Lambda}(e))$.
- Suppose $\{t_1, t_2\} \cap e^* \neq \emptyset$ and $t_1 \sqcup t_2 \in Clos^{\Lambda}(E)$. Then $\{t_1, t_2\} \cap Clos^{\Lambda}(e) \neq \emptyset$ so then $t_1 \sqcup t_2 \in Clos^{\Lambda}(e)$. Then $t_1 \sqcup t_2 \in e^*$ since $e^* \in Qmax(Clos^{\Lambda}(e))$.
- Suppose $t \in e^*$ and $!t \in Clos^{\Lambda}(E)$. Then $t \in Clos^{\Lambda}(e)$ so $!t \in Clos^{\Lambda}(e)$ so $!t \in e^*$ since $e^* \in Qmax(Clos^{\Lambda}(e))$.
- e^* is quasi closed under subterms6 since it is qmax.
- Maximality.
  Suppose $e^* \subseteq e' \in Qmax(Clos^{\Lambda}(E))$.
  First we prove:

(*) $e' \cap E \subseteq e$
Suppose $t \in e' \cap E$ then $t \not\lightning_{Clos^{\Lambda}(E)} e'$ so $t \not\lightning_{E} (e' \cap E)$ so $t \not\lightning_{E} (e^* \cap E)$ since $e^* \cap E \subseteq e' \cap E$.

- Suppose $t = c_\phi$ then there is no $t'_\perp \in Clos^E((e^* \cap E) \cup \{c_\phi\})$. Then, by definition 2.2.22, all $e'' \in qmax^{\uparrow E}(e^* \cap E)$ have $c_\phi \in e''$. Since $e^* \cap E \subseteq Clos^{\Lambda}(e) \cap E = e$ we have that $e \in qmax^{\uparrow e^* \cap E}$. So $c_\phi \in e$.
- Suppose $t = t_1 \cdot t_2$ then $\{t_1, t_2\} \subseteq e^* \cap E$. By I.H. $t_1, t_2 \subseteq e$ so $t_1 \cdot t_2 \in e$ since $t \in E$ and $e \in Qmax(E)$.
- other cases similarly follow from I.H.

(**) $e \in qmax^{\uparrow E}(e' \cap E)$
Since $e' \cap E \subseteq e$ (*), we have that $e \in qmax^{\uparrow E}(e' \cap E)$.

Suppose now, for contraposition, that $t \notin e^*$.
If $t \in Clos^{\Lambda}(e)$ then $t \lightning_{Clos^{\Lambda}(e)} e^*$. Then $t \lightning_{Clos^{\Lambda}(E)} e'$ since $e \subseteq e'$ and $Clos^{\Lambda}(e) \subseteq Clos^{\Lambda}(E)$ so then $t \notin e'$.

If $t \notin Clos^{\Lambda}(e)$ then $t \notin e$ so $t \notin Clos^{\Lambda}(e'')$ for some $e'' \in qmax^{\uparrow E}(e' \cap E)$ because of (**) and then $t \notin e'$ because of proposition 2.5.5 ($e' \subseteq Clos^{\Lambda}(e'')$ for $e'' \in qmax^{\uparrow E}(e' \cap E)$).

By contraposition we have that $e' \subseteq e^*$. So e^* is maximal.

Then $e^* \in Qmax(Clos^{\Lambda}(E))$.

■

Lemma 2.5.7. if $e^* \in Qmax(Clos^{\Lambda}(E))$ then there is some $e \in Qmax(E)$ such that $e^* \in Qmax(Clos^{\Lambda}(e))$.

Proof. Suppose $e^* \in Qmax(Clos^{\Lambda}(E))$ and $e \in qmax^{\uparrow E}(e^* \cap E)$.

- $e^* \subseteq Clos^{\Lambda}(e)$.
  proposition 2.5.5.
- $e^* \in Qmax(Clos^{\Lambda}(E))$ so it is quasi-consistent.
- e^* is closed under commposition w.r.t. $Clos^{\Lambda}(e)$ since it is closed under composition w.r.t. $Clos^{\Lambda}(E)$.
- e^* is closed under (qmax) subterms since it is qmax.
- Maximal.
  Suppose $e^* \subseteq e' \in Qmax(Clos^{\Lambda}(e))$. Suppose $t \notin e^*$, we will show that $t \notin e'$ from which it follows that $e' \subseteq e^*$.
  If $t \notin Clos^{\Lambda}(e)$ then $t \notin e'$ since $e' \subseteq Clos^{\Lambda}(e)$. So suppose $t \in Clos^{\Lambda}(e)$.
  Then, since $t \notin e^*$, we have that $t \lightning_{Clos^{\Lambda}(E)} e^*$.
  - If $t = c_\phi$ then there is some $t' \in Clos^{\Lambda}(E)$ such that $t' \in Clos^{Clos^{\Lambda}(E)}(e^* \cup \{c_\phi\}) = Clos^{\Lambda}(e^* \cup \{c_\phi\})$. Since $e^* \subseteq Clos^{\Lambda}(e)$ and $c_\phi \in Clos^{\Lambda}(e)$ we have that $e^* \cup \{c_\phi\} \subseteq Clos^{\Lambda}(e)$ so $t' \in Clos^{\Lambda}(Clos^{\Lambda}(e)) = Clos^{\Lambda}(e)$. Then $c_\phi \lightning_{Clos^{\Lambda}(e)} e^*$ in which case $c_\phi \lightning_{Clos^{\Lambda}(e)} e'$ so $c_\phi \notin e'$.
  - If $t = t_1 \cdot t_2$ then $\{t_1, t_2\} \not\subseteq e^*$. If either t_1 or t_2 is not contained in e^* then, by I.H. we have that it is not contained in e' in which case $t_1 \cdot t_2 \notin e'$.
  - If $t = t_1 \sqcup t_2$ then $\{t_1, t_2\} \cap e^* = \emptyset$. Then, by I.H. we have that $\{t_1, t_2\} \cap e' = \emptyset$ in which case $t_1 \sqcup t_2 \notin e'$.
  - If $t = !t'$ then $t' \notin e^*$ so by I.H. we have that $t' \notin e'$ so then $!t' \notin e'$

So $e^* \in Qmax(Clos^{\Lambda}(e))$.

∎

Theorem 1. $Qmax(Clos^{\Lambda}(E)) = \cup\{Qmax(Clos^{\Lambda}(e)) \mid e \in Qmax(E)\}$.

Proof.

(⊆) Suppose $e^* \in Qmax(Clos^{\Lambda}(E))$. By lemma 2.5.7 we then have that there is $e \in Qmax(E)$ such that $e^* \in Qmax(Clos^{\Lambda}(e)) \subseteq \cup\{Qmax(Clos^{\Lambda}(e)) \mid e \in Qmax(E)\}$.

(⊇) Suppose $e^* \in \cup\{Qmax(Clos^{\Lambda}(e)) \mid e \in Qmax(E)\}$. Then $e^* \in Qmax(Clos^{\Lambda}(e))$ for some $e \in Qmax(E)$. By lemma 2.5.6 we know that $Qmax(Clos^{\Lambda}(e)) \subseteq Qmax(Clos^{\Lambda}(E))$ so then $e^* \in Qmax(Clos^{\Lambda}(E))$.

∎

This also tells us that any potential explicit belief is also an implicit belief.

Proposition 2.5.8. $\mathcal{B}_E^{\mathcal{P}} \subseteq \mathcal{B}_E^{i(ev)}$.

Proof. Suppose $t_\phi \in e^*$ for each $e^* \in Qmax(Clos^{\Lambda}(E))$. Since $Qmax(Clos^{\Lambda}(e)) \subseteq Qmax(Clos^{\Lambda}(E))$ for $e \in Qmax(E)$ we have that $t_\phi \in e'$ for $e' \in Qmax(Clos^{\Lambda}(e))$ for $e \in Qmax(E)$. Since each $e' \in Qmax(Clos^{\Lambda}(e)$ has that $e \subseteq Clos^{\Lambda}(e)$ we have that $t_\phi \in Clos^{\Lambda}(e)$ for $e \in Qmax(E)$.

∎

So we do have that a potential explicit belief is also an implicit belief. The converse does not necessarily hold however. Since, as seen in example 2.4.13, we can have an implicit belief in a contradiction which we may never have a potential explicit belief for.

However, this relies on unknown contradictions. Whenever we are aware of all contradicitions however, we have that the two types of implicit belief are equivalent.

Proposition 2.5.9. Whenever we are aware of all contradictions in $E \cup \{c_\phi \mid \phi$ is a Λ axiom $\}$ then $\mathcal{B}_E^{\mathcal{P}} = \mathcal{B}_E^{i(ev)}$.

Proof. Suppose we are aware of all contradictions in $E \cup \{c_\phi \mid \phi$ is a Λ axiom $\}$. Then any Λ-inconistent subset $e \subseteq E$(i.e. there is some $t_\perp \in Clos^{\Lambda}(e)$) has that there is $t_\perp \in Clos^{E}(e)$. Then :

(*) $Qmax(Clos^{\Lambda}(e)) = \{Clos^{\Lambda}(e)\}$ for $e \in Qmax(E)$

Because $Clos^{\Lambda}(e)$ is quasi consistent.

$$\begin{aligned}
&\mathcal{B}_E^{\mathcal{P}} \\
=&\Phi(\bigcap Qmax(Clos^{\Lambda}(E))) \\
=&\Phi(\bigcap(\cup\{Qmax(Clos^{\Lambda}(e)) \mid e \in Qmax(E)\}) && \text{(theorem 1)} \\
=&\Phi(\bigcap(\cup\{\{Clos^{\Lambda}(e)\} \mid e \in Qmax(E)\}) && (*) \\
=&\Phi(\bigcap(\{Clos^{\Lambda}(e) \mid e \in Qmax(E)\}) \\
=&\mathcal{B}_E^{i(ev)}
\end{aligned}$$

■

2.6 Chapter conclusion

In this chapter we have covered the mechanisms of deriving explicit beliefs from our defined evidence bases. Deriving such explicit beliefs relies on considering maximal quasi-consistent (qmax) sets of evidence. The manner in which we define such qmax sets builds on the work by Gogoladze on explicit evidence models [Gog16] by replacing *logical formulas* as explicit evidence with *justification terms* as explicit evidence.

When we covered explicit evidence models in section 1.3, we noted on two examples which may occur in this formal setting. The first (example 1.3.9) showed an agent automatically becoming aware of all contradictions in their evidence (and future evidence) once they become aware of one contradiction in their evidence. The other (example 1.3.10) showed an agent gaining a belief in some formula q by applying Modus ponens to p and $p \rightarrow q$, after which the agent obtains some evidence against p, thus losing their belief in p, but not in q. Even though this belief in q was obtained using evidence for p.

