Commitments, beliefs and expectations in conversation

Understanding the effect of speech acts in logic

MSc Thesis *(Afstudeerscriptie)*

written by

Thomas van der Leer
(born 16 January 2001 in Utrecht)

under the supervision of **Dr. Maria Aloni** and **Prof. Dr. Corien Bary**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract. This thesis departs from the idea that, by performing speech acts, we undertake commitments to act in accordance with their content (Brandom, 1983; Geurts, 2019; Hamblin, 1971; Tuzet, 2006). Three effects that speech acts have on the discourse and its participants will be studied through this lens. First, the commitments speech acts invoke manipulate interlocutors' normative attitudes, since commitments are social obligations. Second, they affect their beliefs, since after the performance of a speech act discourse participants believe that certain commitments are made and may also come to believe in the content of these commitments. Lastly, the commitments introduced by a speech act shape our expectations regarding the discourse development, given that we prefer to be cooperative by making the same commitments and do not expect someone to violate their commitments by making inconsistent claims. Whereas existing commitment-based theories typically seem to be concerned with either the relation between commitment and belief or between commitment and expectation, the goal of this thesis is to unite all three effects of speech acts in one dynamic logical system. In the Speech Act Logic (**SAL**) that this thesis introduces, both commitment and belief are defined as modal operators that are updated on the basis of speech act performance. Moreover, by following Commitment Space Semantics (CSS; Cohen & Krifka, 2014; Krifka, 2015) in representing the discourse as a commitment space, **SAL** captures not only the speech act's effect on the current state of the discourse but on its projected future states as well. However, whereas CSS models only one type of discourse expectation (similarly to Farkas and Bruce (2010)), **SAL** is able to represent different layers of expectations, by formally distinguishing between commitment violation and uncooperative behaviour. Although the thesis mainly focuses on discussing and formalising assertions, **SAL** is applied to polar questions and speech act embeddings as well and could be further extended to a model of yet other speech act types. To motivate the formal work, the thesis will along the way offer a state-of-the-art theoretical overview of the commitment-based approach to speech act theory.

Contents

Chapter 1

Introduction

1.1 Why communicate?

Speech acts, such as assertions, questions, promises, commands, greetings and many others, are the building blocks of communication. In any conversation, each speech act builds on the conversational work that has previously been done and in turn influences what speech acts will follow. Speech acts not only transform our perception of the conversation but also the context in which the conversation takes place and the information discourse participants have at their disposal. Understanding their profuse function means understanding why we communicate: what is it that we achieve with performing speech acts, and what is it about this achievement that makes it worth pursuing?

Traditionally, it has been thought that the function of communication is to share and recognise each other's mental attitudes, such as intentions, beliefs and desires. Speech acts aid this goal by enabling discourse participants to express these attitudes (Bach & Harnish, 1979; Grice, 1957; Owens, 2006; Strawson, 1964; Williams, 2002).[1] This view is known as 'expressivism' (Lauer, 2013), 'inferentialism' (Elder, 2021), 'mentalism' (Asher & Lascarides, 2008a; Geurts, 2019) or, as I will call it, 'intentionalism' (Harris, 2019).

Arguably the most discussed type of speech act in the philosophical debate on speech acts (and hence in the intentionalist tradition) is assertion. Intentionalism - as I understand and use the term here - defines assertion simply as a recognised expression of a belief. The following definition, originating from Bach and Harnish (1979), is often taken as representative for the intentionalist's view on assertion:

> In uttering [an expression] e, [the speaker] S asserts that P if S expresses:
>
> (i) the belief that P, and
>
> (ii) the intention that [the hearer] H believe that P.
>
> (Bach & Harnish, 1979, p. 42)

Thus, by asserting a proposition, the speaker conveys that she believes this proposition and that she intends for the addressee to have this belief as well. It is hence the speaker's hope that the addressee, in turn, recognises this intention and comes to share the belief expressed (cf. Grice, 1957).

Of course, the belief that the speaker expresses with her assertion need not be genuine. For instance, she may lie. This possibility is acknowledged by the intentionalist, who could argue that it is still the

---

[1] To be precise, Strawson (1964) and Bach and Harnish (1979) do not uphold that all speech acts express mental attitudes, but rather only those that are 'communicative', such as assertions, promises and orders.

speaker's intention to make the addressee believe in her lie (this seems to be the point of lying in the first place). Since a lie is an expression of a belief with the intention to instill this belief in the addressee, it fits the intentionalist's definition of assertion. Although lying is thus accounted for in the intentionalist framework, other cases of insincerity provide more of a challenge. More specifically, intentionalism starts to struggle in situations where both the speaker and addressee are aware of the fact that the speaker is insincere (MacFarlane, 2011). Such contexts of open insincerity can be found, for instance, in fiction or pretend play. Geurts (2017) gives the example of a school practicing for an emergency by carrying out an evacuation plan. In such a situation, the teacher could assert the following to her students:

(1)   This is an emergency.

It is implausible to think that either the teacher or her students will actually believe (1). Moreover, we cannot conclude that the teacher *intends* for her students to believe (1). Thus, although arguably (1) constitutes an assertion, it does not fit Bach and Harnish's (1979) definition.

For a similar example, consider an actor playing the part of Claudius in Hamlet - this actor could unproblematically assert (2) to his audience:

(2)   I am the king of Denmark.

Again, we should not assume that either the actor or the audience actually believes (2), nor that the actor intends his audience to have this belief. Similarly to (1), the second clause of Bach and Harnish's (1979) definition is not met, leading to the disputable conclusion that (2) is not an assertion.

If one finds these examples of open insincerity not convincing enough because they do not form part of the kind of everyday conversations we naturally think of when discussing the function of assertion, then consider the following more reality-based scenario: suppose that you are walking through a market place in Marrakech and your eye catches a nice souvenir. You ask the marketeer what the price is, after which he asserts:

(3)   This souvenir costs 100 dirham.

Being aware of haggling practices, you both know that the marketeer does not truly believe that the souvenir costs this much. Him knowing that you know this, he surely did not intend for you to believe this either. Rather, his assertion merely functioned as an opening offer, with the expectation that you eventually settle for a price the marketeer is happy with. Again, we here find a linguistic expression that seems to be an assertion, yet does not conform to Bach and Harnish's (1979) definition.

Even worse for the intentionalist, the counterexamples do not seem to be limited to contexts of open insincerity only. In more default linguistic situations as well, assertions often seem to do something else than (just) express beliefs. Krifka (2024b), for instance, gives the following pair of examples, arguing that intuition tells us that (4a) is a stronger statement than (4b):

(4)   a.   The buffet is open.
       b.   I believe that the buffet is open.

<div align="right">(Krifka, 2024b, p. 52)</div>

If we follow the intentionalist, however, and take assertions to equal expressions of beliefs, then we would have to say that the meaning of the two sentences is identical (assuming that a belief in a belief in a proposition p equals a belief in p), which is clearly not the right conclusion. Similarly, the following expression seems more than natural:

(5)   I won't say that the elections were rigged. But I still believe it.

If assertions were just expressions of belief, then (5) would constitute a contradiction, while in actual fact the utterance seems to be perfectly acceptable.

Krifka (2014, 2015) moreover points out that, contrary to what Bach and Harnish's (1979) definition tells us, assertions do not always presuppose the speaker's intention to have the addressee believe in their content, even in sincere contexts. In fact, it seems even possible for the speaker to explicitly announce that she does not have this intention, as shown in (6).

(6)   Believe it or not, I never cheated on you.

<div align="right">(Krifka, 2014, p. 65)</div>

All of the above examples involve assertions that do not seem to be mere expressions of belief. Mac-Farlane (2011) therefore concludes that intentionalism *undergenerates*: it does not correctly identify an utterance as an assertion in all cases, or it wrongfully predicts two distinct assertions to be identical. He moreover argues that intentionalism *overgenerates*. That is, intentionalism sometimes wrongly predicts an expression to be an assertion. Specifically, MacFarlane notices this happening in situations where a belief is expressed implicitly, such as through a conversational implicature. Consider, for instance, a professor writing a recommendation letter for one of her students (based on Grice (1975, p. 52) and MacFarlane (2011, p. 3)):

(7)   The student has nice handwriting.

It seems obvious that the professor expresses the belief that she does not find the student a suitable candidate. Moreover, it is quite plausible that she intended the receiver of her letter to obtain the same belief. The two clauses of Bach and Harnish's (1979) definition are thus met and so the authors would have to conclude that the professor *asserted* that she found the student unsuitable. However, the way the professor conveyed her belief was clearly different from asserting it. Intuitively, an assertion is an explicit, non-cancelable statement. Conversational implicatures such as the one discussed above, however, are implicit and cancelable. The intuition that the professor did not assert her belief that the student is unsuitable is further evidenced by the fact that she quite naturally can deny she ever intended to convey this belief (Camp, 2018), as the following conversation shows (based on MacFarlane, 2011, p. 3):

(8)   A: The student has nice handwriting.
       B: So you're saying that the candidate is unsuitable?
       A: No no, I was careful not to say that.

Intentionalism is thus standing on loose ground: it not only fails to correctly identify certain linguistic

expressions as assertions, it also incorrectly labels implicit expressions of belief as assertions. What the above counterexamples seem to suggest is that there is more to communication than the reciprocal expression and recognition of beliefs. In the next section, I introduce an alternative perspective on communication and the role of speech acts in it, which will form the foundation of the current thesis.

1.2 Commitments

A view that has gained traction in recent years in the philosophy of language is that speech acts do not necessarily express beliefs but first and foremost induce normative *commitments* (Brandom, 1983; Geurts, 2019; Hamblin, 1971; Krifka, 2015; MacFarlane, 2011; Shapiro, 2020, although this conception already finds its roots in the work of Peirce; see Tuzet, 2006). According to this view, performing a speech act comes with normative consequences. If I promise to be home by noon, for instance, then from that point on I have a social obligation to being home by noon. In other words, we could say that I have undertaken a *commitment* to being home by noon. Others can hold me responsible for keeping my promise; if I do not keep my promise, i.e. if I violate my commitment, then I can be held liable, and social penalties may follow (Bary, 2025; Hamblin, 1970; Kibble, 2006).

Since commitments are normative, they are very different from beliefs and other mental attitudes: you do not have an obligation to hold consistent beliefs, yet you are responsible to make coherent commitments. Moreover, whereas your mental attitudes are private and do not have to be shared with others, commitments are always public (Gunlogson, 2001; Lascarides & Asher, 2009; Sacks, 1992). Everyone is thus aware of the commitments you have made. In fact, if someone does not accept your commitment, then arguably it cannot be called a commitment anymore (Geurts, 2019; Kibble, 2006) - this is a point I will return to in Chapter 2.

Moreover, commitments always relate to actions: a commitment is a responsibility to act in a certain way. Consequentially, commitments restrict the courses of action that are normatively acceptable to take up in the future: if I promise to go home, then I cannot proceed (in a normatively acceptable way) by going to the supermarket. Making a commitment can thus be very useful to others: since the committer can be expected to stick to a specific course of action in the future, others can plan their own actions accordingly. Resultingly, commitments are an important ingredient for the coordination of future actions (Michael, 2021). The fact that speech acts invoke commitments could even be seen as the very reason for communicating in the first place. Communication gives us a way to plan for the future, making it easy to collaborate and eventually achieve collective goals (Geurts, 2019).

Different speech acts come with different kinds of commitments. A promise, such as to be home by noon, induces a commitment for the speaker to fulfill the promise, whereas an order can be seen as entailing a commitment for the addressee to carry out that order. We can thus already note an important distinction between speech acts that bring about commitments for the speaker and those primarily involving commitments for the addressee. Moreover, Geurts (2019) argues that the commitments produced by promises and orders can both be categorised as 'telic', meaning that these commitments have a clear endpoint after which they are revoked. For instance, my commitment to be home by noon persists up to the point that it is noon. In the afternoon, I am released from my commitment, since the responsibility it related to now lies in the past.

Contrasting telic commitments are what Geurts (2019) calls 'atelic' commitments. These commitments, as their name suggests, do not naturally end but rather persist indefinitely by default. Commitments resulting from assertions belong to this class. After making an assertion, the speaker is often said to undertake a commitment to the assertion's propositional content (Brandom, 1983; Hamblin, 1971; MacFarlane, 2005; Searle, 1979; Tuzet, 2006; Watson, 2004). This means that an assertion commits the

speaker to act in a way that is consistent with the truth of that propositional content. The speaker is thus held liable if it turns out that the assertion is false and should be ready to defend its truth when challenged to do so by others (Alston, 2000; Brandom, 1983; MacFarlane, 2011; Matheson et al., 2000; Traum & Allen, 1994; Tuzet, 2006). Propositions and their truth values do not cease to exist and so neither do the commitments that are made in their regard.[2] Given their content, atelic commitments are sometimes also called *propositional* commitments (e.g. Kibble, 2006; Yamada, 2012).

I should note that, although propositional commitments are usually understood as obligations to act in a way that is consistent with the *truth* of a proposition, some scholars such as Lauer (2013) and Condoravdi and Lauer (2011) argue that propositional commitments rather relate to a *belief* in a proposition (cf. Gunlogson, 2001; Lyons, 1995). Under this interpretation, after making an assertion with content p, the speaker is committed to act as if she believes that p. However, I am not convinced that such an approach can explain equally well why assertions of belief such as (4b) seems argumentatively weaker than a plain assertion such as (4a), repeated below:

(4) a. The buffet is open.
     b. I believe that the buffet is open.

<div align="right">(Krifka, 2024b, p. 52)</div>

Again assuming that a belief in a belief in a proposition p boils down to just a belief in p, on Lauer's (2013) conception (4a) and (4b) commit the speaker to the same belief (viz. that the buffet is open) and hence have the same normative effect. We would thus have to come up with some other reason for why the two utterances nevertheless differ from each other. However, if we follow the popular approach in thinking of assertions as resulting in commitments to the *truth* of propositions, then the two assertions above clearly do have different normative effects: whereas (4a) marks a commitment to the truth of the proposition that the buffet is open, (4b) entails a commitment to the belief in that proposition. It seems not too far-fetched to think that having to act in a way that is consistent with the truth of a certain proposition p weighs more heavily than merely having to act in a way that is consistent with believing that p, since a belief in p is easier to defend than the truth of p. The view that propositional commitments relate to truth thus gives us a straightforward explanation for why (4a) is an intuitively stronger statement than (4b) (Krifka, 2024b).[3]

On the other hand, regarding the content of commitment in terms of belief does come with the advantage of easily accounting for the infelicity of Moore sentences, such as (9) below (Lyons, 1995). The apparent contradiction in (9) does not immediately follow, however, when we interpret assertions as committing their author to the *truth* of rather than a belief in a proposition. This is a point I will return to in Chapter 2, where I show that even on a truth-based account of commitments the infelicity of (9) can be simulated.

(9) # It is raining, but I don't believe it is raining.

Disregarding whether to interpret propositional commitments in terms of truth or belief, note that in either case the normativist does not require for an assertion to satisfy either of the two clauses of Bach and

---

[2] Even commitments resulting from statements referencing a specific time, such as "It will rain tomorrow", persist indefinitely: after tomorrow, the speaker will still be committed to the proposition that it rained on the day "tomorrow" referred to.

[3] One could argue that thinking of propositional commitments in terms of truth has to deal with a similar problem, by suggesting that "The buffet is open" is different from "I commit myself to (the truth of) the proposition that the buffet is open". However, such expressions of commitment arguably form a problem for any theory of commitment assuming that committing to committing to X equals committing to X, irrespective of whether X represents a belief in or the truth of some proposition. Moreover, I take it that assertions of commitment are much less natural sounding and therefore in lesser need of a linguistic analysis than assertions of belief such as (4b).

Harnish's (1979) definition. It is possible to commit to a proposition without believing in that proposition or intending to have this proposition believed by the addressee (Lyons, 1977). In fact, Geurts (2019) argues that it is sometimes even rational to commit to a proposition both you and the addressee take unlikely to be true. For instance, a commitment to the fact that someone else will carry out some action in the future which they were initially not intending to carry out may persuade them to nevertheless take up this action. This relaxation of what it means to make an assertion comes with important advantages when analysing the natural language examples intentionalism seems to struggle with.

First, we discussed cases of assertions that do not appear to express just a belief or, alternatively, lack the speaker's intention to have her belief recognised by the addressee. Given the existence of such assertions, we concluded that intentionalism undergenerates: it wrongfully predicts that these linguistic expressions are not assertions (MacFarlane, 2011). For instance, making an assertion during a play on behalf of a fictional character, such as (2) repeated below, of course does not reflect any belief or intended belief recognition, forcing the intentionalist to wrongfully conclude that it does not constitute an assertion.

(2)   I am the king of Denmark.

The normativist *can* show how (2) is an assertion by analysing it as invoking a commitment for the actor to the assertion's content. After uttering (2), the actor is now committed to acting in accordance with the fact that he is king of Denmark - although of course in cases of pretend play we should say that commitments introduced during the play are dissolved after the play has ended, in contrast to the persistent nature of other propositional commitments (Geurts, 2017).

Similarly, consider again the haggling example given in (3):

(3)   This souvenir costs 100 dirham.

According to the normativist, the seller of the souvenir should be seen as undertaking a commitment to justify why he initially asked 100 dirham for it and to act as if the souvenir truly is worth this much - both of which he seems perfectly fit to do. Unlike the intentionalist, the normativist does not presuppose that the seller actually intends the addressee to believe his claim. Moreover, since the normativist would argue that we primarily make assertions to take on responsibilities with which we coordinate our future actions rather than to convey beliefs (Geurts, 2019), it is also not odd to assert a sentence such as (6), repeated below:

(6)   Believe it or not, I never cheated on you.

<div style="text-align: right">(Krifka, 2014, p. 65)</div>

Next, we discussed how intentionalism sometimes overgenerates as well, since it incorrectly defines expressions of implicit belief as assertions. Going back to example (7), we concluded that intentionalism incorrectly predicts that the professor *asserts* that she finds the candidate unsuitable.

(7)   The student has nice handwriting.

The same conclusion is not drawn in normativism. Insinuations such as (7) seem to offer a way of packaging information while avoiding commitment to it (Bary, 2025; Mazzarella et al., 2018): although the professor may well express her belief that the student is unsuitable, she is careful not to take any responsibility for this belief. By keeping meaning implicit, we can thus express our beliefs without running the risk of social penalties if others disagree with us (Camp, 2018; Mazzarella et al., 2018). Since no

commitment is made to implicitly conveyed information, the normativist would rightfully conclude that this information was not asserted.[4]

Normativism thus strongly distantiates itself from intentionalism. Whereas intentionalism pays (too much) attention to the role speech acts play in the recognition of private mental attitudes, normativism shifts the focus to an analysis of their overt social effects. As a result, normativism seems to provide a better explanation of cases in which assertions do not reflect any belief, as well as in cases in which beliefs are successfully conveyed without the use of an assertion.

1.3 The current thesis

The commitment-based normative approach thus offers a new, cognition-free perspective on communication with which it arguably explains certain linguistic data better than intentionalism. Rather than analysing discourse purely in terms of the black-box formed by discourse participants' cognitive attitudes, normativism emphasises the publicly accessible changes ensuing in the social context of the discourse. That is not to say, however, that a commitment-based perspective limits us to thinking of speech acts as only affecting the *normative* attitudes of discourse participants. Rather, starting from the assumption that speech acts induce commitments, we can through this lens investigate other effects of speech acts as well.

The basic idea of the current thesis is to develop a commitment-based analysis of two of these effects, in addition to the effect on normative attitudes already mentioned. First, I will suggest that the commitments introduced by speech acts have an effect on discourse participants' *beliefs*. This effect is assumed to be twofold: after a commitment is made, discourse participants will normally believe that the act of committing took place and, under certain circumstances, they will moreover come to believe in the commitment's content. Next to beliefs, I will advocate that the commitments resulting from speech acts also influence discourse participants' *expectations* regarding the way the discourse is set to develop. This effect, again, is taken to be twofold: we expect each other to keep acting in accordance with our commitments, and we also expect others to align their commitments with ours, i.e. make the same commitments. From understanding speech acts as commitment-introducing communicative devices, we thus arrive at a tripartite division of a speech act's contribution to discourse: a speech act affects discourse participants' normative attitudes (the commitments themselves), their beliefs (about these commitments) as well as their expectations (on what commitments will follow).

The commitment-based approach thus not only acknowledges a new kind of relation between speech acts and normativity, it moreover offers us a different perspective on more familiar notions studied in discourse theory. Next to the fact then that the commitment-based approach comes with high explanatory power, investigating it more thoroughly can thus potentially give us new insights on the interplay between the various effects of speech acts as well. A formal theory of discourse built on the assumption that speech acts induce commitments would most rigorously show the consequences of this assumption and the way it fits in with an understanding of the discourse as a whole, shedding more light on the logical properties of speech acts and the commitments they result in. Such a theory would thus allow us to keep track of the commitments introduced in a conversation and observe the effect of these commitments, each step of the way.

The current thesis therefore sets itself as its main goal to develop a commitment-based logical formalisa-

---

[4]An analysis of linguistic expressions only in terms of the commitments they result in does of course not explain how meaning can still be expressed without undertaking any commitment, nor what the conversational effect is of conveying meaning in this way. I will come back to this point in the next chapter, where I argue, following Bary (2025), that we need to keep track of discourse participants' beliefs as well to properly account for such cases.

tion of speech acts that are performed in the context of a discourse. More specifically, the goal is to offer a framework that represents the effect of speech acts on the commitments, beliefs and expectations of discourse participants. These attitudes and how they change by performing speech acts will be captured and combined in one dynamic system called Speech Act Logic (**SAL**). Central to the goal of the thesis will be three desiderata that I formulate for **SAL**. These desiderata do not map one-to-one with the three to-be-modeled effects of speech acts discussed above, but of course do relate to them. Whereas the first desideratum expresses the wish to represent a speech act's influence on both the commitments and beliefs of discourse participants, the second will require us to keep track of a speech act's effect on discourse expectations as well. In addition, the last desideratum concerns the applicability of **SAL** to different types of speech acts. Each desideratum will be presented in a separate chapter, where it is first theoretically motivated given the existing literature on commitments and speech act theory, after which an attempt is made to let **SAL** meet this desideratum. In the brief overview of the thesis that now follows, I will shortly mention each desideratum and the strategy that I pursue to formally integrate it.

In Chapter 2, the effect of speech acts on commitment and belief will be investigated, by first honing in on the difference between these two concepts. Bary (2025) argues that, although it has clear advantages to follow the commitment-based approach, beliefs should not be forgotten in the model of discourse this approach gives rise to. Not only do beliefs play a vital role in our utterance choices and the implicit messages we convey, they are also crucial for understanding the cognitive effect speech acts (and the commitments resulting from them) have on discourse participants' information states. A distinction will be made between the effect of a speech act on discourse participants' normative attitudes (what I will call, following Krifka (2024a), the 'performative update') and their mental attitudes (the 'informative update'). As a first desideratum, I pose that a good model of discourse should account for both effects. To this end, I start the construction of **SAL** on the basis of a possible-worlds model, defining both commitment and belief as modal operators. This model, which in the spirit of Krifka (2001) I call a 'commitment state', can be updated with propositions in a way resembling the performative and informative effect of assertion.

Chapter 3 then turns to discourse expectations by extending the model of assertions developed in Chapter 2 to a model including the discourse this type of speech act projects (i.e. how the discourse participants expect the discourse to develop after an assertion has been made). Founded on the commitment-based approach, a three-way distinction is made between the possibility, admissibility and cooperativity of future discourse moves. Performing a speech act may change the projected discourse by manipulating both what is admissible and what is cooperative. The second desideratum, which is to represent this trichotomy formally, is formalised by raising **SAL**'s models from commitment states to commitment spaces, following the work of Cohen and Krifka (2014) and Krifka (2015). A commitment space contains a representation of the current state of the discourse as well as a set of projected future states, the latter of which I use to represent cooperativity. I then show how assertions can get modeled as updates on commitment spaces, potentially changing both the current commitment state and the set of cooperative continuations.

Once the full formal machinery is in place, it is only in Chapter 4 that I start investigating speech acts other than assertions. This is to satisfy the third desideratum, stating that a model of discourse should in theory be generalisable to many types of speech acts. In particular, the chapter focuses on questions (specifically polar questions) and a class Cohen and Krifka (2014) call 'meta speech acts', which are speech acts obtained by embedding illocutionary forces under natural language connectives. Defining new types of commitment space updates, I show how **SAL** is particularly suited for both types of speech acts and take this as evidence that **SAL** is able to accommodate for more than just assertion. In Chapter 5, the closing chapter, the theoretical motivation for and the formalisation of the three desiderata are summarised once again, and the results and limitations of the framework are discussed.

This is certainly not the first project aiming to give a model of speech acts centering around the notion of commitments. Both in linguistics and in artificial intelligence, attempts at formalising commitments have been made. In linguistics, these formalisation attempts are made with the same general intention as that underlying the current thesis, which is to better understand communication and the effect speech acts have on human interlocutors. Contrastingly, in AI research a formal model of commitments is often strived for to represent protocol states in human-computer, multi-agent systems (Singh, 1999). Nevertheless, work on commitments in this field will still turn out to be useful for many of the modeling choices this thesis will be confronted with. The focus, however, will of course be on the formal theories of commitments falling within the linguistic tradition. These theories not only aim to capture commitments but moreover interpret these commitments in relation to the speech acts that cause them and the discourse participants that make them, providing us with an excellent foundation on which to build **SAL**.

Although I let myself be inspired by many commitment-driven linguistic frameworks, there are three in particular that I will explicitly compare **SAL** to. These are (1) the Commitment Space Semantics developed by (among others) Cohen and Krifka (2014), Krifka (2015) and Zhang (2024); (2) the Table-model of Farkas and Bruce (2010); and (3) what I will call the 'cognitive approach' combining the work of Asher and Lascarides (2008a, 2008b) and Lauer (2013). These three frameworks have all been highly influential in the literature and should give a good representation of the range of existing commitment-based theories of discourse. Moreover, they all strongly relate to at least one of the three desiderata that this thesis centers around.