In both cases the behavior comes as a result of not being able to express derivations that agents have made. This motivated an integration of justification terms into this derivation of belief via qmax subsets of evidence. We have seen that the addition of justification terms, which can be seen as encodings of derivations, succeeds in adding the expressivity needed to overcome these anomalies.

Apart from explicit belief, we have defined three more notions of belief. These are *implicit belief* corresponding to the formulas that logically follow from explicit beliefs, *evidence based implicit belief* which are obtained by considering the composition closure of each qmax set, and *potential explicit belief* which is obtained by first closing the set of evidence and then considering qmax sets. This last notion can be conceived of as the explicit beliefs of a logically omniscient agent.

On that note, we do need to mention that, even though we have described mechanisms to derive such notions of belief, we cannot express and formalize these notions with the language used throughout this chapter (defined in definition 2.1.1) since we only considered propositional formulae, contradiction, implication and justification.

We may then want to add formulas expressing these notions to the language and provide a formal logic and semantics describing the mechanisms to derive these notions of belief. However, note that our notion of explicit belief can intuitively be described as "there is some justification term for ..."[4]
Moreover, the other notions of belief we have defined involve a similar quantification over justification terms.

This quantification over justification terms brings us to a problem. Say we have an explicit belief in ϕ. How do we know which justifications we actually have for ϕ?
For instance, we may have an explicit belief in ϕ which may come from justification $c_{\psi\rightarrow\phi} \cdot c_{\psi}$. But it may also come from justification $c_{\psi\rightarrow\phi} \cdot c_{\chi\rightarrow\psi} \cdot c_{\chi}$ or $c_{\psi_n\rightarrow\phi} \cdot \ldots \cdot c_{\psi_1\rightarrow\psi_2} \cdot c_{\psi_1}$. If were then to formally describe our notion of explicit belief we would have to account for this quantification over terms. Which is something we want to avoid.

In the next chapter we will propose a solution to this problem and define a formal logic and semantics which considers explicit belief and implicit belief. For this we will have to make some changes to the mechanisms we have discussed so far. These changes we will come

[4] Recall section 1.4 where an interpretation of modal operator $\Box$ as "there exists some proof for..." is discussed. Such an interpretation (similar to ours for notions of belief) contains an existential quantification. The non-constructive nature of such an existential operator was part of the original motivation of introducing such proof/justification terms.

back to again in chapter 4, as they bring about some less desirable behavior of their own.

Formalizing the other notions of belief (evidence based implicit belief, potential explicit belief) we discuss further in chapter 4 and leave for future work.

Chapter 3

The logic of Quasi-Consistent Belief

In this chapter we will be considering a formal logic describing explicit and implicit belief. As mentioned in section 2.6 adding formulas for explicit belief $B^e\phi$ and implicit belief $B^i\phi$ to the language leads to a problem involving quantification over justification terms. In section 3.1 we will suggest a way to deal with this problem, which consists of *certification* of evidence bases. This certification will require some changes to the mechanisms defined in previous chapter.

In section 3.2 we propose the logic of quasi-consistent belief. We first define the language of QCB in section 3.2.1 which extends the language in previous chapter with *implicit justifications*[1] which intuitively can be considered as justifications which may be derived from our *explicit justifications*.
After which, in section 3.2.2 we define modified versions of the mechanisms we have discussed in previous chapter to account for *certification*. Moreover, we define models and semantics in section 3.2.3, provide an axiomatization for QCB in section 3.2.4 and prove completeness of QCB w.r.t. finite evidential base models in section 3.2.5.

3.1 Certification

As mentioned before in section 3.1, the manner in which we have defined explicit belief (definition 2.3.1) and implicit belief (definition 2.4.5) in chapter 2 can intuitively be described as "there is some justification for ..."
If we want to formalize such notions of belief it would require us to account for this quantification. This is something we want to avoid as there are infinitely many terms admissible for formulas we may have beliefs in.

However, recalling the set of terms we have considered in chapter 2, we see that, for any formula ϕ, we know of the existence of a specific term which is admissible for ϕ, which is c_ϕ, i.e. the *certificate* for ϕ.

One solution, which we will call *certification*, to the problem we have is to put an extra condition on our evidence bases. Even though we may not be able to determine what all our justifications for our explicit beliefs and implicit beliefs are, if we require an atomic term (certificate) c_ϕ to be in our evidence base whenever we have some argument for ϕ then we can always refer to this atomic term c_ϕ when determining belief. In this manner we

[1] Intuitively, the language in previous chapter can be considered as the *explicit fragment* of the language we will be considering in this chapter.

may define explicit belief as an abbreviation, namely that c_ϕ is an *explicit justification* for our belief in ϕ. Similarly, we may define implicit belief in ϕ as c_ϕ being an *implicit justification* for a belief in ϕ.

In a way we can interpret such an atomic term c_ϕ as a *canonical term* or *canonical evidence* for ϕ. This, in a way, relates to weakening terms as well. These certificates for formulas ϕ can now be interpreted as the weakening of all terms admissible for ϕ. The result of this (as we will discuss later in the chapter as well) is that this certification makes weakening to a certain extent obsolete. We will discuss this further in section 3.2.1.

3.2 Logic of Quasi-Consistent Belief (QCB)

In this section we will consider *the logic of quasi consistent belief (QCB)* which aims to express explicit and implicit beliefs alongside explicit and implicit justifications. The approach is to add a condition to our evidence bases that, whenever we have some argument for ϕ, we have some certificate (i.e. atomic term) for ϕ. Such evidence bases we call *certified evidence bases*. This certificate will then act as a justification we can always refer to when determining explicit beliefs.
However, note that such justifications are derived from our qmax belief computation. So such a condition should not only be required for evidence bases, but also enforced on its qmax sets. So we need some notion of *certified qmax subsets* as well. Moreover, such an approach brings with it some complications when turning to implicit belief, which motivates a model that can take into account background beliefs which may not necessarily follow from our explicit beliefs. Such models we call *evidential base models*.

3.2.1 Language

We will first describe the language we will be considering for the logic of quasi-consistent belief. As noted before, our qmax sets should be *certified* as well, meaning that they will contain a certificate for ϕ whenever they contain an argument for ϕ. Note that this means that all qmax sets will, by definition, have a certain justification in common whenever they all contain some argument for ϕ, namely a certificate for ϕ. This was the reason we introduced weakening terms before. Which, for this section, makes weakening unnecessary. We will therefore set aside weakening for now, but in chapter 4 we will highlight an example that such a *certification* of evidence bases and qmax sets brings about for which weakening terms in combination with well organized evidence bases (definition 2.2.6) may offer a solution.

Moreover, we will now consider two kinds of justification, one explicit ($t :^e \phi$) and one implicit ($t :^i \phi$). These different kinds of justification we will use to define explicit and implicit belief respectively.
If we consider the “old" language in chapter 2, definition 2.1.1 and omit weakening terms, then it corresponds to the explicit fragment of the language defined in this chapter.

For this chapter we let $\mathcal{L}$ refer to the language of QCB and $\mathcal{L}^t$ refer to the set of terms we will be considering for QCB.

Definition 3.2.1. *(Language $\mathcal{L}$, Explicit fragment $\mathcal{L}_e$)*

Given countable set of atomic propositions $Prop$, we define formulas $\phi \in \mathcal{L}$, the *language of quasi-consistent belief*, and $t \in \mathcal{L}^t$ by mutual induction as follows:

$$\phi ::= p \mid \bot \mid \phi \to \phi \mid t :^e \phi \mid t :^i \phi$$

$$t ::= c_\phi \mid t \cdot t \mid !t$$

We further define *the explicit fragment* $\mathcal{L}_e$ and $\mathcal{L}_e^t$ of $\mathcal{L}$ and $\mathcal{L}_e^t$ respectively. These consist of formulas and terms built up without using *implicit justification*. To put it more formal: formulas $\phi \in \mathcal{L}_e$ and terms $t \in \mathcal{L}_e^t$ are defined by mutual induction as follows:

$$\phi ::= p \mid \bot \mid \phi \to \phi \mid t :^e \phi$$

$$t ::= c_\phi \mid t \cdot t \mid !t$$

Moreover, we define the following abbreviations:

$$\begin{aligned}
\neg\phi &:= \phi \to \bot \\
\top &:= \neg\bot \\
\phi \vee \psi &:= (\neg\phi) \to \psi \\
\phi \wedge \psi &:= \neg(\neg\phi \vee \neg\psi) \\
\phi \leftrightarrow \psi &:= (\phi \to \psi) \wedge (\psi \to \phi) \\
B^e\phi &:= c_\phi :^e \phi \\
B^i\phi &:= c_\phi :^i \phi
\end{aligned}$$

We have propositional formulas $p \in Prop$, contradiction $\bot$ which is never true, material implication $\to$, explicit justification $t :^e \phi$ denoting that $t \in \mathcal{L}^t$ is an explicit justification for believing ϕ, and $t :^i \phi$ denoting that t is an implicit justification for ϕ. An implicit justification can be seen as one that we can construct out of our explicit justifications. An explicit/implicit belief in ϕ is now defined as having explicit/implicit atomic justification for ϕ.

Since we will be working with background beliefs, and we will be determining our implicit beliefs in terms of our explicit beliefs and such background beliefs, we will make a distinction between the full language and the explicit fragment[2] of $\mathcal{L}$ and $\mathcal{L}^t$.

As we are now be considering implicit justifications as well, we will use an admissibility relation taking this into account.

Definition 3.2.2. *(Admissibility 2)*
We call *Admissibility* the smallest relation $\gg \subseteq \mathcal{L}^t \times \mathcal{L}$ such that:

- $c_\phi \gg \phi$
- If $t_1 \gg (\phi \to \psi)$ and $t_2 \gg \phi$ then $(t_1 \cdot t_2) \gg \psi$
- If $t \gg \phi$ then $!t \gg (t :^e \phi)$ and $!t \gg (t :^i \phi)$

Again we write $t^{\gg}$ whenever there is some formula ϕ such that $t \gg \phi$, and $t^{\not\gg}$ whenever there is no such ϕ.

We call $\mathcal{L}^{t^{\gg}}$ the set of evidence terms that are admissible, i.e. $\mathcal{L}^{t^{\gg}} = \{t \in \mathcal{L}^t \mid t^{\gg}\}$

We denote $\mathcal{L}_e^{t^{\gg}}$ is the set of all admissible explicit evidence, i.e. $\mathcal{L}_e^{t^{\gg}} = \{t \in \mathcal{L}_e^t \mid t^{\gg}\}$

We define $\Phi : \mathcal{P}(\mathcal{L}^t) \to \mathcal{P}(\mathcal{L})$ and as follows:

$$\Phi(E) = \{\phi \in \mathcal{L} \mid t \gg \phi \text{ for some } t \in E\}$$

[2] This division between explicit and full language is similar to the approach in [Lor20] where such background beliefs are interpreted as the *common ground* of a set of agents.