Nevertheless, I believe that the framework I develop in this thesis offers a truly unique approach on discourse modeling, for two important reasons. First, although existing theories of discourse - most notably the three mentioned above - come close to meeting some of the desiderata, neither theory is apt to satisfy all three. Second, I will repeatedly argue throughout the thesis that, in most cases, the desiderata these theories *do* approximate are still not met to their fullest extent. The goal, of course, is to let neither point of critique apply to **SAL**. In other words, I aim for **SAL** to meet all three desiderata in a uniform and satisfactory way. It does this by combining several fruitful ideas present in the three frameworks while at the same time avoiding their weaknesses. In what follows, I will give a short overview of each of the three frameworks mentioned above. In these overviews, I will also already hint at what I believe is missing in them, which should give a rough first impression of the ground that is available for **SAL** to cover.

1.3.1 Commitment Space Semantics

Perhaps the theory most eminently acting as a reference point for this thesis is Commitment Space Semantics (henceforth CSS). CSS finds its origins in the work of Cohen and Krifka (2014), who develop it as a model of speech act embeddings (so-called 'meta speech acts') with the goal to give a formal account of superlative quantifiers. CSS has later been refined and extended to other speech act types, in particular different kinds of questions, by Kamali and Krifka (2020) and Krifka (2015, 2017, 2021, 2022). Most recently, a version of CSS has moreover been used by Zhang (2024) to cover cases of epistemic contradiction, such as (10) below.

(10)   # It might be raining and it is not raining.

(Zhang, 2024, p. 1060)

A model in CSS is called a commitment space, which is a rooted set of commitment states. A commitment state, in turn, is the set of all propositions that are publicly shared by the discourse participants. A commitment state thus represents the information available to the participants, including information

on the commitments the discourse participants have made (Krifka, 2015, 2021, 2022). The unique commitment state laying at the root of a commitment space represents the information available currently in the discourse. The other commitment states in the commitment space branching from this root represent the expected discourse continuations. Since the amount of information can only increase over time, these projected commitment states are supersets of the root.[5]

In CSS, a speech act is seen as restricting a commitment space: if a speech act is performed, we 'move down' the tree to the projected commitment state representing the performance's result (see Figure 1.1 for a visual illustration). Crucially, however, not all possible commitment states get projected. For instance, if a speech act contradicts the information that is already present in the root commitment state (e.g. by invoking commitments that are inconsistent with existing commitments), then the commitment state this speech act would result in is not reachable from the root. Rather, we have to go back to a previous iteration of the commitment space ('move up' the tree again) until we reach a commitment state that does project the result of performing the speech act. This move of retrieving a previous commitment space is called a retraction in CSS. Using retraction, a speaker can withdraw a commitment she undertook and hence 'delete' information from the current commitment state. Given the correct retraction move, the speech act that was not possible to perform before can now be executed.
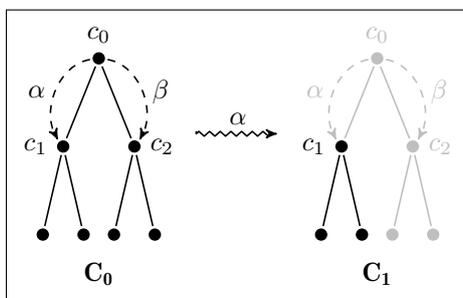

Figure 1.1: A visual representation of a commitment space update in CSS: in the input commitment space C_0, the actual discourse is represented by the commitment state c_0. Two future discourses are projected: one that is the result of performing some speech act α, the other of performing β. After α is performed, we obtain commitment space C_1, which is a restriction of C_0 such that it has c_1 as its root.

With this approach, CSS is able to show how speech acts not only affect the current state of the discourse but also restrict the ways in which a discourse is expected to develop. For instance, CSS models a polar question as projecting two distinct commitment states: one in which the addressee gives (and commits to) an affirmative answer and one in which she gives (and commits to) a negative answer. Since CSS is very successful in providing the tools necessary for dynamically modeling the projected discourse, I will adopt much of CSS's strategy for my own framework. However, I believe CSS also contains two fundamental flaws. First, following Bary (2025), one could argue that CSS fails to capture the 'performative' side of speech acts, since it only models their effect on discourse participants' information states without also showing how speech acts bring about a change in the context of the discourse (namely by actually creating new commitments for the participants). This is the argument I will work out in Chapter 2. Second, since all speech acts are modeled in CSS as transitions within a commitment space, CSS is forced to treat all 'unexpected' speech acts (viz. those resulting in commitment states that are not projected) as mutually identical, namely as requiring a retraction move. However, as I will argue for in Chapter 3, there are important differences in the extent to which a speech act should be seen as unexpected. These differences, I claim, cannot be captured in CSS.

---

[5]Both Krifka (2021) and Zhang (2024) represent commitment states as sets of possible worlds rather than sets of propositions (also see Krifka, 2024a). Since learning new information means deleting possible worlds, on this approach a projected commitment state is of course a *subset* rather than a *superset* of the root.

1.3.2 Farkas and Bruce (2010)

The discourse model of Farkas and Bruce (2010) focuses on assertions and polar questions. It centers around a component the authors denominate 'the Table', reflecting what is at issue in the conversation (also see Farkas, 2010; Farkas & Roelofsen, 2011). On the Table are propositions (together with their linguistic representations) that are currently negotiated among the discourse participants to expand the common ground with. The common ground, in Farkas and Bruce's model, is the collection of all shared commitments together with all mutual background beliefs. By performing speech acts, discourse participants can propose to add certain propositions to the common ground. In the model, speech acts thus change what is on the Table. Speech acts moreover have the ability to create new commitments for the discourse participants. These commitments are registered with the use of a set of individual discourse commitments, one for each participant. This set contains all propositions that the participant has made up until that point in discourse. If a proposition is in each of the participants' commitment sets, then it is the object of a shared commitment and hence part of the common ground.

When a discourse participant makes an assertion, she undertakes a commitment to the assertion's propositional content p and proposes to add p to the common ground. This is reflected in the model by adding p to the participant's commitment set and by placing p on the Table. When, on the other hand, a discourse participant asks a polar question with content p, she proposes to add one of the alternatives raised by the question (in this case p or $\neg p$) to the common ground. Formally, the set $\{p, \neg p\}$ is placed on the Table. Farkas and Bruce assume that it is the discourse participants' goal to settle any pending issues by emptying the Table. Preferably, emptying the Table is done in a way that realises the pending proposal for common ground expansion represented by the propositions on the Table. Thus, when a speaker makes an assertion with content p, the canonical reaction from the addressee would be to confirm this assertion by also committing to p, thereby making p a shared commitment and hence adding it to the common ground. Similarly, an addressee should react to a polar question with content p by answering it, i.e. by committing to either p or $\neg p$, since these two propositions were on the Table.

The addressee may also react to the speaker's proposal in an uncanonical way. For instance, she could deny the speaker's assertion by committing to $\neg p$. This would prevent p from being added to the common ground. To formally distinguish between these 'canonical' and 'uncanonical' ways of reacting to speech acts, Farkas and Bruce (2010) derive a so-called projected set from what is on the Table. In the projected set are all possible future common grounds that include one of the propositions on the Table. They are then able to formally state that a reaction is canonical if it is consistent with one of the common ground expansions represented in the projected set.

Farkas and Bruce's (2010) model has been highly influential to theories on the conversational effect of various linguistic expressions (e.g. Ettinger & Malamud, 2015; Faller, 2019; Hogeweg et al., 2011; Roelofsen & Farkas, 2015). Similarly to CSS, its power resides in the ability to untangle how speech acts restrict the ways in which the discourse is expected to continue. It will therefore play a large role in Chapter 3, where I investigate exactly this phenomenon. However, as was the problem with CSS, the model cannot capture relevant distinctions between speech acts that are not projected. For instance, whereas the model does show how it is somehow less preferred for the addressee to deny a speaker's assertion, it does not concern itself with the fact that such disagreement between two interlocutors still seems more expected than a situation in which the same interlocutor makes inconsistent claims. Moreover, in the same paper featuring her critical comments on CSS, Bary (2025) argues that Farkas and Bruce's common ground does 'double duty' by registering both the content of mutual beliefs and that of shared commitments. This critique, which I will take up in Chapter 2, forms an important motivation for treating commitments and beliefs as strictly separate entities in **SAL**.

1.3.3   The cognitive approach

The last formal commitment-based theory of speech acts I would like to mention is what I will here call 'the cognitive approach', consisting of the work of Asher and Lascarides (2008a, 2008b) and Lauer (2013). In contrast to Farkas and Bruce's (2010) model, the cognitive approach characterises itself by treating normative attitudes (viz. commitments) and mental attitudes (such as beliefs and intentions) as strictly separate entities. The cognitive approach formally integrates both types of attitudes and specifies how they relate to each other in dialogue. It will therefore form the basis of Chapter 2, where I aim to imitate this strategy in **SAL**.

As discussed in section 1.1, our utterances in conversation not always reflect what we believe, nor what we intend others to believe. In line with the normativist tradition, Asher and Lascarides (2008a, 2008b) therefore analyse discourse moves in terms of the commitments they bring about. Accordingly, they argue that 'grounding' (viz. agreement in conversation; see Clark, 1996) may not and often does not reflect any mutual belief and should instead be understood in terms of shared public commitment (cf. Gaudou et al., 2006). At the same time, the authors argue that we need mutual belief in a model of discourse to understand how interlocutors strategically reason about each other's mental attitudes and how this reasoning affects their future discourse moves. According to the cognitive approach, speech acts not only manipulate the interlocutors' commitments but their cognitive states as well: based on what has been said, we can learn about the preferences and intentions of others and adjust our own accordingly.

With this idea in mind, Asher and Lascarides (2008a, 2008b) present an extension of Segmented Discourse Representation Theory (SDRT; Asher & Lascarides, 2003) that not only makes the framework dynamic but also involves a representation of discourse participants' cognitive attitudes. Importantly, the framework features a set of defeasible axioms relating commitment to these cognitive attitudes. Most relevant to the current purposes is an axiom the authors call Sincerity, stating that, normally, one believes what one commits to. A second noteworthy axiom of the system (one that does not necessarily relate commitments to beliefs but is essentially cognitive) is Competence, telling us that, normally, if an addressee believes that a speaker believes a proposition, she herself believes it too (also see Asher & Lascarides, 2003).

Lauer (2013) models the same two defeasible principles (although he does not give them the same names) as pragmatic inferences in a framework he terms Dynamic Pragmatics (DP). Similarly to Asher and Lascarides (2008a, 2008b), Lauer distances himself from intentionalism and believes that communication does not primarily center around the recognition of intentions. Instead, utterances first and foremost have a normative effect. For instance, Lauer assumes that declaratives commit their author to a belief in a proposition (see the discussion above) and that imperatives entail a commitment to a preference. Nevertheless, under specific contextual conditions (in particular Sincerity and Competence) certain mental inferences may follow, which is why DP focuses on the effects speech acts can have on interlocutors' information states (although Lauer does discuss the possibility to integrate a commitment modality in DP for capturing their normative effects as well). In DP, utterances are modeled as removing possible worlds from these information states, reflecting the gradual increase of information available to the interlocutors during conversation. Next to a direct update with the information that an utterance took place, it is also possible that utterances update an information state by adding their content to it - for this second update to take place, however, Lauer assumes that both Sincerity and Competence need to be met in the context of the conversation, mirroring Asher and Lascarides's (2008a, 2008b) theory.

The cognitive approach distinguishes itself from CSS and the work of Farkas and Bruce (2010) by devoting significant attention to the relation between public commitments and private mental attitudes. As a result, the approach is able to show how speech acts affect both and how contextual factors (such as

Sincerity and Competence) determine exactly how commitments translate to beliefs. On the basis of the cognitive approach, I will formalise the intricate interplay between commitments and beliefs in my own framework in Chapter 2. However, an important aspect both the theories of Asher and Lascarides (2008a, 2008b) and Lauer (2013) lack is that speech acts affect the projected discourse as well, contrasting the previously discussed theories. The cognitive approach hence is not able to make predictions regarding the acceptability or preferentiality of different future discourse moves, which is something **SAL** does aim for.

We have now set the stage for the framework that I will be presenting in the rest of the thesis. This framework will couch itself in the normative approach by modeling speech acts as primarily resulting in commitments. In line with the cognitive approach, this does however not mean that I do not at all concern myself with discourse participants' mental attitudes. Rather, beliefs will be assigned an important role in **SAL** as well. The formal integration of commitments, beliefs and their interaction is what I will now turn to first, before formalising the projected discourse and extending **SAL** to different types of speech acts in later chapters.

Chapter 2

Beliefs and commitments

This chapter focuses on the two concepts that will form the skeleton of **SAL**: beliefs and commitments. Bary (2025) argues that both are needed in order to give an accurate analysis of discourse. Her article is mostly concerned with a conceptual review of the discourse model of Farkas and Bruce (2010). As we discussed in the previous chapter, this model captures both the beliefs and commitments of discourse participants, yet Bary notes that it fails to treat the concepts as two separate entities. My goal here will be to work out Bary's argument in detail, first showing how beliefs and commitments are different yet both important concepts for understanding discourse. I will then formalise the two concepts in a first version of **SAL** in section 2.2. In section 2.3 and 2.4, I move towards the formulation of the first desideratum of the to be developed model, discussing the performative and informative update that speech acts induce. In section 2.5, I integrate my findings in **SAL**, which I apply to some examples in section 2.6.

2.1 Two separate entitites

It should be clear from the previous chapter that beliefs alone are not enough to understand our behaviour in communication, since, as we have seen, assertions do not always express belief. I therefore argued for a normative perspective, in which speech acts are seen as introducing new commitments. Commitment, being a public and normative notion, is clearly different from belief or any other cognitive attitude, which all are private and mental in nature. Although commitments may help us better understand communication, Bary (2025) argues that they cannot act as a replacement for belief: just as commitments cannot be reduced to beliefs, beliefs cannot be reduced to commitments. Similarly to how our utterances may not reflect what we believe, we of course can also hold beliefs that we keep to ourselves and hence never commit to. Of course, it does not immediately follow from the fact that belief cannot be explained in terms of commitment that we should assign a significant role to belief in a model of discourse. To motivate that we should, I will give two arguments for attaching importance to beliefs too, roughly following Bary's line of argument.

First, Bary (2025) discusses how we often base our utterances on what we take to be mutually believed, even when we are not committed to these beliefs. The word 'mutual' here should be taken to express recursivity. That is, if we take $\mathbf{B}_x\,p$ to stand for "agent x believes proposition p", then p is mutually believed among two agents a, b iff:

$$\begin{array}{lcl}
\mathbf{B}_a\, p & \text{and} & \mathbf{B}_b\, p \\
\mathbf{B}_a\, \mathbf{B}_b\, p & \text{and} & \mathbf{B}_b\, \mathbf{B}_a\, p \\
\mathbf{B}_a\, \mathbf{B}_b\, \mathbf{B}_a\, p & \text{and} & \mathbf{B}_b\, \mathbf{B}_a\, \mathbf{B}_b\, p \\
& \vdots &
\end{array}$$

(Geurts, 2019, p. 16)

In words, a proposition is mutually believed if everyone believes p, everyone believes that everyone believes p, and so on. Following Stalnaker (1978) (as well as Bary (2025) and many others), I use the term *common ground* to refer to the set of all mutually believed propositions.

Bary (2025) thus argues that keeping track of the common ground is vital for understanding discourse participants' utterance choices (see also Stalnaker, 1978). For instance, if it is mutually believed in our conversation that you like classical music, this might be a good reason for me to start talking to you about Bach. Crucially, I need not commit to this mutual belief first for taking it as a starting point of our conversation. Many uses of presuppositions reveal the same pattern: for uttering the sentence "The king of the Netherlands is at a ceremony", I do not have to make explicit (and hence I need not commit to) the fact that I mean to refer to Willem-Alexander if I take this to be mutually believed among us. And not only background beliefs, but unvoiced mutual beliefs arising from events happening during the discourse may play a part in our utterances, too. Bary discusses Stalnaker's example of a goat walking into the room: without anyone asserting first that a goat just entered, it would be perfectly reasonable for one of the interlocutors to ask what *the* goat is doing in the room, since he may of course assume that everyone witnessed the event and hence that it is a mutual belief.

Moreover, conveying one's beliefs does not always require one's commitment. Recall the example of a professor writing a recommendation letter for one of her students from Chapter 1:

(7)   The student has nice handwriting.

Bary (2025) theorises that by formulating meaning implicitly, it is possible to avoid commitment to it. By packaging her belief as a conversational implicature, the professor in (7) can thus express it without needing to commit to it. One argument for interpreting conversational implicatures as non-committal is that they are cancelable (cf. Mazzarella et al., 2018), as the following conversation, also repeated from Chapter 1, shows (based on MacFarlane, 2011, p. 3):

(8)   A: The student has nice handwriting.
        B: So you're saying that the candidate is unsuitable?
        A: No no, I was careful not to say that.

Similarly, Bary (2025) herself discusses an example from Camp (2018), in which a driver gets pulled over by a police officer for speeding.

(11)   I'm in a bit of a hurry. Is there any way we can settle this right now?

(Camp, 2018, p. 43)

The driver can be seen as insinuating a bribe, although, again, he does not make this insinuation explicit, thereby seemingly avoiding commitment to it:

(12)   A: I'm in a bit of a hurry. Is there any way we can settle this right now?
         B: So you're proposing a bribe?
         A: No no, I was careful not to say that.

While in both examples the speaker does not undertake a commitment to the insinuation she is expressing, she expresses it nevertheless. As a result, the insinuation becomes mutually believed, or is at least intended to be. If the professor's communicational attempt in (7) is successful, it will be mutually believed between her and whoever reads her recommendation letter that she thinks the student is an unsuitable candidate. Similarly, the driver in (11) seems to intend to make it a mutual belief that she is proposing a bribe. Understanding implicatures as non-committal gives us an explanation for why we sometimes package information this way: not having to commit to a message you still want to get across minimises the social risk it entails, which can be useful when this message has the potential to be reputationally costly (Bary, 2025; Camp, 2018; Mazzarella et al., 2018; Pinker et al., 2008).

An approach that studies only the normative effects of linguistic expressions can obviously not take care of meaning not committed to and hence misses the full effect insinuations such as (7) and (11) have on a discourse. For this, beliefs need to be taken into account as well. In fact, not only are beliefs needed for explicating hidden layers of meaning in utterances that normativism can already analyse on a surface level: without them, assertions of belief seem to not be analysable either. For instance, only interpreting assertions in terms of commitment without also acknowledging in some way their relation to belief would not lead us to the conclusion that (4a) is a stronger statement than (4b), repeated below.

(4)   a.   The buffet is open.
        b.   I believe that the buffet is open.

<div align="right">(Krifka, 2024b, p. 52)</div>

To understand the choice for and the full effect of any given discourse move, we thus need beliefs as much as we need commitments. This gets acknowledged in the influential discourse model of Farkas and Bruce (2010), who work with a common ground containing mutual background beliefs. However, as we discussed, their common ground also contains the propositions that the discourse participants have jointly committed to. Bary (2025) hence notes that this makes it impossible to represent the different roles beliefs and commitments may play, since they eventually get treated formally as the same kind of object: the common ground may contain commitments not believed in and one's commitments partly consist of beliefs not committed to (in fact, Bary already spots the same inconsistency in Gunlogson (2001)).

I aim to avoid this conceptual mistake and model the common ground as a device solely reserved for mutual beliefs. That is not to say, however, that mutual commitments should not be taken into consideration as well (cf. Asher & Lascarides, 2008a). According to Geurts (2019), it is in fact the eventual goal of communication to increase the set of mutual commitments (which in his work is called the common ground), since mutual commitments are needed for the realisation of joint future actions. Only if we commit to the same courses of action are we able to reach the same goals. By using speech acts, we are able to communicate our commitments and thus coordinate such actions (see also Geurts, 2017). Thus, if someone makes a commitment, it is most cooperative to respond by making an identical commitment. Arguably, this is also why confirming but not denying an assertion is often seen as the cooperative or canonical response (Farkas & Bruce, 2010; Gunlogson, 2008; Krifka, 2022), and why a conversation is sometimes said to be 'in crisis' (Farkas & Bruce, 2010) or a proposition to be 'unresolved' (Gunlogson, 2008) if one but not all discourse participants have made a commitment to that proposition. I will return to cooperativity as a feature of future discourse moves in Chapter 3.

It could thus be valuable to consider the set of shared commitments in a theory of communication, since increasing this set is what the interlocutors hope to eventually achieve through conversing with each other. Shared commitment, however, is not the same as mutual commitment, if we define the latter similarly to mutual beliefs. That is, if $\mathbf{C}_{a,b}\, p$ means that agent a is committed towards agent b to a proposition p

(more on this notation in the next section), then we could say that p is mutually committed to iff:

$$\begin{array}{rcl} \mathbf{C}_{a,b}\, p & \text{and} & \mathbf{C}_{b,a}\, p \\ \mathbf{C}_{a,b}\, \mathbf{C}_{b,a}\, p & \text{and} & \mathbf{C}_{b,a}\, \mathbf{C}_{a,b}\, p \\ \mathbf{C}_{a,b}\, \mathbf{C}_{b,a}\, \mathbf{C}_{a,b}\, p & \text{and} & \mathbf{C}_{b,a}\, \mathbf{C}_{a,b}\, \mathbf{C}_{b,a}\, p \\ & \vdots & \end{array}$$

<div align="right">(Geurts, 2019, p. 17)</div>

The reason that Geurts assumes a shared commitment (i.e. $\mathbf{C}_{a,b}\, p$ and $\mathbf{C}_{b,a}\, p$) to also always be a mutual commitment is because of a general law he calls Acceptance:

> *Acceptance.* $\mathbf{C}_{x,y}\, p$ entails $\mathbf{C}_{z,z'}\, \mathbf{C}_{x,y}\, p$, for all agents x, y, z, z'.

<div align="right">(based on Geurts, 2019, p. 19)</div>

His argument for assuming Acceptance is that a commitment can only truly be called a commitment if it gets accepted by all interlocutors, i.e. if everyone commits to the commitment. Only when everyone accepts a commitment can we actually assume the committer to have a responsibility to keeping his commitment, which is of course exactly what it means to have a commitment. It is thus impossible to have commitments if we do not know of each other that we have them (Elder, 2021; Geurts, 2024; Kibble, 2006).

2.2   Commitment states

Now that we have a clear idea in mind of the notions a model of discourse should at least consist of (viz. belief and commitment), we can already lay a first hand at our logic of commitments **SAL**. Since we want to formalise communication, we need a way of representing different discourse participants. I will do this by fixing a (finite) set $\mathcal{A}$ of agents (labeled a, b, ...). Together with a countable set of propositional atoms Atom (labeled p,q,...), this gives rise to the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ of **SAL**, containing sufficient material for satisfying Desideratum 1 that is to be introduced in section 2.4. I will then extend this language in succeeding chapters. $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ is defined as follows:

> **Definition 1.** (Basic language) Let a finite set of agents $\mathcal{A}$ and a countable set of propositional atoms Atom be given. For $p \in$ Atom, the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ of **SAL** is given by the following BNF:
>
> $$\pi ::= p \mid \neg\pi \mid \pi \wedge \pi \mid \{\mathbf{B}_a\, \pi\}_{a\in\mathcal{A}} \mid \{\mathbf{C}_{a,b}\, \pi\}_{(a,b)\in\mathcal{A}^2}$$

Thus, the basic language enables us to express atomic propositions, as well as the standard connectives. I take negation ($\neg$) and conjunction ($\wedge$) as primitives and from there derive the other connectives.[6] Belief is a binary relation: the belief of an agent a in a proposition p is represented as $\mathbf{B}_a\, p$. Commitment is a trinary relation: the language contains formulas of the form $\mathbf{C}_{a,b}\, p$, denoting that an agent a is committed towards some agent b to a proposition p (cf. Geurts, 2019; Singh, 2000). Normally, a and b will be different agents, but the framework in principle allows agents to make 'private' commitments to themselves as well (see Cohen & Levesque, 1990; Geurts, 2018). The motivation for considering a receiving agent b as well in the commitment relation is the observation that the same agent may without incident make commitments towards different agents that taken together would be inconsistent (though we probably expect each other to avoid such scenarios as much as possible (Geurts, 2019)), as example (13) shows:

---

[6]That is: $\psi \vee \chi := \neg(\neg\psi \wedge \neg\chi)$, $\psi \to \chi := \neg\psi \vee \chi$, $\psi \leftrightarrow \chi := (\psi \to \chi) \wedge (\chi \to \psi)$, $\bot := p \wedge \neg p$ and $\top := \neg\bot$.

(13) a. Alice to Charlie: you are smart.
      b. Alice to Bob (Charlie out of earshot): Charlie is stupid.

(based on Gaudou et al., 2006, p. 225; Asher & Lascarides, 2008a, p. 35)

Still, choosing to represent commitments as a relation between two agents is not completely indisputable. For instance, Liau (2001) only considers the author of the commitment, Asher and Lascarides (2008a, 2008b) model commitment as a relation between an agent and a group, and Kibble (2006) takes the relation to involve a 'hearer' different from the speaker and addressee as well. Moreover, cases seem to exist in which more than one agent takes on a certain commitment (Portner, 2004), or in which the agent who is meant to make the commitment is not specified, as in (14).

(14) Somebody get this table out of the way.

(Mastop, 2005, p. 83)

To keep the current framework as clear and simple as possible, I choose to only model two single (fully specified) agents in any commitment relation. That being said, the system will be formulated in a general enough way as to accommodate for the other options discussed as well.

I will model both belief and commitment as modal operators. As we will shortly see, this creates an elegant formal representation of cases where belief and commitment operators embed under themselves or under each other. Since commitment is a normative notion stating the obligations of discourse participants, I follow Dignum and van Linder (1997) and others in defining commitment in terms of the deontic obligation relation $O_{x,y}$ (one for each agent pair $(x, y) \in \mathcal{A}^2$). If a world w is $O_{a,b}$-accessible, this can be interpreted as w being a deontic possibility for agent a towards agent b. That is, it is permitted for a to commit (towards b) to the state of affairs described in w. I take the $O_{x,y}$-relation to be transitive and Euclidean (cf. Asher & Lascarides, 2008a),[7] ensuring positive and negative introspection of commitment:

1. *Positive introspection of commitment.* If $\mathbf{C}_{x,y}\,\pi$ then $\mathbf{C}_{x,y}\,\mathbf{C}_{x,y}\,\pi$, for all agents $x, y \in \mathcal{A}$.

2. *Negative introspection of commitment.* If $\neg\,\mathbf{C}_{x,y}\,\pi$ then $\mathbf{C}_{x,y}\,\neg\,\mathbf{C}_{x,y}\,\pi$, for all agents $x, y \in \mathcal{A}$.