We define $\Phi_e : \mathcal{P}(\mathcal{L}^t) \to \mathcal{P}(\mathcal{L}_e)$ as follows:

$$\Phi_e(E) = \{\phi \in \mathcal{L}_e \mid t \gg \phi \text{ for some } t \in E\}$$

Note that two different functions for extracting formulas from sets of terms are necessary as we have internalization terms which are admissible for formulas of the form $t :^e \phi$ and $t :^i \phi$. If a set of terms E contains no internalization terms then the two are equivalent.

Fact 3.2.3. When $E \subseteq \mathcal{L}_e^t$ contains no internalization terms then $\Phi(E) = \Phi_e(E)$.

Definition 3.2.4. *(Subterm (in)consistency)*
We say a term t is subterm inconsistent if there is some $t' \in subterm(t)$ s.t. $t' \gg \bot$. We say a term t is subterm consistent if it is not subterm inconsistent i.e. there is no $t' \in subterm(t)$ s.t. $t' \gg \bot$.

3.2.2 Certified Evidence base, qmax and explicit belief

As mentioned before, we will be adding a condition to our evidence bases and qmax sets which will ensure that we always have some justification for ϕ we can refer to when we want to determine if we have explicit belief in ϕ, namely the certificate c_ϕ.
However, we get into troublesome situations when we do this for all arguments. Recall section 2.5 where we noted that the beliefs of an omiscient agent can be interpreted as taking the closure of our evidence base and then considering the explicit beliefs of that closure.
If we add certificates for all of our arguments/derivations we can get into the following situation:

Example 3.2.5. Suppose we have an evidence base $E = \{c_p, c_q, c_{q\to\bot}\}$. We have some inconsistent evidence of which we have not yet become aware of: namely c_q and $c_{q\to\bot}$. However, we have no evidence against p so whatever derivations we make with the evidence we have, we should keep a belief in p.
Suppose we were to close E with the condition that we have to add certificates for each derivation we make.
Since $c_{q\to\bot}$ and c_q are contained in E we will have $c_{q\to\bot} \cdot c_q$ in its closure. Then $c_\bot$ is contained in its closure as well. But then, due to ex falso, we have $c_{\bot\to\phi}$ in the closure of E as well, for any ϕ. So for $\phi = p \to \bot$ as well. Then we can construct term $c_{\bot\to p\to\bot} \cdot c_\bot$ in which case we have a certificate $c_{p\to\bot}$ in E's closure as well.
But then we have c_p and $c_{p\to\bot}$ in the closure of E which means we have $c_{p\to\bot} \cdot c_p$ in the closure of E as well.
If we then consider the qmax sets of the full closure of E then there will be one containing $c_{p\to\bot}$ and not c_p. Which means we have no (potential) explicit belief in p.

If we certify all derivations then it seems that a logically omniscient agent can not have any non-trivial beliefs if we have some contradictory evidence of which we have not yet become aware of. This is something we want to avoid.

However, do note that this loss of belief in p resulted from adding certificates for derivation made using the principle of explosion (ex falso). If we limit certification to those derivations that do not use this, we avoid such a situation.

We therefore define certified evidence base as follows:

Definition 3.2.6. *(Certfified Evidence Base)*
A *Certified evidence base* is a set $E \subseteq \mathcal{L}^t$ of evidence terms such that:

- E is closed under subterms[a].
- E consists only of admissible evidence: (each $t \in E$ has that $t^{\gg}$).
- If $t \in E$ is subterm consistent then there is some ϕ such that $t \gg \phi$ and $c_\phi \in E$.

We call such a set a *certified explicit evidence base* if $E \subseteq \mathcal{L}^t_e$ i.e. E consists only of explicit evidence terms.

We call the family of certified evidence bases **EB-C**, moreover we call the family of certified explicit evidence bases **EB-C**$_e$.

[a] note that this need not be quasi-closed anymore since we are not considering weakening terms.

Since **EB-C** and **EB-C**$_e$ are closed under arbitrary unions we have that there is a largest certified evidence base which is exactly the set of all admissible terms:

$$\bigcup \textbf{EB-C} = \mathcal{L}^{t^{\gg}} \qquad \bigcup \textbf{EB-C}_e = \mathcal{L}_e^{t^{\gg}}$$

Again, we consider a logic Λ which we assume the agent uses to reason. Later in section 3.2.4 we define this logic Λ.

Definition 3.2.7. *(Certified E-closure)*

For a certified evidence base $E \in$ **EB-C**, we say that $e \subseteq E$ is certified E-closed iff

- If $t_1 \cdot t_2 \in E$ then $(\{t_1, t_2\} \subseteq e$ iff $t_1 \cdot t_2 \in e)$.
- If $!t \in E$ then $(t \in e$ iff $!t \in e)$.
- If $c_\phi \in E$ and ϕ is a Λ-axiom then $c_\phi \in e$.
- If $c_\phi \in E$ and $t_\phi \in e$ is term consistent then $c_\phi \in e$.

We denote $Clos_c^E(e)$ the smallest certified E-closure containing e.

Moreover, we define here quasi-consistency and maximally quasi-consistency with the condition of certification.

Definition 3.2.8. *(Certified Quasi-consistency, maximally certified quasi-consistent)*
We call a set of explicit evidence $e \subseteq E$ *certified quasi-consistent w.r.t.* E if there is no term $t \in Clos_c^E(e)$ s.t. $t \gg \bot$.

We call a set $e \subseteq E$ *maximally certified quasi-consistent* whenever e is certified quasi-consistent w.r.t. E but none of its proper extensions are.

We denote $Qmax_c(E) = \{e \subseteq E \mid e$ is maximally certified quasi-consistent w.r.t. $E\}$ as the set of all maximally certified quasi consistent sets w.r.t. E.

Similar as before if we have that E is quasi-consistent w.r.t. itself then E is the only certified qmax set.

Fact 3.2.9. If $E \in$ **EB-C** is certified quasi-consistent w.r.t. E then $Qmax_c(E) = \{E\}$

Moreover, since all certified qmax set are subterm closed and consist only of admissible evidence they are also certified evidence bases.

Fact 3.2.10. For each $e \in Qmax_c(E)$ we have $e \in$ **EB-C**.

And certified qmax sets of E are maximal so they are certified closed w.r.t. E.

Fact 3.2.11. For $e \in Qmax_c(E)$ we have that $Clos_c^E(e) = e$.

With that we have defined all the essential mechanisms to derive our explicit beliefs, which are then defined as before, but now using certified qmax. For reasons that become apparent later when considering implicit belief and background beliefs we require explicit beliefs to consists only of explicit formulas $\phi \in \mathcal{L}_e$.

> **Definition 3.2.12.** *(Certified Explicit belief)*
>
> Given certified explicit evidence base $E \in$ **EB-C**$_e$ we define set of explicit beliefs $\mathcal{B}^e_{E_c}$.
>
> $$\mathcal{B}^e_{E_c} := \Phi_e(\cap Qmax_c(E))$$

Recall proposition 2.3.4 where we noted that, for collected evidence bases, we have that the set of explicit beliefs is equivalent to the set of formulas for which we have some argument in all our qmax sets. This came as a result of having an evidence base that has some form of closure under weakening, since it will ensure that the weakening of terms for ϕ is a justification for ϕ whenever each qmax set contains an argument for ϕ.
As noted before, the addition of *certification* in evidence bases and qmax sets results in weakening terms being unnecessary. A certificate for ϕ now takes the role of the weakening of all arguments for ϕ. As a result we again have that, whenever we have an argument for ϕ in all our certified qmax sets, we have some justification for ϕ i.e. an explicit belief in ϕ.

Proposition 3.2.13. For certified evidence base $E \in$ **EB-C** we have that:

$$\Phi_e(\cap Qmax_c(E)) = \cap\{\Phi_e(e) \mid e \in Qmax(E)\}$$

i.e. we have some justification for ϕ (explicit belief in ϕ) iff each certified qmax set contains some argument for ϕ.

> ***Proof.***
>
> ($\subseteq$) Suppose $\phi \in \Phi_e(\cap Qmax_c(E))$. Then there is term t such that $t \in e$ for $e \in Qmax_c(E)$ and $t \gg \phi$. Then $\phi \in \Phi_e(e)$ for $e \in Qmax_c(E)$ so therefore $\phi \in \cap\{\Phi_e(e) \mid e \in Qmax_c(E)\}$.
>
> ($\supseteq$) Suppose $\phi \in \cap\{\Phi_e(e) \mid e \in Qmax_c(E)\}$. Then each $e \in Qmax_e(E)$ has some term $t \in e$ such that $t \gg \phi$. Then each $e \in Qmax_c(E)$ has that $c_\phi \in e$ since each $e \in Qmax_e(E)$ is certified. Then $c_\phi \in \cap Qmax_c(E)$ so therefore $\phi \in \Phi_e(\cap Qmax_c(E))$.
>
> ■

For implicit beliefs we will have to take into account background beliefs, which do not follow necessarily from our explicit beliefs.

3.2.3 Evidential base models

We will consider evidential base models since we want to incorperate a notion of background belief into our models. Such background beliefs are implicit beliefs that do not necessarily

follow from our implicit beliefs.
To do this we set aside the fully syntactic approach and take an approach which takes possible atomic valuations into account, these will serve as our variant of possible worlds. Background beliefs can then be encoded by eliminating certain atomic valuation. It bears many similarities with the approach in [Lor20]. However, we will be considering a fully introspective variant.

Definition 3.2.14. *(Evidential base model)*

An evidential base model is a tuple $M = (W, E)$ where

- $W \subseteq \mathcal{P}(Prop)$ is a set of possible atomic valuations
- $E \in$**EB-C**$_e$ is a certified explicit evidence base.

we say an evidential base model $M = (W, E)$ is full when $W = \mathcal{P}(Prop)$, i.e. all valuations are considered possible.

Such full evidential base models can be considered as having no background beliefs. Recall section 1.2.2 and definition 1.2.4 where we noted that, if the *Context* (LDA's version of possible worlds) consists of all belief bases, then implicit belief is just that which follows from explicit belief. Here we have the same for full evidential base models.

Since we are defining implicit belief in ϕ as having a certificate for ϕ as implicit justification for ϕ we will use it to inductively define what implicit justifications are.