Positive introspection states that if someone has a commitment to a proposition, they are also committed to that commitment. Similarly, negative introspection states that if someone does not have a commitment to a proposition, they are committed to not being committed to that proposition. Introspection seems appealing, given the normative nature of commitment: having to act in a way consistent with the truth of a proposition of course also means having to act in a way consistent with having to act in a way consistent with the truth of a proposition (cf. Geurts, 2024). Not having to act in a way consistent with a proposition similarly obliges one to not having to act in a way consistent with a proposition.

For belief, I make use of the belief relation B_x (one for each agent $x \in \mathcal{A}$) from doxastic logic (see e.g. Fagin et al., 1995). If a world w is B_a-accessible, then agent a sees w as an epistemic possibility. As is standard practice, I take this relation to be transitive, Euclidean and serial (cf. Asher & Lascarides, 2008a).[8] This gives us the result that agents have positive and negative introspection on their beliefs, and that their beliefs cannot be inconsistent:

1. *Positive introspection of belief.* If $\mathbf{B}_x\,\pi$ then $\mathbf{B}_x\,\mathbf{B}_x\,\pi$, for all agents $x \in \mathcal{A}$.

2. *Negative introspection of belief.* If $\neg\,\mathbf{B}_x\,\pi$ then $\mathbf{B}_x\,\neg\,\mathbf{B}_x\,\pi$, for all agents $x \in \mathcal{A}$.

---

[7] A relation R is transitive if for any worlds u, v, w, uRv and vRw entail uRw. It is Euclidean if uRv and uRw entail vRw.

[8] A relation R is serial if for any world v there is some world w such that vRw.

3. *Consistency of belief.* It is never the case that $\mathbf{B}_x \perp$, for any agent $x \in \mathcal{A}$.

Crucially, we do *not* have consistency of commitment. This is intentional, since we would like to have a way of expressing violations of commitment, which take place precisely when an agent makes inconsistent commitments. I will return to the importance of having a way to express violation in Chapter 3.

Since we are working within a modal framework, **SAL** will be based on a Kripke-style semantics. I will borrow the term *commitment state* from Commitment Space Semantics (CSS; Cohen & Krifka, 2014; Krifka, 2015) to refer to this set of possible worlds, together with an interpretation function and the accessibility relations necessary for defining belief and commitment. Note, however, that this makes my use of the term different from its use in CSS, where it either refers to a set of propositions (Cohen & Krifka, 2014; Krifka, 2015, 2017, 2022) or just a set of possible worlds (Krifka, 2021; Zhang, 2024).

> **Definition 2.** (Commitment states) A *commitment state* over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ is a tuple $c = \langle W^c, \{B^c_a\}_{a \in \mathcal{A}}, \{O^c_{a,b}\}_{(a,b) \in \mathcal{A}^2}, I^c \rangle$, such that:
>
> - W^c is a non-empty set of possible worlds
>
> - $\{B^c_a\}_{a \in \mathcal{A}} \subseteq W^c \times W^c$ is a belief relation (one for each agent) that is transitive, Euclidean and serial.
>
> - $\{O^c_{a,b}\}_{(a,b) \in \mathcal{A}^2} \subseteq W^c \times W^c$ is a commitment relation (one for each agent pair) that is transitive and Euclidean.
>
> - $I^c : \mathsf{Atom} \to \mathcal{P}(W^c)$ is a valuation function.
>
> We write $c, w \models \pi$ to say that π is true in a world w in c. We write $c \models \pi$ to say that $c, w \models \pi$ for all $w \in W^c$.

We now interpret commitment and belief as deontic and doxastic necessity respectively, arriving at the following truth conditions on the level of commitment states:

> **Definition 3.** (State satisfaction) Let c be a commitment state over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $w \in W^c$ a world in c, $a, b \in \mathcal{A}$ two agents, and $p \in \mathsf{Atom}$ and $\pi, \tau \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ all propositions. Then:
>
> - *Atomic propositions.*
> $c, w \models p$ iff $w \in I^c(p)$
>
> - *Connectives.*
> $c, w \models \neg\pi$ iff $c, w \not\models \pi$
> $c, w \models \pi \wedge \tau$ iff $c, w \models \pi$ and $c, w \models \tau$
>
> - *Modal operators.*
> $c, w \models \mathbf{B}_a \pi$ iff $c, v \models \pi$ for all worlds $v \in W^c$ such that $wB^c_a v$
> $c, w \models \mathbf{C}_{a,b} \pi$ iff $c, v \models \pi$ for all worlds $v \in W^c$ such that $wO^c_{a,b} v$

Now, we can easily define mutual belief as follows: a proposition π is mutually believed in a commitment state c and world $v \in W^c$ iff $c, w \models \pi$ for all $w \in W^c$ such that vBw, where $B = \bigcup_{a \in \mathcal{A}} B_a$.

Similarly, a proposition π is mutually committed to in a commitment state c and world $w \in W^c$ iff $c, w \models \pi$ for all $w \in W^c$ such that vOw, where $O = \bigcup_{(a,b) \in \mathcal{A}^2} O_{a,b}$.

Notice how going modal also allows us to express believing in a commitment or committing to a belief, without needing any additional machinery. As we will see in section 2.5, this will come in useful when formalising the performative and informative update potential of speech acts in the framework.

Before we move on, let me make a quick note on the entailment of belief and commitment. Since in the current system both belief and commitment are defined as modalities, both sets are deductively closed. For belief, this kind of 'logical omniscience' is often regarded as an unwelcome result (e.g. Stalnaker, 1991). For instance, I could believe (15a) without needing to believe (15b): although all mammals are vertebrates, I might not be aware of this fact and hence not believe that (15a) entails (15b).

(15)   a.   Dogs are mammals.
       b.   Dogs are vertebrates.

This reasoning, arguably, does not apply to commitments. Since commitments are not private or mental but public and normative, there is something to say for the argument that commitments *are* deductively closed. If I commit to (15a), I promise to act in a way that is consistent with dogs being mammals. Since the fact that dogs are mammals entails that dogs are vertebrates, this means that I should also act in a way that is consistent with dogs being vertebrates. Even if I am not aware of the logical consequences of my commitment, I am still liable for any inconsistencies existing between my earlier commitments and what is entailed by my present commitment (Lance & Kremer, 1994; Walton, 1992). Hence, by overtly committing to (15a), I am implicitly committed to (15b), too.

Although most scholars seem to agree that one's set of commitments should be deductively closed (Cohen & Krifka, 2014; Geurts, 2019; Gunlogson, 2008; Krifka, 2015, 2024a; Yamada, 2012), others argue it is not. An agent might, for instance, commit to a contradiction, in which case it seems an unlikely result that she is committed to everything (Lance and Kremer, 1994; note however how this *is* the result our model produces). Nevertheless, those who argue against full deductive closure of commitments, such as Hamblin (1970) and Kibble (2006), still believe that agents should avoid commitments that immediately can be seen to contradict each other. There is thus much discussion to be had on the extent to which commitments can be implicit. To keep the current project feasible, I will follow what appears to be a majority of authors in simply assuming that an overt commitment comes with a commitment to all that it logically entails.

2.3   The performative and informative update

In section 2.1, I gave two arguments for taking into account beliefs next to commitments in a model of discourse: beliefs influence our utterance choices and moreover can be conveyed to others in conversation without being the object of any commitment, for example with the use of conversational implicatures. Discourse participants not only update their beliefs based on what is conveyed implicitly, however. They also do this in light of propositions that are explicitly asserted. For instance, if your boss tells you that the office is closed tomorrow, then you have good reason to add this information to your pile of beliefs. A speech act thus has an effect on both the normative and the mental attitudes of discourse participants, as both commitments and beliefs may change in light of that speech act. These two changes constituted in the discourse are what Krifka (2024a, 2024b) calls the *performative* and *informative* update, respectively. Although I borrow the terms and general idea from Krifka, I will also diverge from his analysis in a number of important ways. Specifically, Bary (2025) points out that Krifka's performative and informative update are essentially two components of the informative update. Using her argument, I will arrive at three different kinds of updates: a performative update and two informative updates, which

I will call the informative update 1 and 2.

2.3.1 The performative update

First and foremost, speech acts can be seen as updates or transformations of the discourse *context* (Asher & Lascarides, 2003; Krifka, 2024a; Recanati, 1987; Szabolcsi, 1982). Upon performance, they change what is true in the world in which the discourse takes place (Austin, 1975). This is what we can call the *performative* update (Krifka, 2024a). The performative update a speech act brings about is largely due to its illocutionary contribution.[9] Given that different speech acts make different illocutionary contributions, the specific way in which they change the context differs as well. Perhaps the most clear case of a performative update is one that is the result of performing what Austin (1975) called 'conventional' speech acts, also known as 'declaratives' (Vanderveken, 1994). Consider, for instance:

(16)   I hereby pronounce you husband and wife.

If declared by someone with the authority to perform a marriage, (16) becomes true in the context of the discourse: by making the declaration, the addressees are married, whereas they weren't before the performance of the act.

This is not the only way in which the discourse context may change as a result of performing speech acts. In Chapter 1, we have discussed the normativist stance that assertions result in commitments: by asserting a proposition, the speaker commits to the truth of that proposition (Brandom, 1983; MacFarlane, 2005; Searle, 1979; Tuzet, 2006; Watson, 2004). Crucially, undertaking a commitment comes with a change in one's normative status. If I undertake a new commitment, I have from that point on an obligation to act in accordance with that commitment, whereas I did not have this obligation before. Note how this change in normative status stands independent of any mental attitudes: irrespective of what goes on in our minds, we are socially bounded to the commitments we make. Thus, because an assertion brings about commitments, it changes the world: it updates the (social) context in which the discourse takes place (Bary, 2025).

We could extend our commitment-based analysis of assertions by arguing that *any* speech act bares at least some relation to commitments (Geurts, 2019; Tuzet, 2006).[10] Most speech acts, including assertions, promises and orders, *directly* manipulate the commitments of discourse participants: if someone asserts p, they become committed to p, with immediate effect. Alternatively, a speech act may *indirectly* induce commitments by influencing the actions that some discourse participant will or should undertake in the future of the discourse. In Chapter 4, I will analyse questions and speech act denegations as speech act types that *project* commitments while leaving the discourse participants' current commitments untouched. For instance, asking a question whether p can be seen as a request to the addressee to commit to either p or $\neg p$ without changing what either the speaker or addressee is presently committed to. Thus, although perhaps not every speech act creates commitments instantaneously, all speech acts do eventually bear upon commitments.

Although we have seen that a speech act such as a declaration can affect the context *beyond* changes in normative attitudes, I will generally assume that commitments are always at least part of the effect

---

[9]Technically, a speech act also changes the physical world because of its locutionary effect, e.g. because uttering a sentence creates sound waves (Krifka, 2024a). Since this effect is obviously of little linguistic interest, I omit it here in my discussion of the performative update.

[10]This may be somewhat of an overgeneralisation. For instance, Lauer (2013, p. 213) argues that exclamatives such as "My, how high this building is!" arguably do not result in any commitment. I would like to think that such a position is worth considering but also certainly refutable (see e.g. Geurts (2019), who analyses exclamatives and related speech acts as commitments to a psychological state). In what follows, I will therefore simply assume that any speech act is linked to commitment.

of performing that speech act. That is, although a declaration primarily changes the context insofar the proposition it expresses becomes true, it indirectly causes all discourse participants to commit to the proposition as well. Hence, the performative update always resides in the creation of (current or projected) commitments.

> **The performative update.** After performing a speech act, commitments are created in the context of the discourse.

This informal definition of the performative update seems also to be the one Krifka (2024a, 2024b) envisions. Yet in his model, he represents this update as a change in discourse participants' information states. That is, after performing a speech act, all discourse participants add the information that a new commitment has been made to their representation of how the world could be like, i.e. to their information state. Bary (2025) critically remarks that this is not a performative but rather another type of informative update, one that comes in addition to the informative update Krifka already acknowledges himself. We thus obtain a two-faceted informative update - this is what I will turn to next.

2.3.2   The informative update

Speech acts not only bring about a change in the factual state of affairs. They also manipulate the information available to discourse participants (what Austin (1975) calls the 'descriptive' use of language). The information a speech act introduces can be grouped into two categories, giving rising to two ancillary updates: the informative update 1 and the informative update 2.

The informative update 1 abides in the assumption that when a speech act gets performed, all discourse participants believe that it was performed. From our assumption that speech acts induce commitments, it follows that all participants come to believe that the act of committing took place.[11] For instance, when I say to you that penguins live on Antarctica, we both update our stock of beliefs with the information that I am now committed to penguins living on Antarctica. Of course, this is somewhat of an idealisation, given that 'uptake', as Austin (1975) calls it, may not be perfect: you could have misunderstood what I said or be out of hearing range (Clark, 1996; Geurts, 2019), in which case you miss the fact that I have made a new commitment.[12,13] I omit such possibilities here and instead focus on the more interesting cases in which all interlocutors do witness the performance of a speech act (cf. Asher & Lascarides, 2008a).

> **The informative update 1.** After performing a speech act, everyone believes that the speech act is performed.

Similarly to how we defined Acceptance, we could formally denote the informative update 1 as the following rule, which I call Acknowledgment:

> *Acknowledgment.* $\mathbf{C}_{x,y}\, p$ entails $\mathbf{B}_z\, \mathbf{C}_{x,y}\, p$, for all agents $x, y, z \in \mathcal{A}$.

Whereas both the performative update and informative update 1 relate to a speech act's illocutionary effect, the informative update 2 is linked to its perlocutionary effect. This effect could be seen as the

---

[11] As part of the locutionary act, all participants will also come to believe that the utterance containing the speech act was produced (Bary, 2025). This is true of course, but not something I will spend much attention on, for reasons similar to why I chose to ignore the speech act's locutionary contribution in the discussion of the performative update.

[12] Going back to our previous discussion on Acceptance, we could argue that uptake is not an idealisation but a precondition for committing, since a commitment is only deserving of the name if it gets *accepted* (i.e. witnessed) by everyone in the conversation.

[13] Clark (1996) defines uptake as grounding on the highest level, i.e. agreeing with and taking up the speaker's project. This is of course not what I mean with the word here: uptake as I use it refers to fully receiving and understanding the speaker's utterance, which constitutes the second highest level of grounding in Clark's model.

ultimate goal of the speech act, since arguably the reason for communicating anything at all is to bring about some kind of effect on the addressee's cognitive attitudes. For assertions, the perlocutionary effect is often to persuade the addressee on the truth of some proposition and thus get her to believe the assertion's content (Krifka, 2022, 2024a). If this goal is achieved, we obtain a second kind of informative update: the informative update 2, plainly called the informative update in Krifka (2024a, 2024b).

> **The informative update 2.** After performing an assertion, the addressee believes its propositional content to be true.

Unlike the informative update 1, the informative update 2 does not always take place (Bary, 2025; Jary, 2007; Stalnaker, 1978). This is something Krifka (2024a) notes as well, who therefore regards it as a cancelable update. One of the reasons for the defeasability of the informative update 2 is that not all assertions are made with the intention that the addressee comes to believe in their content, as discussed in Chapter 1. Recall, for instance, the case of being openly insincere, such as the actor playing in Hamlet asserting (2) to his audience.

(2)   I am the king of Denmark.

Moreover, even if the aim is to persuade the addressee, the informative update 2 may not take place: the addressee might suspect that the speaker is lying or that she is not knowledgeable on the subject matter. Thus, whether the informative update 2 goes through or not seems to depend on contextual conditions. I will here focus on two of these conditions, which were already mentioned in section 1.3.3, where I introduced the cognitive approach of Lauer (2013) and Asher and Lascarides (2008a, 2008b). Following Asher and Lascarides (2003, 2008b), I call these conditions Sincerity and Competence.

1. *Sincerity.* The addressee takes the speaker to be honest.

2. *Competence.* The addressee takes the speaker to be well-informed.

Now, suppose an assertion is made in a context where both Sincerity and Competence are assumed. Because the addressee thinks the speaker is honest (Sincerity), she will reason that the speaker himself believes the assertion's content. And because the addressee additionally takes the speaker to be well-informed (Competence), she will infer that his belief is true and so adopt it. It seems that if we drop either of these two assumptions, the informative update 2 does not go through.

Although there are cases in which Sincerity and Competence are lacking, it seems safe to assume that at least Sincerity is true by default (Asher & Lascarides, 2008a). More often than not, we say things that we truly believe in. After all, under normal circumstances it doesn't make much sense to commit to things you know or believe to be false, since this would require you to follow a course of action that is not in your best interest (MacFarlane, 2005). Moreover, committing involves a certain risk: if a speaker intentionally commits to a falsehood, she would do so knowing that she shall have to defend her claim when asked to do so and that she will be held liable in case others discover it is not true (Bary, 2025; Brandom, 1983; Geurts, 2019; Krifka, 2015; Lauer, 2013).

2.4   Desideratum 1

We so far have seen how beliefs as well as commitments play an important role in discourse and how speech acts manipulate both concepts. Commitments change because of the performative update a speech

act brings about, whereas beliefs change as a result of the (possibly twofold) informative update. This forms the basis for formally understanding speech acts: by finding a way of integrating the performative and informative update in the framework, we can model the discourse as a 'scorekeeping game' (Lewis, 1979), where participants add beliefs and commitments to their own and each other's scoreboard. We need both the performative and informative update to fully understand how speech acts may alter not only participants' normative but also their cognitive scores. I therefore pose that any good theory of discourse should meet Desideratum 1:

> **Desideratum 1.** A good model of discourse should capture both the performative and informative effect of speech acts.

Note that if a framework wants to model the performative and informative update, it has to have the ability to express discourse participants' commitments and beliefs as separate entities. Desideratum 1 thus also ensures that Bary's (2025) wish (viz. to distinctly keep track of both commitments and beliefs) is met. Ideally then, a theory satisfying Desideratum 1 will be able to account for differences between plain assertions and assertions of belief, such as the fact that (4a) is a stronger statement than (4b).

(4)   a.   The buffet is open.
      b.   I believe that the buffet is open.

(Krifka, 2024b, p. 52)

It would not be a completely new achievement if **SAL** manages to satisfy Desideratum 1. Asher and Lascarides (2008a, 2008b), for instance, give a model of speech acts that takes into account their effect both on the commitments and the cognitive states of the participants, acknowledging the cancellability of the informative update 2 by formulating Sincerity and Competence as defeasible axioms of the system. The other theory belonging to the cognitive approach, Lauer's (2013) Dynamic Pragmatics (DP), comes close to meeting Desideratum 1 as well. Since DP models utterances as events updating discourse participants' information states with Sincerity and Competence as contextual conditions that may or may not be operative, Lauer can represent the informative update of assertions (which he calls declaratives) and other types of speech acts. The performative update is present less so in Lauer's work, but he does mention a possible extension of his framework in which commitments play a more prominent role.

The other two commitment-based theories of discourse talked about in Chapter 1 seem to fail Desideratum 1. In CSS, a speech act's effect on the current state of the discourse is registered as an update of a commitment state. In Krifka (2015, 2021, 2022), this commitment state seems to represent the discourse participants' collective information state, given that commitment states are here interpreted as "non-empty sets of propositions that represent the information about the world and time at which the conversation takes place - more specifically the information that the interlocutors assume to be shared" (Krifka, 2022, p. 95). Similarly to Krifka (2024a, 2024b) then, CSS seems to focus too much on the informative effect of speech acts, not considering their performative effect as well (Bary, 2025). Moreover, by only modeling the shared information and not also the discourse participants' individual beliefs, I believe CSS misses out on the informative update 2. After all, it could be that a speaker is successfully lying, in which case she does not believe what she says yet does manage to have the addressee believe in it. In this case, the speaker's utterance is not a shared belief, yet the informative update 2 *does* take place. In other work on CSS (e.g. Cohen & Krifka, 2014; Krifka, 2001, 2017), a commitment state is taken to be "the set of public commitments that the interlocutors have accumulated up to the current point in conversation" (Cohen & Krifka, 2014, p. 48). It might be argued that, on this rendition of CSS, commitment states should be interpreted as tracking the performative update. If this is the case, however, then CSS loses a representation of the informative update. The problem, it seems, is that CSS is unable to capture

both the performative and informative update simultaneously: given that a commitment state is simply modeled as just one set of propositions (or possible worlds), CSS can either represent the performative update or (part of) the informative update with that set, but never both.

The model of Farkas and Bruce (2010) is not able to cover Desideratum 1 either. Their model interprets speech acts in terms of commitment changes - although, confusingly, when these commitments are shared they end up in the common ground which also contains participants' (background) beliefs (again, see Bary's (2025) critical comments). In any case, their model does not seem to make a distinction between the performative effect of speech acts (viz. creating individual or mutual commitments) and either of the two informative updates speech acts bring about (viz. creating individual or mutual beliefs). Additionally, since both CSS and Farkas and Bruce's (2010) model do not capture beliefs distinctly from commitments, it should come as no surprise that neither theory seems able to cover cases such as (4): neither theory has a way of expressing commitments to a belief, let alone of explaining how such commitments differ from commitments to plain propositions.

If **SAL** manages to achieve Desideratum 1, it will thus have an explanatory advantage in comparison to the theories not apt for representing both the performative and informative update of discourse. The cognitive approach, which does account for Desideratum 1 (especially Asher & Lascarides, 2008a, 2008b), will be shown to miss out on the other two desiderata, which will be introduced in the following chapters. In the following section, I will formally implement Desideratum 1 into **SAL**. Since **SAL** enriches CSS's notion of a commitment state by interpreting it as a full model of both commitments and beliefs rather than simply a set of possible worlds, it can avoid CSS's problem of not being able to capture the performative and informative update simultaneously. After having shown how **SAL** concretises this idea, I apply the newly extended system to some examples presented in Chapter 1, including (4) above, and show that it is indeed able to give the right predictions.

2.5 Dynamic commitment states

The framework as it has been sketched so far is static in nature: it is able to capture agents' beliefs and commitments, but not yet how both may change in dialogue as a result of performing speech acts. To meet Desideratum 1, we thus need to make **SAL** dynamic. In Chapter 3 and 4, I will make use of Propositional Dynamic Logic, where modal operators are used to define a class of actions (Fischer & Ladner, 1979), which in our case will be speech acts. For now however, given that we have not yet reached the model-theoretic entities to which these actions can be applied, I only show how commitment states can get updated with a given proposition from our basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$. This will already be enough to model the effect of simple assertions on the level of commitment states and thus to demonstrate how the performative and informative update are captured in the framework.

Updating a commitment state with a proposition imitates the performative update induced by an assertion. Since typical assertions involve a speaker and an addressee, so will the propositional update. According to our discussion on the performative update, if a speaker a asserts a proposition π to an addressee b, a becomes committed towards b to π. Thus, when updating a commitment state with proposition π, indexed by the speaker a and addressee b, $\mathbf{C}_{a,b}\,\pi$ will be true in the updated commitment state. Given the truth definitions, this means that, for any world w in the new commitment state, all worlds $O_{a,b}$-accessible from w have to satisfy π. Thus, we take the $O_{a,b}$-relations already present in the original commitment state and delete all $O_{a,b}$-relations not pointing to π-worlds (cf. Yamada, 2012, p. 189). Formally:

> **Definition 4.** (Commitment state update) Let c be a commitment state over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition. Updating c with π results in the commitment state:
>
> $$c \lceil \pi \rceil_{a,b} = \langle W^c, \{B_x^c\}_{x \in \mathcal{A}}, \{O_{x,y}^{c \lceil \pi \rceil_{a,b}}\}_{(x,y) \in \mathcal{A}^2}, I^c \rangle$$
>
> ... such that
>
> $$O_{a,b}^{c \lceil \pi \rceil_{a,b}} = O_{a,b}^c \cap \{(v,w) \in (W^c)^2 \mid c, w \models \pi\}$$
>
> ... and $O_{x,y}^{c \lceil \pi \rceil_{a,b}} = O_{x,y}^c$ for all $(x,y) \in \mathcal{A}^2$ such that $(x,y) \neq (a,b)$.

Given that commitment state updates represent the act of assertion, we can prove that $c \lceil \pi \rceil_{a,b} \models \mathbf{C}_{a,b}\, \pi$, for any commitment state c and proposition $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, provided that π is $\mathbf{C}_{a,b}$-free (see Theorem 25).[14] From this, it already follows that **SAL** is able to formally capture the performative update resulting from assertions.

Moreover, notice that $\mathbf{C}_{a,b}\, \pi$ being true globally in $c \lceil \pi \rceil_{a,b}$ (again, on the condition that π is $\mathbf{C}_{a,b}$-free) also ensures that $\mathbf{C}_{x,y}\, \mathbf{C}_{a,b}\, \pi$ and $\mathbf{B}_x\, \mathbf{C}_{a,b}\, \pi$ are true, for all agents $x, y \in \mathcal{A}$. These results, reminiscent of the Acceptance and Acknowledgment rules discussed previously, entail that we can represent the informative update 1 for assertions in **SAL** as well, without the need of any additional machinery.

Furthermore observe that Definition 4 does not always give us the informative update 2. We could easily give an instance of a commitment state c such that $c \lceil p \rceil_{a,b} \not\models \mathbf{B}_b\, p$.[15] This is precisely what we would like, given that there are contexts that avoid the informative update 2. As we discussed, two conditions need to be met for the informative update 2 to be instantiated, which we called Sincerity and Competence (Asher & Lascarides, 2003, 2008b; Lauer, 2013). Formally, we can interpret Sincerity and Competence as additional constraints on the accessibility relations in a commitment state (cf. Asher & Lascarides, 2008b):

> **Definition 5.** (Sincerity and Competence) Let c be a commitment state over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$.
>
> 1) c satisfies Sincerity (notation: c^{sin}) iff:
>
> For all $v, w \in W^c$ and $x, y \in \mathcal{A}$: $v B_x^c w$ entails $v O_{x,y}^c w$.
>
> 2) c satisfies Competence (notation: c^{comp}) iff:
>
> For all $v, w \in W^c$ and $x, y \in \mathcal{A}$: $v B_y^c w$ entails $v B_x^c w$.
>
> If c satisfies both Sincerity and Competence, we use the notation $c^{\mathsf{sin,comp}}$.