Definition 3.2.15. *(Truth conditions)*

For evidential base model $M = (W, V)$ we define truth conditions for formula in $\mathcal{L}$ as follows:

$$
\begin{array}{ll}
M, w \vDash p & \text{iff } p \in w \\
M, w \nvDash \bot & \\
M, w \vDash \phi \to \psi & \text{iff } M, w \nvDash \phi \text{ or } M, w \vDash \psi \\
M, w \vDash t :^e \phi & \text{iff } t \in \cap Qmax_c(E) \text{ and } t \gg \phi \\
 & \\
M, w \vDash c_\phi :^i \phi & \text{iff } M, v \vDash \phi \text{ whenever } M, v \vDash \mathcal{B}^e_{E_c} \\
M, w \vDash t_1 \cdot t_2 :^i \phi & \text{iff } M, w \vDash t_1 :^i (\psi \to \phi) \text{ and } t_2 :^i \psi \text{ for some } \psi \\
M, w \vDash !t :^i \phi & \text{iff } M, w \vDash \phi \text{ and } !t \gg \phi
\end{array}
$$

Where $M, v \vDash \Phi$ is an abbreviation for $M, v \vDash \phi$ for $\phi \in \Phi$.

Notice that, implicit and explicit belief ($c_\phi :^e \phi$ and $c_\phi :^i \phi$) do not depend on the atomic valuation where it is evaluated, thus we have they are global modalities.

Fact 3.2.16. $M, w \vDash t :^e \phi$ iff $M \vDash t :^e \phi$ and $M, w \vDash t :^i \phi$ iff $M \vDash t :^i \phi$

Moreover, since implicit belief ($c_\phi :^i \phi$) is defined by considering all valuations that satisfy all explicit beliefs, we do not terms for implicit beliefs to occur in our evidence bases, i.e. our evidence bases should consist of terms in the explicit fragment if the language. Otherwise we allow for paradox.

Consider again definitions 1.2.3 and 1.2.4 where, on multi-agent belief models in LDA, implicit belief is defined by considering all doxastic alternatives which are determined by

considering all “possible worlds" that make all explicit beliefs true. We have the same here for our notion of implicit belief, that being $c_\phi :^i \phi$.
We can as well define relation $R \subseteq W \times W$ which behaves the same way:

Definition 3.2.17. *(Doxastic alternatives)*

For evidential base model $M = (W, E)$ we define doxastic alternatives relation $R \subseteq W \times W$ as follows:

$$wRv \text{ iff } M, v \vDash \mathcal{B}^e_{E_c}$$

In which case it is not hard to see that:

$$M, w \vDash B^i\phi \text{ iff } M, v \vDash \phi \text{ whenever } wRv$$

We will define models using such relations to derive belief, and use these to prove completeness.

3.2.4 Axiomatization QCB

We propose the following rules and axioms for the logic of quasi-consistent belief (QCB).

Definition 3.2.18. *(Proof system QCB)*

Rules

- MP
- B^i-Necessition B^i: From $\vdash \phi$ infer $\vdash B^i\phi$

Propositional logic axioms

Explicit Justification

- $\neg B^e \bot$
- $\neg(t :^e \phi)$ whenever $t \notin \mathcal{L}_e^{t\gg}$ or $\phi \notin \mathcal{L}_e$
- $(t_1 \cdot t_2 :^e \psi) \rightarrow (t_1 :^e (\phi \rightarrow \psi) \wedge t_2 :^e \phi)$ whenever $t_1 \gg \phi \rightarrow \psi$ and $t_2 \gg \phi$
- $(!t :^e (t :^e \phi)) \rightarrow (t :^e \phi)$ whenever $t \gg \phi$
- $t :^e \phi \rightarrow B^e\phi$

Implicit Justification

- $\neg(t :^i \phi)$ whenever $t \not\gg \phi$
- $(t_1 \cdot t_2 :^i \psi) \leftrightarrow (t_1 :^i (\phi \rightarrow \psi) \wedge t_2 :^i \phi)$ whenever $t_1 \gg \phi \rightarrow \psi$ and $t_2 \gg \phi$
- $(!t :^i (t :^i \phi)) \leftrightarrow (t :^i \phi)$
- $\neg t :^i \phi \rightarrow B^i(\neg t :^i \phi)$
- $t :^i \phi \rightarrow B^i\phi$

Connecting

- $t :^e \phi \rightarrow t :^i \phi$

- $!t :^i (t :^e \phi) \leftrightarrow (t :^e \phi)$
- $\neg t :^e \phi \to B^i(\neg t :^e \phi)$

Proposition 3.2.19. The following formulas are provable theorems in the logic QCB:

- $B^i(\phi \to \psi) \to B^i\phi \to B^i\psi$
- $B^iB^i\phi \leftrightarrow B^i\phi$
- $B^i(B^e\phi) \leftrightarrow (B^e\phi \vee B^i\bot)$
- $B^i\neg B^i\phi \leftrightarrow (\neg B^i\phi \vee B^i\bot)$
- $B^i\neg B^e\phi \leftrightarrow (\neg B^e\phi \vee B^i\bot)$
- $\neg B^e\phi$ for $\phi \notin \mathcal{L}_e$.
- $\neg B^eB^i\phi$
- $\neg B^e(\neg B^i\phi)$

Definition 3.2.20. *QCB-Maximally Consistent (sub)Set*

We say a set of formulas $\Gamma \subseteq \mathcal{L}$ is a QCB-maximally consistent set when it is QCB-consistent and no proper extensions of it are QCB-consistent. We call such a set a QCB-MCS.

We say a set of formulas $\Gamma \subseteq \Sigma \subseteq \mathcal{L}$ is a QCB maximally consistent subset of Σ if Γ is QCB-consistent and no proper extension of Γ is QCB-consistent and a subset of Σ.
We will call such a set a QCB-MCS$^\Sigma$.

Lemma 3.2.21. *Lindenbaum Lemma*
Any QCB-consistent set of formulas Γ_0 can be extended to a QCB-MCS Γ^*. If $\Gamma_0 \subseteq \Sigma$ then Γ_0 can be extended to a QCB-MCS$^\Sigma$.

3.2.5 Completeness

Here we will prove completeness of QCB w.r.t. our intended evidential base models. Recall that we could define doxastic alternatives relation from which our implicit beliefs can be derived. We will first prove completeness w.r.t. a non-standard semantics, in terms of pseudo models (where the doxastic alternatives relation is taken as primitive and not so tightly connected to explicit beliefs), and then show that the pseudo models are equivalent to our intended models.

We first need to define an auxiliary notion (pre-models), which facilitates the definition of pseudo-models.

Definition 3.2.22. *(Pre-model)*

A pre-model is a tuple $M = (W, E, R, ||.||)$ where

- W is a set of worlds,
- $E \in \mathbf{EB\text{-}C}_e$ is a certified explicit evidence base
- $R \subseteq W \times W$ is an accesibility relation

- $\|.\| : Prop \to \mathcal{P}(W)$.

Definition 3.2.23. *(Pre-model truth conditions)*

For pre-model $M = (W, E, R, V)$ we have the following:

$$\begin{array}{ll}
M, w \vDash p & \text{iff } w \in \|p\| \\
M, w \nvDash \bot & \\
M, w \vDash \phi \to \psi & \text{iff } M, w \nvDash \phi \text{ or } M, w \vDash \psi \\
M, w \vDash t :^e \phi & \text{iff } t \in \cap Qmax_c(E) \text{ and } t \gg \phi \\
 & \\
M, w \vDash c_\phi :^i \phi & \text{iff } M, v \vDash \phi \text{ whenever } wRv \\
M, w \vDash t_1 \cdot t_2 :^i \phi & \text{iff } M, w \vDash t_1 :^i (\psi \to \phi) \text{ and } t_2 :^i \psi \text{ for some } \psi \\
M, w \vDash !t :^i \phi & \text{iff } M, w \vDash \phi \text{ and } !t \gg \phi
\end{array}$$

Notice that the cases for implicit justification w.r.t. our intended models is only different for the case of implicit belief i.e. $c_\phi :^i \phi$.

Again we have that implicit and explicit justifications are global modalities. As a result of this, any two worlds satisfying the same propositional formulas (called p-equivalence) have that they agree on all formula in $\mathcal{L}$:

Proposition 3.2.24. If $M, w \vDash p$ iff $M, w' \vDash p$, then $M, w \vDash \phi$ iff $M, w' \vDash \phi$.

Proof. By induction on ϕ

Suppose $M, w \vDash p$ iff $M, w' \vDash p$ for $p \in Prop$

- $\phi = p \in Prop$. From assumption we have $M, w \vDash p$ iff $M, w' \vDash p$
- $\phi = \bot$. $M, v \nvDash \bot$ for any $v \in W$.
- $\phi = \psi \to \chi$.
  Follows from I.H.
- $\phi = t :^e \psi$.
  Since $t :^e \phi$ is a global modality we have that $M, w \vDash t :^e \phi$ iff $M, w \vDash t :^e \phi$
- $\phi = t :^i \psi$.
  Since $t :^i \phi$ is a global modality we have that $M, w \vDash t :^i \phi$ iff $M, w \vDash t :^i \phi$

■

Definition 3.2.25. *(certified quasi-consistent pre-models, pseudo-models, strong pseudo-models)*

A pre-model $M = (W, E, R, \|.\|)$ is certified quasi consistent if E is certified quasi-consistent w.r.t. E, meaning that $Qmax_c(E) = \{E\}$.

A *pseudo-model* is a pre-model $M = (W, E, R, \|.\|)$ satisfying the following:

1. wRv implies wRv for all $w, w', v \in W$.
2. wRv implies $v \vDash B^e_{E_c}$.

A pseudo model is *strong* if any two p-equivalent worlds are identical, i.e. if ($w \vDash p$ iff $w' \vDash p$) then $w = w'$.

For our completeness proof we will need a non-standard notion of complexity of formulas and terms since we can not use the standard notion of complexity as "subterms/subformulas". This notion of complexity we define next.

Definition 3.2.26. *($\mathcal{L}$-formula complexity C)*

We define complexity function $C : \mathcal{L}^t \cup \mathcal{L} \to \mathbb{N}$ as follows:

$$\begin{aligned}
C(c_\phi) &:= C(\phi)\\
C(t_1 \cdot t_2) &:= 1 + max(C(t_1), C(t_2))\\
C(!t) &:= 1 + C(t)\\
C(p) &:= 0\\
C(\bot) &:= 0\\
C(\phi \to \psi) &:= 1 + max(C(\phi), C(\psi))\\
C(t :^e \phi) &:= 1 + max(C(t), C(\phi))\\
C(t :^i \phi) &:= 1 + max(C(t), C(\phi))
\end{aligned}$$

For which we have the following:

Proposition 3.2.27. For any term $t \in \mathcal{L}^t$ and formula $\phi \in \mathcal{L}$ we have the following:

1. If $t \gg \phi$ then $C(t) \geq C(\phi)$.
2. If ϕ is a proper subformula of ψ then $C(\phi) < C(\psi)$.
3. If t is a proper subterm of t' then $C(t) < C(t')$.
4. If $t \gg \phi$ and $t' \gg \psi$ and t is a proper subterm of t' then $C(t :^e \phi) < C(t' :^i \psi)$ and $C(t :^i \phi) < C(t' :^i \psi)$.

Proof.

1. Suppose $t \gg \phi$. We will prove that $C(t) \geq C(\phi)$ by induction on the (standard) subterm-complexity of t.
   - $C(c_\phi) = C(\phi)$
   - If $t = t_1 \cdot t_2$ then $t_1 \gg \psi \to \phi$ and $t_2 \gg \psi$ for some ψ.

$$\begin{aligned}
C(t_1 \cdot t_2) &= 1 + max(C(t_1), C(t_2))\\
&\geq 1 + max(C(\psi \to \phi), C(\psi)) && \text{(I.H)}\\
&= 1 + max(1 + max(C(\psi), C(\phi)), C(\psi))\\
&= 1 + 1 + max(C(\psi), C(\phi))\\
&> max(C(\psi), C(\phi))\\
&\geq C(\phi)
\end{aligned}$$

   - If $t =!t'$ then $\phi = t :^e \psi$ or $\phi = t :^i \psi$. Then $C(\phi) = C(t' :^e \psi) = C(t' :^i \psi) = 1 + max(C(t'), C(\psi))$ By I.H. we have that $C(t') \geq C(\psi)$ so then

$max(C(t'), C(\psi)) = C(t')$ so then:

$$\begin{aligned} C(!t') &= 1 + C(t') \\ &= 1 + max(C(t'), C(\psi)) \\ &= C(\phi) \end{aligned}$$

2. Suppose ϕ is a proper subformula of ψ. We will prove that $C(\phi) < C(\psi)$ by induction on the (standard) subformula complexity of ψ.
   - $\psi = p$ or $\psi = \bot$ cannot be the case since ϕ is a proper subformula of ψ. So the base case is vacuously true.
   - Suppose $\psi = \chi \to \sigma$. Then ϕ is a subformula of χ or a subformula of σ. If $\phi = \chi$ we have that $C(\phi) = C(\chi)$. If $\phi = \sigma$ we have that $C(\phi) = C(\sigma)$. By I.H. we have that $C(\phi) < C(\chi)$ if ϕ is a proper subformula of χ and $C(\phi) < C(\sigma)$ if it is a proper subformula of σ.
     In all these cases we have that $C(\phi) \leq max(C(\chi), C(\sigma)) < 1+max(C(\chi), C(\sigma)) = C(\psi)$.
   - Suppose $\psi = t :^e \chi$ or $\psi = t :^i \chi$. Then ϕ is a subformula of χ. If $\phi = \chi$ then $C(\phi) = C(\chi)$. If ϕ is a proper subformula of χ, then by I.H. we have that $C(\phi) < C(\chi)$. So then $C(\phi) \leq C(\chi) \leq max(C(t), C(\chi)) < 1 + max(C(t), C(\chi)) = C(\psi)$.
3. Suppose t is a proper subterm of t'. We will prove that $C(t) < C(t')$ by induction on the (standard) subterm complexity of t'.
   - $t' = c_\phi$ cannot be the case since t is a proper subterm of t'. So the base case is vacuously true.
   - Suppose $t' = t_1 \cdot t_2$. Then t is a subformula of t_1 or a subformula of t_2. If $t = t_1$ then $C(t_1) = C(t)$, if $t = t_2$ then $C(t) = C(t_2)$. If t is a proper subterm of t_1 then, by I.H. we have that $C(t) < C(t_1)$. Similarly, when t is a proper subterm of t_2 we have that $C(t) < C(t_2)$. In all cases we have that $C(t) \leq max(C(t_1), C(t_2)) < 1 + max(C(t_1), C(t_2)) = C(t')$.
   - Suppose $t' =!t''$. Then t is a subterm of t''. If $t = t''$ then $C(t) = C(t'')$. If t is a proper subterm of t'' then, by I.H., we have that $C(t) < C(t'')$. In either case we have that Then $C(t) \leq C(t'') < 1 + C(t'') = C(t')$.
4. Suppose $t \gg \phi$ and $t' \gg \psi$ and t is a proper subterm of t'. Then:

$$\begin{aligned} C(t :^e \phi) = C(t :^i \phi) &= 1 + max(C(t), C(\phi)) \\ &= 1 + C(t) && (t \gg \phi, 1.) \\ &< 1 + C(t') && (t \text{ prop subterm of } t', 3.) \\ &= 1 + max(C(t'), C(\psi)) && (t' \gg \psi, 1.) \\ &= C(t' :^e \psi) = C(t' :^i \psi) \end{aligned}$$

■

Theorem 2. QCB is sound and complete w.r.t. finite strong quasi-consistent pseudo-models.

Proof. Suppose ϕ^* is QCB-consistent formula. We will construct a finite pre model M as follows:

Take smallest sets $\Sigma^\Phi \subseteq \mathcal{L}$ and $\Sigma^t \subseteq \mathcal{L}^t$ such that:

- $\phi^* \in \Sigma^\Phi$.
- Σ^Φ is closed under subformulas.

- If $\phi \in \Sigma^\Phi$ and ϕ is not of the form $\psi \to \bot$ then $\phi \to \bot \in \Sigma^\Phi$, i.e. Σ^Φ is closed under single negation.
- Σ^t is closed under subterms.
- If $t \in \Sigma^t$ and $t \gg \phi$ then $\{(t :^e \phi), (t :^i \phi)\} \subseteq \Sigma^\Phi$ and $c_\phi \in \Sigma^t$.
- If $\{(t :^e \phi), (t :^i \phi)\} \cap \Sigma^\Phi \neq \emptyset$ then $t \in \Sigma^t$.
- If $c_\phi \in \Sigma^t$ then $\phi \in \Sigma^\Phi$.

Note that such Σ^Φ and Σ^t exist for any ϕ^* and are finite.

Since ϕ^* is consistent we have that, by Lindenbaum lemma, there is QCB-maximally consistent subset of Σ^Φ we denote as Γ s.t. $\phi^* \in \Gamma$.

Let $M = M_\Gamma^{\Sigma^\Phi} = (W, E, R, \|.\|)$ as follows:

- $W = \{\Theta \subseteq \Sigma \mid \Gamma$ and Θ agree on all $t :^e \phi$ and $t :^i \phi\}$.
- $E = \{t \mid (t :^e \phi) \in \Gamma$ for some $\phi \in \mathcal{L}_e \cap \Sigma\}$.
- $\Theta R \Delta$ iff $c_\phi :^i \phi \in \Theta$ implies $\phi \in \Delta$ for each $\phi \in \Sigma$.
- $\|p\| = \{\Theta \in W \mid p \in \Theta\}$.

Then:

1. E is a certified quasi-consistent evidence base.
    - E is closed under subformulas:
      Suppose $t \in E$ then $t :^e \phi \in \Gamma$ for some $\phi \in \mathcal{L}_e \cap \Sigma^\Phi$. If $t = c_\phi$ then t has no subformulas. If $t = t_1 \cdot t_2$ then $t_1 \cdot t_2 :^e \phi \in \Gamma$ so $t_1 :^e \psi \to \phi \in \Gamma$ and $t_2 :^e \psi \in \Gamma$ so then $\{t_1, t_2\} \subseteq E$. If $t = !t'$ then $!t :^e \phi \in \Gamma$ so then $t' :^e \psi \in \Gamma$ so $t' \in E$.
    - $E \subseteq \mathcal{L}^{t^\gg}$ since, suppose that $t \notin \mathcal{L}^{t^\gg}$. Then $t \not\gg \phi$ for any ϕ so then we have $\neg(t :^i \phi) \in \Gamma$ for all ϕ. So then by contraposition we have $\neg(t :^e \phi) \in \Gamma$ for all ϕ, i.e. $(t :^e \phi) \notin \Gamma$ for all ϕ. So then $t \notin E$.
    - E consists of explicit evidence due to $\neg(t :^e \phi)$ whenever $t \notin \mathcal{L}_e^t$ or $\phi \notin \mathcal{L}_e$.
    - E is quasi-consistent since it consists of explicit justifications which cannot be admissible for $\bot$, and the fact that E is closed under subterms.
    - E is certified since $t :^e \phi \to B^e \phi$ so then $t_\phi \in E$ implies $c_\phi \in E$. Moreover, all such t are term consistent since E is quasi-consistent.
2. Truth lemma:

   We will prove the following claim:

   $$\text{for } \phi \in \Sigma^\Phi \text{ and } \Theta \in W \text{ we have that } M, \Theta \vDash \phi \text{ iff } \phi \in \Theta.$$

   By induction on complexity measure C (definition 3.2.26).
   induction hypothesis: For $\psi \in \Sigma^\Phi$ and $\Theta \in W$ such that $C(\psi) < C(\phi)$ we have that $M, \Theta \vDash \psi$ iff $\psi \in \Theta$.
    - $p \in \Theta$ iff $\Theta \in \|p\|$ iff $M, \Theta \vDash p$
    - $\bot \notin \Theta$ since Θ is consistent. Moreover, $\bot$ is never true on pre-models.
    - $t :^e \psi \in \Theta$ iff $t :^e \psi \in \Gamma$ iff $t_\psi \in E = \cap Qmax(E)$ iff $M, \Theta \vDash t :^e \psi$
    - $\psi \to \chi$ follows from I.H.

- Suppose $\phi = c_\psi :^i \psi$.

(→) Suppose $c_\psi :^i \psi \in \Theta$ and take any Δ s.t. $\Theta R \Delta$. Then, for any $c_\chi :^i \chi \in \Theta$ we have that $\chi \in \Delta$. We have $c_\psi :^i \psi \in \Theta$ so then $\psi \in \Delta$. Since $C(\psi) < C(c_\psi :^e \psi)$ we have, by I.H. that $M, \Delta \vDash \psi$. So then, since $\Delta \in W$ is arbitrary we have that $M, \Theta \vDash c_\psi :^i \psi$.