Given two agents a and b and proposition π, we can show for any commitment state c^{sin} satisfying Sincerity that $c^{\mathsf{sin}} \models \mathbf{C}_{a,b}\, \pi \to \mathbf{B}_a\, \pi$.[16] Similarly, for any commitment state c^{comp} satisfying Competence, it holds that $c^{\mathsf{comp}} \models \mathbf{B}_a\, \pi \to \mathbf{B}_b\, \pi$.[17] So, for a commitment state $c^{\mathsf{sin,comp}}$ satisfying both Sincerity and Competence, we obtain $c^{\mathsf{sin,comp}} \models \mathbf{C}_{a,b}\, \pi \to \mathbf{B}_b\, \pi$, i.e. the informative update 2 (see Theorem 26).

---

[14]A proposition π is $\mathbf{C}_{a,b}$-free if $\mathbf{C}_{a,b}$ does not occur anywhere in π. The reason for this constraint is that the result does not hold for propositions containing instances of $\mathbf{C}_{a,b}$ (for a proof, see Yamada, 2007, p. 9).

[15] For instance, take $B_b^c = W^c \times W^c$ and $I^c(p) = \emptyset$.

[16]This is different from the way Geurts (2019, p. 25) understands Sincerity, who models it as $\mathbf{C}_{a,b}\, \pi \to \neg \mathbf{C}_{a,a}\, \neg \pi$.

[17]My interpretation of Competence is not completely identical to that of Asher and Lascarides (2008b), since they represent it (roughly) as $\mathbf{B}_b\, \mathbf{B}_a\, \pi \to \mathbf{B}_b\, \pi$ (p. 24). Following Asher and Lascarides (2003, p. 398), I choose to simplify the antecedent here to make it more aligned with the formulation of Sincerity and to demonstrate in a concise manner how the two conditions together can lead to a model of the informative update 2.

For updating commitment states satisfying Sincerity, Definition 4 does not suffice. After all, if an update changes a sincere agent's commitments, it should change her beliefs accordingly. In Definition 4, however, an agent's beliefs stay unchanged in the updated state. As a result, a sincere input commitment state could become insincere when it is updated. We should thus specify a new kind of update, specific to sincere commitment states:[18]

> **Definition 6.** (Sincere commitment state update) Let c^{sin} be a commitment state over $\mathcal{L}^{\mathcal{A}}_{\text{Atom}}$ satisfying Sincerity, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\text{Atom}}$ a proposition. Updating c^{sin} with π results in the commitment state:
>
> $$c^{\text{sin}} \lceil \pi \rceil^{\text{sin}}_{a,b} = \langle W^c, \{ B_x^{c^{\text{sin}} \lceil \pi \rceil^{\text{sin}}_{a,b}} \}_{x \in \mathcal{A}}, \{ O_{x,y}^{c^{\text{sin}} \lceil \pi \rceil_{a,b}} \}_{(x,y) \in \mathcal{A}^2}, I^c \rangle$$
>
> ... such that
> $$B_a^{c^{\text{sin}} \lceil \pi \rceil^{\text{sin}}_{a,b}} = B_a^{c^{\text{sin}}} \cap O_{a,b}^{c^{\text{sin}} \lceil \pi \rceil_{a,b}}$$
>
> ... and $B_x^{c^{\text{sin}} \lceil \pi \rceil^{\text{sin}}_{a,b}} = B_x^{c^{\text{sin}}}$ for all $x \in \mathcal{A}$ such that $x \neq a$.

Of course, I here have shown only a representation of the performative and informative effect of assertions. However, given the way the system is set up, other speech act types may be treated in a similar fashion. After all, a speech act can be modeled as inducing commitments (the performative update) after which a mutual belief in these commitments automatically follows. For speech acts different from assertions, this latter effect constitutes the full informative update.

2.6 Testing the model

In the introduction, I discussed some examples of utterances made in discourse that seem to support a commitment-based analysis. I will now go over these examples once again and formally treat them in **SAL**. What I hope to show is that **SAL** generates satisfying predictions in all cases.

Let us start with a simple case. Consider scenarios of being openly insincere, such as the example based on Geurts (2017) where a teacher asserts (1) to her students during the school's yearly emergency practice. Geurts notes how nobody believes what the teacher is saying, including the teacher herself: that is, the teacher is said to be *openly insincere*. Yet, it seems perfectly possible for her to assert (1), given the context.

(1)   This is an emergency.

The felicity of (1) is rightly predicted in **SAL**, if we consider a commitment state c that does not satisfy Sincerity. Let p be the proposition expressed in (1), a denote the teacher and b the addressee (i.e. one of the students). Then, it's possible to specify c such that $c\lceil p \rceil_{a,b} \not\models \mathbf{B}_a\, p$ (see footnote 15). This causes no problems: the only change required is that a commits to p, i.e. $c\lceil p \rceil_{a,b} \models \mathbf{C}_{a,b}\, p$. But commitment does not have anything to do with belief in a commitment state that does not satisfy Sincerity, and hence (1) is correctly modeled.

Things become more interesting when we shift our attention to example (4), already referred back to in section 2.4.

---

[18]I should note that, technically, this update is a partial function, since the belief relation may not be serial anymore after the update. Conceptually, however, it should not be possible for such updates (i.e. commitments to contradictions) to occur in a sincere context. We thus need to further restrict the propositions π a sincere commitment state can get updated with. I leave this task to future work, hoping to at least have demonstrated here that speech acts performed in sincere contexts can in principle get represented in **SAL**.

(4)   a.   The buffet is open.

        b.   I believe that the buffet is open.

<div align="right">(Krifka, 2024b, p. 52)</div>

Intuitively, the plain assertion in (4a) comes over as a stronger statement than the assertion of a belief in (4b). Krifka uses this observation to argue that it is inadequate to think of assertions as expressing a belief. If this were true, after all, then (4a) and (4b) should be equally strong statements. In **SAL**, we can interpret the difference between the two assertions by studying the different results when updating a commitment state with them. Suppose that a is the author of each utterance and b the addressee. Let p be the proposition that the buffet is open. It is possible to come up with a commitment state c such that $c\lceil \mathbf{B}_a\, p \rceil_{a,b} \not\models \mathbf{C}_{a,b}\, p$.[19] At the same time, no matter what c we choose, it will always be the case that $c\lceil p \rceil_{a,b} \models \mathbf{C}_{a,b}\, p$ (Theorem 25). Thus, updating a commitment state with (4a) or with (4b) does not always lead to the same result, indicating that the two assertions are indeed different.

However, Krifka's (2024b) point is not only that (4a) and (4b) are different but moreover that (4a) is assertorically *stronger* than (4b). Even this, I claim, can be explained in **SAL**, provided we assume Sincerity. I take this to be an acceptable assumption in the context of (4), where there seems no direct reason to believe the speaker is insincere. Thus, consider any arbitrary commitment state $c^{\sin}$. Then $c^{\sin}\lceil p \rceil_{a,b}^{\sin} \models \mathbf{B}_a\, p$ (since $c^{\sin}\lceil p \rceil_{a,b}^{\sin} \models \mathbf{C}_{a,b}\, p$), and so $c^{\sin}\lceil p \rceil_{a,b}^{\sin} \models \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$. Thus, in a commitment state satisfying Sincerity, $\mathbf{C}_{a,b}\, p$ implies $\mathbf{C}_{a,b}\, \mathbf{B}_a\, p$. However, since as before we can construct $c^{\sin}$ in such a way that $c^{\sin}\lceil \mathbf{B}_a\, p \rceil_{a,b}^{\sin} \not\models \mathbf{C}_{a,b}\, p$, the reverse implication does not hold, i.e. $\mathbf{C}_{a,b}\, \mathbf{B}_a\, p$ does not imply $\mathbf{C}_{a,b}\, p$. In words, the commitment resulting from asserting (4a) implies the commitment resulting from asserting (4b), but not vice versa. This logically explains the fact that (4a) is stronger than (4b).

From this analysis, we in fact arrive at a more general interesting result:[20]

> **Theorem 27.** *Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $p \in \mathsf{Atom}$ a proposition. Moreover let $\models^{\sin} \pi$ denote that $c^{\sin} \models \pi$ for all commitment states $c^{\sin}$ satisfying Sincerity. Then:*
>
> *1)* $\models^{\sin} \mathbf{C}_{a,b}\, p \rightarrow \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$
>
> *2)* $\not\models^{\sin} \mathbf{C}_{a,b}\, \mathbf{B}_a\, p \rightarrow \mathbf{C}_{a,b}\, p$

Thus, in a sincere context, a commitment to a proposition entails a commitment to a belief in the proposition, but not vice versa. This or a similar result explaining (4) is not present in Farkas and Bruce's (2010) model or CSS, which both are not equipped with a way of expressing belief embedded under commitment. Theorem 27 also goes against Lauer's (2013) theory stating that a commitment *just is* a commitment to a belief (see also Condoravdi & Lauer, 2011), on the basis of which we concluded in section 1.2 that the theory cannot capture the difference between (4a) and (4b). This conception of commitment would moreover not explain the apparent felicity of (5), also discussed in the introduction.

(5)   I won't say the elections were rigged, but I still believe it.

"I won't say the elections were rigged" can be seen as a so-called speech act denegation. In Chapter 4, I will motivate that we should interpret denegation as an act of non-commitment. Thus, the speaker in (5) expresses a refrainment from committing to the proposition that the elections were rigged while also expressing a commitment to the belief that they were. On Lauer's (2013) approach, (5) would thus be a contradiction. **SAL** however, given Theorem 27, rightly predicts that it is not.

---

[19]Consider, for instance: $W^c = \{v, w\}$, $\mathbf{B}_a^c = \{(v,w),(w,w)\}$, $\mathbf{C}_{a,b}^c = \{(w,v),(v,v)\}$ and $I^c(p) = \{w\}$.

[20]Proofs for all theorems stated in this thesis can be found in the Appendix (A.2).

Examples such as (5) seem to stand in contrast to Moore sentences such as (9) repeated below. Whereas it is unproblematic to state that you believe in something you do not want to assert, it is unacceptable to assert a proposition you say you don't believe in.

(9)   # It is raining, but I don't believe it is raining.

On Lauer's (2013) approach, the contradiction in (9) arises more or less immediately. On our conception of propositional commitment, however, further explanation of why (9) is infelicitous is required. Theorem 27 gives us this explanation. As soon as we assume Sincerity (which is, as we have argued for, true by default), uttering "It is raining" entails a commitment to the belief that it is raining, which is incompatible with the commitment induced by saying "I don't believe it is raining". However, a pressing problem for **SAL** seems to be that (9) should *always* be predicted to be infelicitous, not only on the condition that the speaker is sincere. Consider, for instance, the actor from example (2) asserting a Moore sentence to his audience:

(17)   # I am the king of Denmark, but I don't believe I am the king of Denmark.

Such an utterance seems to make sense only in the very strange context in which the actor plays his character during the first part of the utterance and jumps out of his role in the second part. If he remains his character throughout uttering (17), however, then (17) seems unacceptable, despite the fact that the actor is clearly being insincere. What this observation perhaps calls for is a finer grained assumption on discourse contexts, one that still generates a result similar to Theorem 27 yet does not require a speaker to be sincere. Perhaps this would boil down to something like acting *as if* you are sincere, even if you are not. I will return to this issue in Chapter 5. For now, I hope to have successfully demonstrated that **SAL** can be used to properly distinguish between commitments and beliefs, and that representing certain contextual conditions (such as Sincerity) as additional constraints on our model enables **SAL** to give accurate predictions regarding situations in which the two concepts interact.

Chapter 3

Discourse developments

In the previous chapter, we discussed how speech acts change the current discourse and the context it takes place in by inducing two kinds of update. The performative update manipulates the discourse participants' commitments, the informative update their beliefs. This chapter discusses the supposition that speech acts not only change the current discourse but also influence our *expectations* on how the discourse will develop. In fact, in Chapter 4 we will see that some speech acts, such as (polar) questions and so-called speech act denegation, *only* affect the projected discourse and do not manipulate the current state of the discourse at all.

At any point in time, we might have very clear expectations on how the discourse is set to continue. These expectations not only arise from a general preference to endorse each other's commitments but also from the basic wish to not violate one's own commitments. This chapter focuses on both of these factors and brings them together in a three-level hierarchy of future speech acts: the higher up a speech act is on this hierarchy, the more expected or preferred it is predicted to be in the discourse. In section 3.1, I introduce and motivate this trichotomy. What follows from it is the second desideratum of this thesis, presented and discussed in section 3.2. In section 3.3, I propose a formalisation of the projected discourse based on the notion of a 'commitment space' (Cohen & Krifka, 2014; Krifka, 2015). This proposal extends the formal work done in Chapter 2 and leaves enough room for the integration of other speech acts to be introduced in Chapter 4.

3.1 Possibility, admissibility, cooperativity

In this section, I distinguish three ways in which a speech act might relate to future discourse expectations. According to this distinction, a speech act can be *possible*, *admissible* or *cooperative* to perform. This classification is hierarchical: any cooperative speech act is admissible, and any admissible speech act possible, but not vice versa. Although I base the distinction on the existing literature, to the best of my knowledge there is no theory yet that acknowledges all three notions explicitly as forming part of the same hierarchical structure of discourse expectations. I will discuss some theories that come close to this goal in section 3.2, where I postulate it as a second desideratum. First, I will introduce each of the three notions separately and show how they relate to one another.