(←) Suppose for contraposition that $c_\psi :^i \psi \notin \Theta$. Let

$$\begin{aligned}\Delta_0 = &\{\chi \mid c_\chi :^i \chi \in \Theta\}\\ &\cup \{t :^e \xi \mid t :^e \xi \in \Gamma\} \cup \{\neg t :^e \xi \mid \neg t :^e \xi \in \Gamma\}\\ &\cup \{t :^i \iota \mid t :^i \iota \in \Gamma\} \cup \{\neg t :^i \iota \mid \neg t :^i \iota \in \Gamma\}\\ &\cup \{\neg\psi\}\end{aligned}$$

We will first show that Δ_0 is QCB-consistent:
Suppose, for contradiction, that Δ_0 is inconsistent. Then there is:

- $X := \{\chi_1, ..., \chi_n\} \subseteq \{\chi \mid c_\chi :^i \chi \in \Theta\}$
- $\Xi^+ := \{t_1 :^e \xi_1, ..., t_m :^e \xi_m\} \subseteq \{t :^e \xi \mid t :^e \xi \in \Gamma\}$
- $\Xi^- := \{\neg t_1 :^e \xi_1, ..., \neg t_l :^e \xi_l\} \subseteq \{\neg t :^e \xi \mid \neg t :^e \xi \in \Gamma\}$
- $I^+ := \{t_1 :^i \iota_1, ..., t_o :^i \iota_o\} \subseteq \{t :^i \iota \mid t :^i \iota \in \Gamma\}$
- $I^- := \{\neg t_1 :^i \iota_1, ..., \neg t_k :^i \iota_k\} \subseteq \{\neg t :^i \iota \mid \neg t :^i \iota \in \Gamma\}$

s.t. $\vdash (\bigwedge X) \to (\bigwedge \Xi^+) \to (\bigwedge \Xi^-) \to (\bigwedge I^+) \to (\bigwedge I^-) \to \psi$. Then:

$$\begin{aligned}&\vdash (\bigwedge X) \to (\bigwedge \Xi^+) \to (\bigwedge \Xi^-) \to (\bigwedge I^+) \to (\bigwedge I^-) \to \psi\\ &\vdash B^i((\bigwedge X) \to (\bigwedge \Xi^+) \to (\bigwedge \Xi^-) \to (\bigwedge I^+) \to (\bigwedge I^-) \to \psi) && (N)\\ &\vdash B^i(\bigwedge X) \to B^i(\bigwedge \Xi^+) \to B^i(\bigwedge \Xi^-) \to B^i(\bigwedge I^+) \to B^i(\bigwedge I^-) \to B^i\psi && (K)\end{aligned}$$

We have $B^i \bigwedge X \in \Theta$. Moreover, by positive $(t :^e \psi \to B^i(t :^e \psi)$, $t :^i \psi \to B^i(t :^i \psi))$ and negative introspection $(\neg t :^e \psi \to B^i(\neg t :^e \psi)$, $t \neg :^i \psi \to B^i(\neg t :^i \psi))$ we have implicit belief for the other conjunctions as well. So then $c_\psi :^i \psi \in \Theta$ which cannot be the case because of our initial assumption that $c_\psi :^i \psi \notin \Theta$. So Δ_0 is QCB-consistent.

Since Δ_0 is QCB-consistent we can, by lindenbaum lemma, extend Δ_0 to QCB-maximal consistent subset of Σ^0 which we denote as Δ. We have that Δ agrees with Γ on all explicit and implicit justifications. Therefore $\Delta \in W$. Moreover, $\{\chi \mid c_\chi :^i \chi \in \Theta\} \subseteq \Delta$ so, by construction of M we have that $\Theta R \Delta$. Moreover, since $\psi \notin \Delta$ and $C(\psi) < C(\phi)$ we have that, by I.H. $M, \Delta \nvDash \psi$. Then, since $\Theta R \Delta$ we have that $M, \Theta \nvDash c_\psi :^i \psi$.

Then, by contraposition, we have that $M, \Theta \vDash c_\psi :^i \psi$ implies that $c_\psi :^i \psi \in \Theta$.

- Suppose $\phi = t_1 \cdot t_2 :^i \psi$.

$$\begin{aligned}t_1 \cdot t_2 :^i \psi \in \Theta &\text{ iff } t_1 :^i (\chi \to \psi) \wedge t_2 :^i \chi \in \Theta && (\text{ implicit application})\\ &\text{ iff } t_1 :^i (\chi \to \psi) \in \Theta \text{ and } t_2 :^i \chi \in \Theta\\ &\text{ iff } M, \Theta \vDash t_1 :^i \chi \to \psi \text{ and } M, \Theta \vDash t_2 :^i \chi && (\text{I.H.})\\ &\text{ iff } M, \Theta \vDash t_1 \cdot t_2 :^i \psi\end{aligned}$$

- Suppose $t = !t'$. Then $\psi = (t' :^i \chi)$ or $\psi = (t' :^e \chi)$.

– If $\psi = t' :^i \chi$

$$!t' :^i (t' :^i \chi) \in \Theta \text{ iff } t' :^i \chi \in \Theta$$
$$\text{iff } M, \Theta \models t' :^i \chi \quad \text{(I.H.)}$$
$$\text{iff } M, \Theta \models !t' :^i (t' :^i \chi)$$

– If $\psi = t' :^e \chi$

$$t' :^i (t' :^e \chi) \in \Theta \text{ iff } t' :^e \chi \in \Theta$$
$$\text{iff } M, \Theta \models t' :^e \chi \quad \text{(I.H.)}$$
$$\text{iff } M, \Theta \models !t' :^i (t' :^e \chi)$$

3. M is a pseudo model
   - $\Theta R \Delta$ iff $\Theta' R \Delta$
     Suppose $\Theta R \Delta$ and take any $\Theta' \in W$. We have that $\phi \in \Delta$ for all $c_\phi :^i \phi \in \Theta$. Since $\Theta' \in W$ we have that $c_\phi :^i \phi \in \Theta'$ iff $c_\phi :^i \phi \in \Theta$. So then $\phi \in \Delta$ for all $c_\phi :^i \phi \in \Theta'$ meaning that $\Theta' R \Delta$. The other direction goes analogues.
   - $\Theta R \Delta$ implies $M, \Delta \models \Phi_e(\cap Qmax(E))$
     Suppose $\Theta R \Delta$. Take any $t \in \cap Qmax_c(E) = E$ such that $t \gg \phi$ with $\phi \in \mathcal{L}_e$. Then $t :^e \phi \in \Gamma$ so then $t :^e \phi \in \Theta$ then $c_\phi :^e \phi \in \Theta$ and then $c_\phi :^i \phi \in \Theta$ so then $\Delta \models \phi$. Since ϕ is arbitrary we have $M, \Delta \models \Phi_e(\cap Qmax_c(E))$.
4. M is strong.
   Suppose Θ and Θ are p-equivalent. Then proposition 3.2.24 tells us that $M, \Theta \models \phi$ iff $M, \Theta' \models \phi$. By truth lemma we have that $\phi \in \Theta$ iff $\phi \in \Theta'$ i.e. $\Theta = \Theta'$.

So we have that $\phi^* \in \Gamma$ is satisfiable on a finite quasi-consistent strong pseudo model.

■

This gives us completeness w.r.t. finite strong quasi-consistent pseudo models. Now we will show that we can convert such a model into an evidential base model.

Theorem 3. Given $\Sigma \subseteq \mathcal{L}$ we have that, for any finite strong quasi-consistent pseudo model $M = (W, E, R, \|.\|)$ such that $\|.\| : Prop \cap \Sigma \to \mathcal{P}(W)$ and $E \subseteq \{t \mid t :^e \phi \in \mathcal{L}_e \cap \Sigma\}$, there exists an evidential base model $M^\bullet = (W^\bullet, E^\bullet)$ of the same size such that:

$$M, w \models \phi \text{ iff } M^\bullet, w^\bullet \models \phi \text{ for } \phi \in \Sigma^\Phi$$

Proof. Suppose $\Sigma \subseteq \mathcal{L}$ and $M = (W, E, R, \|.\|)$ is a finite strong quasi-consistent pseudo model such that $\|.\| : Prop \cap \Sigma \to \mathcal{P}(W)$ and $E \subseteq \{t \mid t :^e \phi \in \mathcal{L}_e \cap \Sigma\}$.

Take fresh variable $q \in Prop \setminus \Sigma$ and define $M^\bullet = (W^\bullet, E^\bullet)$ as follows

- $W^\bullet = \{w^\bullet \mid w \in W\}$ where:

$$w^\bullet = \{p \in Prop \cap \Sigma \mid w \in \|p\|\} \cup \begin{cases} \{q\} & \text{if there is } v \in W \text{ s.t. } vRw \\ \emptyset & \text{otherwise} \end{cases}$$

- $E^\bullet = E \cup \{c_q\}$

We have the following:

1. $E^\bullet$ is a certified quasi consistent evidence base.
   E is a certified quasi-consistent evidence base and q does not occur in Σ so $c_q \notin \Sigma^t$ so it cannot occur in any derivations in E. So then $E^\bullet$ is quasi-consistent,

explicit, and certified.

2. $M^\bullet$ has the same size as M
   Since M is strong we have that the map $w \mapsto w^\bullet$ is a bijection, thus $M^\bullet$ has the same size as M.

3. $M, w \models \phi$ iff $M^\bullet, w^\bullet \models \phi$ for $\phi \in \Sigma$

   Suppose $\phi \in \Sigma$. Induction on ϕ. *Induction Hypothesis:* for $\psi \in \Sigma$ such that $C(\psi) < C(\phi)$ we have that $M, w \models \psi$ iff $M^\bullet, w^\bullet \models \psi$.

   - $\phi = p$
     $M, w \models p$ iff $w \in \|p\|$ iff $p \in \{q \in Prop \cap \Sigma^\Phi \mid w \in \|q\|\}$ iff $p \in w^\bullet$
   - $\phi = \bot$ is never true on both kinds of models.
   - Boolean cases follow from I.H.
   - $M, w \models t :^e \psi$ iff $t \in \cap Qmax_c(E) = E$ iff $t \in E \cup \{q\}$ iff $M^\bullet, w^\bullet \models t :^e \psi$.

   (*) note that the previous cases already show that $M, w \models \psi$ iff $M^\bullet, w^\bullet \models \psi$ for $\psi \in \Sigma^\Phi \cap \mathcal{L}_e$.

   - First we prove the following:

     (**) wRv iff $M^\bullet, v^\bullet \models \Phi_e(\cap Qmax_c(E^\bullet))$

     (→) Suppose wRv. Then, since $M_\Gamma^{\Sigma^\Phi}$ is a pseudo model we have that $M, v \models \Phi_e(\cap Qmax_c(E))$. By (*) we have that $M^\bullet, w^\bullet \models \Phi_e(\cap Qmax_c(E))$ since $\Phi_e(\cap Qmax_c(E)) \subseteq \mathcal{L}_e \cap \Sigma^\Phi$. Moreover, since there is w s.t. wRv we have that $M^\bullet, v^\bullet \models q$. Since $\Phi_e(\cap Qmax_c(E)) = \Phi_e(E)$ and $\Phi_e(c_q) = q$ we have that $M^\bullet, v^\bullet \models \Phi_e(E \cup \{c_q\})$. Since $\Phi_e(E \cup \{c_q\}) = \Phi_e(\cap Qmax_c(E \cup \{c_q\}))$ we have that $M^*, v^* \models \Phi_e(\cap Qmax_c(E \cup \{c_q\}))$

     (←) Suppose $M^\bullet, v^\bullet \models \Phi_e(\cap Qmax_c(E^\bullet))$ then $M^\bullet, v^\bullet \models \Phi_e(E \cup \{c_q\})$. Since $q \in \Phi_e(E \cup \{c_q)$ we have that $M^\bullet, v^\bullet \models q$. Then by definition of $v^\bullet$ we have that wRv for some $w \in W$. Since M is a pseudo model we then have that wRv for each $w \in W$.

     Then:

$$\begin{aligned} M, w \models c_\psi :^i \psi &\text{ iff } wRv \text{ implies } M, v \models \psi \\ &\text{ iff } wRv \text{ implies } M^*, v^* \models \psi && \text{(I.H.)} \\ &\text{ iff } M^*, v^* \models \Phi_e(\cap Qmax_c(E^*)) \text{ implies } M^*, v^* \models \psi && (**) \\ &\text{ iff } M^*, w^* \models c_\psi :^i \psi \end{aligned}$$

     Composite cases are defined the same so follow directly from inductive argument.

■

Theorem 4. QCB is sound and complete w.r.t. finite evidential models.

Proof. Follows from previous two theorems.

Suppose Γ_0 is consistent. Then we can take some $\phi_0 \in \Gamma_0$ and construct Σ^Φ and Σ^t and we have that $M_\Gamma^{\Sigma^\Phi}, \Gamma \models \phi$ with $\phi \in \Gamma$, Γ an MCS$^{\Sigma^\Phi}$. Then $M^* = (W^*, E^*)$ has $M^*, \Gamma^* \models \phi$ since $\phi \in \Sigma^\Phi$.

So ϕ is satisfiable on a finite evidential base model.

■

Chapter 4

Discussion and future work

If this thesis were a term, this would be the formula it is admissible for.

In this thesis we set out to integrate justification terms as pieces of evidence into a belief base framework to model beliefs of non-omniscient agents. Such sets of justification terms as evidence we call *evidence bases*.
We have taken inspiration from explicit evidence models [Gog16] where a notion of *explicit evidence*, in the form of logical formulas, is used to compute *explicit beliefs* and, in turn, *implicit belief*.
Such models use a notion of quasi-consistency to deal with logical omniscience in a way that an agent may have contradictory beliefs, but the agent is able to fix these beliefs later when they have *become aware of the contradiction* in their belief.

In this formal setting however, we have noted on two examples (examples 1.3.9 and 1.3.10) which motivated the need for a method to encode derivations that we would then be able to keep track of.

The first example showed that, in the formal setting of Gogoladze [Gog16], an agent may become aware of some contradiction, p and $p \to \bot$ for example, but when gaining new pieces of evidence q and $q \to \bot$, the agent automatically is aware of this contradiction as well, even if q and p are completely unrelated.
The reason this occurs is because there is no mechanism which can tell us that the contradiction the agent has become aware of originates from p and $p \to \bot$.

The second example showed that an agent may derive some formula p from formulas $q \to p$ and q, in which case they may believe p, q and $p \to q$. But If the agent were now to obtain some evidence for $q \to \bot$ and become aware of the contradiction $q \to \bot$ and q, they will still have a belief in proposition p, even though they have derived this from q, which is now not trusted anymore.
Again, this phenomenon occurs due to the fact that the formal setting does not have a mechanism which can tell us that our derivation of p and thus our belief in p actually originates from q.

The formal account in this work has shown that the addition of justification terms, which can be seen as encodings of derivations that an agent has constructed, succeeds in adding the expressivity needed to overcome these anomalies. In chapter 2 we have seen that an adapted version of the qmax calculation in [Gog16] to compute explicit beliefs succeeds in allowing agents to become aware of a *certain* contradiction whilst remaining oblivious with respect to others, and thus still having inconsistent beliefs.
Moreover, since these justification terms allow us to access sub-derivations, the addition of justification terms allows us to disregard certain conclusions if they were made using derivations which we have disregarded due to conflict with other pieces of evidence/derivations.

As a result, we have defined a mechanism that derives explicit belief from explicit evidence in a more fine-grained way, allowing for more expressivity. Moreover, due to the nature of justification terms, we can pinpoint exactly where our explicit beliefs originate from.

Apart from explicit beliefs, we have defined three more notions of belief. The first being *implicit belief*, which corresponds to the formulas which logically follow from our explicit beliefs. In practical terms these can be derived by taking a closure of our justifications (the set of terms occuring in all our qmax sets), i.e. composing all terms possible from our set of justifications and some set of axiom terms.
Here we noted that, when we consider only our set of justifications, we may be disregarding some pieces of evidence which may be distributed among the qmax sets, but which all entail some common conclusion that the agent may not yet be aware of.
As opposed to only considering the set of justification we may then consider looking at all qmax sets and closing these under composition, resulting in our notion of *evidence based implicit belief*. Since justifications occur in all qmax sets of a set of evidence, we have as a result that our implicit beliefs are contained in our evidence based implicit beliefs, and thus we also have that an agent their explicit beliefs are contained in their evidence based implicit beliefs.

Since our explicit beliefs are allowed to be inconsistent, which is the case if we have some set of justifications which are mutually inconsistent, we have as a result that these two notions of implicit belief are allowed to be inconsistent as well. This would allow an agent to implicitly believe a contradiction and thus have an (evidence based) implicit belief in everything.
However, the way explicit beliefs are derived does not allow for an explicit belief in a contradiction $\perp$, since any evidence admissible for $\perp$ can not occur in any qmax set. This means we may have some (evidence based) implicit belief that no rational agent will ever come to explicitly believe, thus steering away from the interpretation of implicit beliefs as potential beliefs.
We therefore considered another notion of belief which we call *potential explicit belief*. This we derive by first closing our set of evidence under composition, after which we consider its qmax sets. Intuitively, this notion of belief can be seen as the explicit beliefs of an omniscient agent, who has unlimited time and resources to derive their explicit beliefs.

We further have shown the relation between evidence based implicit belief and potential explicit belief. That being that any potential explicit belief is also an evidence based implicit belief, but the converse need not be true. The converse is however true when an agent is aware of all contradictions in their evidence.
Furthermore, we have shown that, whenever we compute the qmax sets of an evidence base E and then close these qmax sets under composition after which we compute the qmax sets of these closures, they actually correspond to the qmax sets of the closure of E. This intuitively tells us that we can keep constructing new derivations from the evidence we have and update our explicit beliefs with qmax computation to eventually (given unlimited time and resources) arrive at our set of potential explicit beliefs. Thus agreeing with the idea that a potential explicit belief can be seen as a form of belief revision to the limit.
The relation between implicit belief and potential explicit belief we aim to explore further in the future.

Apart from defined notions of belief and added expressivity by considering derivations in the form of justification terms, a novel contribution in this thesis is the introduction of *weakening terms*. These can intuitively be thought of as arguments which tell us that, whichever of the possible sub-derivations (which may be mutually inconsistent) may be true, they all tell us that at least ϕ is the case. When relating justification terms as pieces of evidence to evidence models where evidence comes in the form of sets of worlds, such a weakening term appears to correspond to the union of pieces of (combined) evidence.

In chapter 3 we have considered a formal logic meant to express the mechanisms we have described in the previous chapter. Here we have noted issues when considering explicit and implicit belief, namely that these notions of belief involve some quantification over justification terms, i.e. we have an explicit/implicit belief in ϕ iff we have *some* explicit/implicit justification for ϕ.

In section 3.2, we consider a solution to this problem, namely *certification*. Here we consider *certified evidence bases* which have the extra condition that, whenever we have some argument for ϕ in our evidence base, we require an atomic term (certificate) for ϕ in our evidence base as well. We further defined *certified qmax sets* which have the same property, i.e. any certified qmax set e of certified evidence base E has that $c_\phi \in e$ whenever there is some argument for ϕ contained in e. In this case, we can always refer to our atomic term c_ϕ whenever we have some term and/or justification for ϕ.
We then considered the logic of quasi-consistent belief (QCB) which involves explicit and implicit justification and belief, where explicit and implicit belief are defined as abbreviations $c_\phi :^e \phi$ and $c_\phi :^i \phi$ respectively. Intuitively these certificates can now be seen as *canonical evidence* for ϕ and, moreover, take the role of weakening terms as we can interpret a certificate for ϕ as the weakening of all our arguments for ϕ when we are considering certified evidence bases.
We have proven completeness of QCB w.r.t. finite evidential base models, which consist of a set of possible valuations and a certified evidence base, taking into account background beliefs.

Such certified evidence bases, together with certified qmax computation have proved useful when considering explicit and implicit belief due to the availability of atomic terms. However, we argue that this property of *certification* has its downsides too.

Consider for instance the following example:

Example 4.0.1. Suppose $E = \{c_{p\to q}, c_{r\to v}, c_p, c_r, c_{v\to q\to\bot}\}$
We have some evidence for p and r but also that p entails q, r entails v, and v and q are mutually inconsistent. So we have inconsistent evidence of which we are not yet aware of. Specifically, we have evidence that tells us that both v and q are the case but these are mutually inconsistent as well.
Suppose we want to model our agent becoming aware of the contradiction. The agent adds $c_{p\to q} \cdot c_p, c_{r\to v} \cdot c_r, c_{q\to v\to\bot} \cdot (c_{p\to q} \cdot c_p)$ and $c_{q\to v\to\bot} \cdot (c_{p\to q} \cdot c_p) \cdot (c_{r\to v} \cdot c_r)$. In which case the agent would be aware of the contradiction present and should lose their belief in q and v.
However, due to certification, we would need to have that c_q, c_v, $c_{v\to\bot}$ and $c_\bot$ are present in the updated evidence base as well. In this case we have:

$$\begin{aligned} E' = \{&c_{p\to q}, c_{r\to v}, c_p, c_r, c_{q\to v\to\bot}, \\ &c_{p\to q} \cdot c_p, c_{r\to v} \cdot c_r, c_q, c_v, \\ &c_{q\to v\to\bot} \cdot (c_{p\to q} \cdot c_p), c_{v\to\bot}, \\ &c_{q\to v\to\bot} \cdot (c_{p\to q} \cdot c_p) \cdot (c_{r\to v} \cdot c_r), c_\bot\} \end{aligned}$$

If we consider the qmax sets of E' we see that, since c_q and c_v and $c_{v\to\bot}$ are not involved in any derivation, we have these terms occuring in all of our qmax sets. Thus we have an explicit belief in q, v and $v \to \bot$.

In the above example we see that, with the addition of certification, an agent may combine their evidence to become aware of a certain contradiction in their evidence, but does not lose belief in the contradiction. This example is somewhat analogues to example 1.3.9 which was one of the motivating examples for introducing justification terms. In this case however, we have become aware of a certain contradiction, but we do not lose contradictory beliefs in the contradiction that we have become aware of. Which is less desirable behavior.

This however can be solved with mechanisms we have already defined before. Consider again well organized evidence bases (definition 2.2.6). As noted before, an agent with a well organized evidence base will have that a derivation of a contradiction made with *some term* for ϕ is equivalent to having derived the contradiction with all pieces of evidence for

ϕ when considering explicit beliefs.

The following example however is more troublesome:

Example 4.0.2. Suppose $E = \{c_p, c_{p\rightarrow q}, c_{p\rightarrow q} \cdot c_p\}$.
We have some evidence for p and some evidence for $p \rightarrow q$, which we have combined. However, E is not certified since $c_q \notin E$. So if we certify E we have:

$$E' = \{c_p, c_{p\rightarrow q}, c_{p\rightarrow q} \cdot c_p, c_q\}$$

In which case we have explicit beliefs in $p, p \rightarrow q, q$.
But now consider the following: we obtain evidence against p ($c_{p\rightarrow\bot}$) which we combine with c_p.

$$E'' = \{c_p, c_{p\rightarrow q}, c_{p\rightarrow q} \cdot c_p, c_q, c_{p\rightarrow\bot}, c_{p\rightarrow\bot} \cdot c_p, c_\bot\}$$

If we now take the qmax sets of E'' we see that c_q is contained in all of them. And thus we have an explicit belief in q.

This example is analogues to example 1.3.10, which is the other motivating example to introduce justification terms. However, we see here that the same anomaly occurs. We gain some belief in q due to deriving it from p and $p \rightarrow q$, but losing our trust in c_p does not result in us losing belief in q.

The reason for this is that we have added c_q due to $c_{p\rightarrow q} \cdot c_p$, but we have no way to 'track' where c_q has come from. In a way certification breaks a derivational link that is inherent to justification terms. This may be a reason we would not want to consider certified evidence bases. Alternatives to certified evidence bases in order to formally address explicit and implicit belief we leave as future work.

The formalization of the other notions of belief discussed in chapter 2, those being evidence based implicit belief and potential explicit belief, we also leave for future work. But we briefly suggest a step towards this formalization in the following paragraph.
In the language we have considered in chapter 3, we have taken explicit and implicit justifications as primitive. As a result, our logic can only express quasi-consistent (certified) evidence bases since we can not have a formula of the form $t :^e \bot$ that is consistent.
Because of this we do not have to account for the qmax computation in the logic as the only qmax set of a quasi-consistent evidence base E is E itself. This is not an issue when considering explicit and implicit beliefs since these are only concerned with explicit and implicit justifications.
However, when moving towards the other forms of belief (in a non-trivial way) this becomes a problem as we will want consider individual qmax sets which may contain arguments (terms) that are not justifications (trusted terms). This would require us to look at evidence bases which contain some derivation of $\bot$.
As opposed to using justifications (explicit and implicit) as primitive we may instead consider pieces of evidence (terms) in our evidence base as primitive.
For instance, we may use a formula Et denoting that t is contained in our (current) evidence base. Moreover, if we would consider an alternative to certified evidence bases and qmax, we may want to introduce formula $E\phi$ which denotes that we have *some* term for ϕ in our (current) evidence base.
Now we will want to define justifications using Et. We can then introduce operator $[\mathbb{Q}]\phi$ which, for evidence base E, intuitively states that in any $e \in Qmax(E)$ we have that ϕ is the case. Then we can define $t :^e \phi$ as $[\mathbb{Q}]Et$ and $B^e\phi$ as $[\mathbb{Q}]E\phi$.
Turning to evidence based implicit belief and potential explicit belief, we can introduce operator $[Cl]\phi$ which intuitively denotes that, in the full closure of current evidence base, we have that ϕ is the case.
With this we can define evidence based implicit belief in ϕ as $[\mathbb{Q}][Cl]E\phi$ i.e. we first consider the qmax sets of our evidence base and if all their closures contain some term for ϕ we have an evidence based implicit belief in ϕ.
Moreover, if we flip these operators around we actually arrive at our definition of potential

explicit belief in ϕ which is $[Cl][\mathbb{Q}]E\phi$, i.e. we first close our evidence base and if each qmax set of the closure contains some argument for ϕ then we have a potential explicit belief in ϕ.
The question whether this is a fruitful avenue for formally describing these two notions of belief we leave for future work.

Bibliography

[AF19] Sergei Artemov and Melvin Fitting. *Justification logic: Reasoning with reasons*, volume 216. Cambridge University Press, 2019.

[AFS24] Sergei Artemov, Melvin Fitting, and Thomas Studer. Justification Logic. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2024 edition, 2024.

[Art95] Sergei N Artemov. *Operational modal logic*, volume 95. Mathematical Sciences Institute, Cornell University, 1995.

[Art01] Sergei N Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic logic*, 7(1):1–36, 2001.

[Art08] Sergei Artemov. The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513, 2008.

[BBG19] Alexandru Baltag, Nick Bezhanishvili, and Saúl Fernández González. Multi-agent topological evidence logics. *SYSMICS2019*, page 34, 2019.

[BBÖS16] Alexandru Baltag, Nick Bezhanishvili, Aybüke Özgün, and Sonja Smets. Justified belief and the topology of evidence. In *Logic, Language, Information, and Computation: 23rd International Workshop, WoLLIC 2016, Puebla, Mexico, August 16-19th, 2016. Proceedings 23*, pages 83–103. Springer, 2016.

[BDRV01] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic: graph. Darst*, volume 53. Cambridge University Press, 2001.

[BMS16] Alexandru Baltag, Lawrence S Moss, and Sławomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Readings in Formal Epistemology: Sourcebook*, pages 773–812. Springer, 2016.

[BN23] Francesco Berto and Daniel Nolan. Hyperintensionality. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2023 edition, 2023.

[BR16] Alexandru Baltag and Bryan Renne. Dynamic Epistemic Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition, 2016.

[BRS14] Alexandru Baltag, Bryan Renne, and Sonja Smets. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81, 2014.

[BS08] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the foundations of game and decision theory (LOFT 7)*, 3:9–58, 2008.

[Cre75] Max J Cresswell. Hyperintensional logic. *Studia Logica: An International Journal for Symbolic Logic*, 34(1):25–38, 1975.

[ELML+25] Thorsten Engesser, Thibaut Le Marre, Emiliano Lorini, François Schwarzentruber, and Bruno Zanuttini. A simple integration of epistemic logic and reinforcement learning. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*, pages 686–694, 2025.

[FH87] Ronald Fagin and Joseph Y Halpern. Belief, awareness, and limited reasoning. *Artificial intelligence*, 34(1):39–76, 1987.

[FHMV04] Ronald Fagin, Joseph Y Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT press, 2004.

[Fit05] Melvin Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132(1):1–25, 2005.

[Get63] Edmund L Gettier. Is justified true belief knowledge? *Analysis*, 23(6):121–123, 1963.

[GG06] Paul Gochet and Pascal Gribomont. Epistemic logic. In *Handbook of the History of Logic*, volume 7, pages 99–195. Elsevier, 2006.

[Göd33] Kurt Gödel. Eine interpretation des intuitionistischen aussagenkalkuls. *Ergebnisse eines mathematisches Kolloquiums*, 4:39–40, 1933.

[Göd38] K Gödel. Vortrag bei zilsel. lecture, published, with an english translation. *Gödel (1995, 86–113)*, 1938.

[Gog16] Kristina Gogoladze. Evidence-based belief revision for non-omniscient agents. 2016.

[Het16] Stephen Hetherington. *Knowledge and the Gettier problem*. Cambridge University Press, 2016.

[Hey31] Arend Heyting. Die intuitionistische grundlegung der mathematik. *Erkenntnis*, 2:106–115, 1931.

[Hin62] Kaarlo Jaakko Juhani Hintikka. Knowledge and belief: An introduction to the logic of the two notions. 1962.

[Hin79] Jaakko Hintikka. Impossible possible worlds vindicated. In *Game-theoretical semantics: Essays on semantics by Hintikka, Carlson, Peacocke, Rantala, and Saarinen*, pages 367–379. Springer, 1979.

[HM90] Joseph Y Halpern and Yoram Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM (JACM)*, 37(3):549–587, 1990.

[Kol32] Andrej Kolmogoroff. Zur deutung der intuitionistischen logik. *Mathematische Zeitschrift*, 35(1):58–65, 1932.

[Leh84] Daniel Lehmann. Knowledge, common knowledge and related puzzles (extended summary). In *Proceedings of the third annual ACM symposium on Principles of distributed computing*, pages 62–67, 1984.

[Lev84] Hector J Levesque. A logic of implicit and explicit belief. In *AAAI*, volume 84, pages 198–202, 1984.

[Lor20] Emiliano Lorini. Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282:103233, 2020.

[LR22] Emiliano Lorini and Éloan Rapion. Logical theories of collective attitudes and the belief base perspective. In *21st International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2022)*, pages 833–841. ACM, 2022.

[LS21] Emiliano Lorini and Francois Schwarzentruber. Multi-agent belief base revision. In *Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021)*, number article 270, pages 1959–1965. International Joint Conferences on Artificial Intelligence Organization, 2021.

[MT48] John CC McKinsey and Alfred Tarski. Some theorems about the sentential calculi of lewis and heyting. *The journal of symbolic logic*, 13(1):1–15, 1948.

[Nel00] Dana K Nelkin. The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109(3):373–409, 2000.

[ON04] Mauricio Osorio and Juan Antonio Navarro. Answer set programming and s4. In *Advances in Artificial Intelligence–IBERAMIA 2004: 9th Ibero-American Conference on AI, Puebla, Mexico, November 22-26, 2004. Proceedings 9*, pages 353–363. Springer, 2004.

[OS24] Ivan Orlov and Werner Stelzner. The calculus of the compatibility of propositions. *The Australasian Journal of Logic*, 21(1):20–46, 2024.

[Pac06] Eric Pacuit. A note on some explicit modal logics. *Proceedings of the Fifth Panhellenic Logic Symposium, pp. 117 - 125*, 2006.

[RSW25] Rasmus Rendsvig, John Symons, and Yanjing Wang. Epistemic Logic. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2025 edition, 2025.

[Rub06] Natalia Rubtsova. On realization of s5-modality by evidence terms. *Journal of Logic and Computation*, 16(5):671–684, 2006.

[Sol17] Anthia Solaki. *Steps out of logical omniscience*. PhD thesis, PhD thesis, ILLC, University of Amsterdam, 2017.

[Sta91] Robert Stalnaker. The problem of logical omniscience, i. *Synthese*, pages 425–440, 1991.

[TVD14] Anne Sjerp Troelstra and Dirk Van Dalen. *Constructivism in Mathematics, Vol 2*. Elsevier, 2014.

[Var86] Moshe Y Vardi. On epistemic logic and logical omniscience. In *Theoretical aspects of reasoning about knowledge*, pages 293–305. Elsevier, 1986.

[vB07] Johan van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.

[vB19] Johan van Benthem. Implicit and explicit stances in logic. *Journal of Philosophical Logic*, 48(3):571–601, 2019.

[vBP11] Johan van Benthem and Eric Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99:61–92, 2011.