3.1.1 Possibility

Before we can determine whether a speech act is expected to be performed or not, we should ask whether it is even *possible* to perform given the current state of the discourse. Usually, two conditions are identified

that would need to be fulfilled for a speech act to be possible, which are best known under the headers Non-redundancy and Consistency (e.g. Cohen & Krifka, 2014).

A speech act is said to be non-redundant when it creates new commitments compared to the (implicit or explicit) commitments already existing (Cohen & Krifka, 2014; Krifka, 2015). For instance, it is redundant to repeat an assertion or, returning to one of our previous examples (15), to assert that dogs are vertebrates after asserting that dogs are mammals. From a Gricean standpoint, being redundant goes against the communicative goal of being as informative as required for effective communication (i.e. the Maxim of Quantity; Grice, 1975), which is perhaps the reason why Non-redundancy is sometimes taken as a precondition for possibility. In Commitment Space Semantics (CSS), for instance, it is not possible to perform a speech act if this speech act does not introduce any new commitments (Cohen & Krifka, 2014; Krifka, 2015). In real life, however, it certainly seems possible to perform a speech act that has been performed before or in any other way does not introduce any new changes to the discourse. In fact, it sometimes seems quite natural to repeat yourself, e.g. to emphasise your point or to help others better understand what you are saying (cf. Rubio-Fernández, 2016). Even if we assume redundancy is to be avoided, however, this still does not take away the fact that redundant speech acts are always *possible* to perform.[21]

The second and perhaps more interesting limitation discussed in the literature is Consistency. Farkas and Bruce (2010), for instance, seem to assume that the individual commitment sets of discourse participants must always be internally consistent. A similar presumption is made in CSS, in which it is not possible to perform a speech act that results in inconsistent commitments (Cohen & Krifka, 2014). Many other (formal) theories stipulate that commitments can never be inconsistent with one another as well (e.g. Geurts, 2019; Gunlogson, 2008), or that at least 'direct' contradictions (viz. blatant inconsistencies between overt commitments) should be avoided (Kibble, 2006; Krifka, 2015). Similarly to how Non-redundancy relates to Grice's Maxim of Quantity, Consistency is represented by the Maxim of Quality: one should not say things one believes to be false (Grice, 1975), which of course in particular means that one should be consistent. A second, commitment-based argument for deprecating inconsistency that will be of importance later on is that being inconsistent causes a violation of commitments and is thus normatively dispreferred as well. However, similarly to redundant speech acts, the fact that inconsistent speech acts are undesirable (or 'inadmissible', as they will be called in what follows) does not mean that they are impossible to perform. In real life we constantly make inconsistent commitments (either intentionally or ignorantly) and hold each other accountable for doing so. Precisely because we are all aware of the possibility of violating a commitment and the social penalties this may bring is it valuable to make commitments in the first place (Vullioud et al., 2017). We thus shouldn't confuse (normative) disapproval with outright impossibility: speech acts that cause inconsistent commitments are eligible to the former, but not the latter attribute.

What I believe this brief discussion on Non-redundancy and Consistency shows is that it is quite difficult to come up with situations in which we would have to say that it is truly impossible to perform a speech act.[22] Of course, a speech act may be *unsuccessful* (e.g. declaring a marriage without the authority to do so), but similarly to redundancy or inconsistency this does not make the act impossible - the very fact that the speech act can turn out to be unsuccessful is evidence of this possibility. I will therefore assume that it is possible to perform *any* speech act at *any* time in *any* discourse.

---

[21]Note that I take a redundant speech act to still be a speech act: the fact that its illocutionary contribution does not constitute any change to the discourse context does not mean that it has no illocutionary contribution. From a commitment-based perspective, we could be more specific and say that redundant utterances still generate commitments, albeit commitments that are already in place.

[22]I am of course not referring to cases where it is *physically* impossible to perform a speech act, e.g. because your mouth is taped off or because an utterance contains words phonologically impossible to produce by the speaker.

Possibility. A speech act is always possible to perform.

That being said, what the above discussion (in particular the one on Consistency) should also highlight is that not all possible speech acts are equally expected given the prior discourse. Rather, speech acts can be ordered on the basis of how likely they are to be performed next. This is the idea that will be developed in the following two subsections.

3.1.2 Admissibility

Although a given speech act is possible to perform at any time during a discourse, performing it may result in inconsistent commitments, as we have just discussed. Inconsistency can arise on the surface level, by making an overt commitment that stands in direct contradiction to one's earlier commitments, or implicitly, as a result of the logical consequences of one's commitments. Crucially, what counts as inconsistent behaviour is context-dependent. Whether a speech act results in inconsistent commitments is determined, of course, by the prior commitments taken on by the discourse participants. The effect a speech act may have on the consistency of discourse participants' sets of commitments is captured by admissibility:

Admissibility. A speech act is admissible to perform if it does not result in inconsistent commitments for any of the discourse participants.

Note that admissibility does not necessarily require a speech act to introduce new commitments. That is, it is possible for a redundant speech act to be admissible. I thus assume that we more strongly expect each other to be consistent than to be non-redundant. This assumption deserves motivation, given that both Consistency and Non-redundancy relate to Gricean maxims, which do not appear to come in some order of importance. The reason for taking an alternative stance here is commitment-based. Since we assume that speech acts induce commitments and since commitment is a normative notion, being inconsistent constitutes a *norm violation*. By making inconsistent claims, you undertake a commitment to two courses of action that are mutually incompatible. Norm violations are of course to be avoided in a social context and may result in social penalties (Bary, 2025; Hamblin, 1970; Kibble, 2006). If you violate a commitment, you are blameworthy, and it is possible that you lose the trust of others (Vullioud et al., 2017).

In contrast, being redundant does *not* produce a violation of commitments, since by definition the normative consequences of a redundant speech act were already manifested by one's earlier utterances. Thus, although being redundant will certainly not always align with our discourse expectations or preferences, it can be argued from a normative perspective that this deviation is less severe than the deviation caused by being inconsistent, since it does not influence the speaker's social reputation or trustworthiness. To represent the assumption that redundant speech acts are more acceptable than inconsistent speech acts yet still are unexpected, I render redundant speech acts as admissible but *uncooperative* discourse moves in the next subsection.

Interestingly, the conjecture that admissibility is a strong predictor for discourse expectations because of its relation to social norms has recently received empirical support as well. In a pilot study, Yates et al. (2025) studied participants' physical reactions to linguistic norm violations (viz. making inconsistent claims) and compared them with their reactions to norm compliances (viz. making consistent claims). More specifically, using facial electromyography (fEMG) the researchers measured participants' frowning muscle activity while they witnessed either a norm violation or a norm compliance. They found significantly higher activity in the violating condition, matching the results of a similar fEMG-study in which

reactions to non-linguistic social norm violations (e.g. cutting in line) were studied (Bartholow et al., 2001). This seems to indicate that we indeed keep track of the normative commitments arising from our utterances, and that we condemn violating these commitments.

3.1.3 Cooperativity

So far, I have argued that all speech acts are possible to perform and that within this all-encompassing class of possible speech acts a distinction can be made between admissible and inadmissible speech acts. I will now go one step further, by distinguishing between different kinds of admissible speech acts. In section 2.1, we discussed how sharing commitments is a normative goal, since shared commitments are needed to achieve objectives that are of common interest (Geurts, 2017, 2019). Given the great value of shared commitments, we should have a strong desire to maximise the set of shared commitments in a conversation: the more your fellow interlocutors make the same commitments as you do, the more you can rely on their cooperation in the future. In the spirit of Grice, we could accordingly call speech acts contributing to an increase of shared commitments *cooperative* (Geurts, 2019).

> **Cooperativity.** A speech act is cooperative if it is admissible and directly contributes to increasing the set of shared commitments.

Our notion of cooperativity is highly similar to that of Grice. It is based on the idea that discourse participants aim to achieve common goals, which, in the normative framework, are realised by undertaking shared commitments (Geurts, 2019). The word 'directly' in the definition above can be interpreted as reflecting the Gricean maxim of Relevance (or Relation, as Grice (1975) himself called it). To understand what I mean exactly by the term, an example might be of good help.

Suppose somewhere in a conversation a participant a has made a commitment to p and no one else is committed to p or $\neg p$ yet. Following Kibble (2006), we could call the possible future commitment to p for the other interlocutors a *pending* commitment. What would be a cooperative course of action for some interlocutor b if she has a pending commitment to p? Evidently, this should be committing to p as well, typically achieved by asserting p. After all, when b commits to p, p becomes a shared commitment (or, when more discourse participants are part of a's and b's conversation, b's move *contributes* to p becoming a shared commitment).

Now, what if b does not take up her pending commitment to p but instead chooses to commit to some new proposition q that plays no role in any of the participants' commitments yet? This move could contribute to a shared commitment, would the other participants decide to also commit to q. However, this contribution is not a *direct* one, since it does not expand on a's already existing proposal to make p a shared commitment. In this case, we could say that the proposal to add p to the set of shared commitments comes about because a committed to p and p is a pending commitment for b. However, commitment is not necessary for making a proposal: as we will see in the next chapter, asking a question with content p (which does not immediately result in any commitments) is also interpreted as a proposal to make p a shared commitment. When a proposal is made - however this is done - the only cooperative moves are those that help realise this proposal, i.e. those that contribute to making the proposal's content a shared commitment.

The second important requirement for cooperativity is that a speech act must always be admissible. After all, inadmissible moves could in principle also contribute to an expansion of the set of shared commitments, yet they wouldn't be cooperative. For instance, if you are committed to $\neg p$ and everyone else is committed p, then if you also commit to p this proposition would become a shared commitment. However, your move could hardly be called cooperative, since you cannot live by all of your commitments

simultaneously and hence do not help anyone in coordinating their collective future actions. Alternatively, from a Gricean perspective, we could repeat that being inconsistent goes against the Maxim of Quality and so is misaligned with the Cooperative Principle (Grice, 1975).

Of course, in reality you might be persuaded to retract your commitment to p so that you can make $\neg p$ a shared commitment. Under certain circumstances, such a retraction might be the cooperative thing to do. To keep this discussion and the formalisation following from it confined, I will not consider retraction here and simply assume that commitments persist: if a commitment is made, it can never be 'unmade' (cf. Geurts, 2019). The phenomenon of retraction and what could be gained from implementing it in the model is further discussed in Chapter 5.

If there are no unanswered proposals, then any admissible speech act instigating a new proposal is deemed cooperative. Such a speech act, after all, contributes to increasing the set of shared commitments in a consistent way. This also means that redundant speech acts are never cooperative (yet are admissible, as discussed before): you can never make progress in the objective of undertaking more shared commitments if the normative consequences of your utterances are already in effect. Again, this is in line with Grice's (1975) use of the term cooperativity, given that redundancy goes against the Maxim of Quantity and hence the achievement of common communicative goals.

Similarly to admissibility, the cooperativity of a speech act is dependent on prior discourse, and not just because all cooperative speech acts are admissible. Consider assertions: although confirmation of an assertion is always a cooperative move (assuming this move is also admissible), whether a specific speech act embodies confirmation depends on the assertion's content. For instance, "Yes, he did" is a confirmation of the assertion "Ed won the race" but a denial of the assertion "Ed didn't win the race" (Krifka, 2015). Moreover, just as all admissible speech acts are possible but not all possible speech acts are admissible, all cooperative speech acts are admissible (by definition) yet there exist admissible speech acts that are not cooperative. For an example, consider again the case where b has a pending commitment to p: although b can respond to a in an admissible way both by asserting p or by asserting $\neg p$, only the former also constitutes a cooperative move.

Possibility, admissibility and cooperativity thus form a hierarchy. Although all speech acts are possible, only a subset will be admissible, given the context. From the speech acts that are admissible, only some will be cooperative. Moreover, all cooperative speech acts are admissible, and all admissible speech acts are possible. If we let $\mathbb{P}$, $\mathbb{A}$ and $\mathbb{C}$ stand for the sets of possible, admissible and cooperative speech acts respectively, then this hierarchy is represented by the inclusion relation:

$$\mathbb{C} \subseteq \mathbb{A} \subseteq \mathbb{P}$$

3.2 Desideratum 2

We now have a three-level scale with which we can capture the extent to which discourse participants expect a speech act to be performed. Importantly, performing a speech act may have an effect on the position of future speech acts on this scale. For instance, suppose that no one is committed to either p or $\neg p$ yet. If the speaker then asserts p, this changes the admissibility and cooperativity of other moves: it is now inadmissible for the speaker to assert $\neg p$ whereas it wasn't before, and it is now uncooperative for the addressee to assert $\neg p$ whereas it wasn't before. Thus, even to fully grasp the meaning of simple speech acts such as assertions, we need an account of the projected discourse, ideally one that involves the hierarchy of possibility, admissibility and cooperativity. Moreover, as we will see in the next chapter, other speech acts such as polar questions and speech act denegations seem to affect only future discourse developments. We can thus not begin to understand such speech acts if we constrain ourselves to a model

of only the current discourse state.

For these reasons, I believe it is well justified to posit the following as a second desideratum of a model of discourse:

Desideratum 2. A good model of discourse should capture a speech act's effect on both the current and the projected state of the discourse, in which a distinction is made between the possibility, admissibility and cooperativity of future speech acts.

Perhaps the most prominent theory dealing with discourse projection is that of Farkas and Bruce (2010). As discussed in section 1.3.2, their projected set represents the 'canonical' ways in which the Table can get emptied, i.e. those that constitute an expansion of the common ground. Since their common ground includes the set of shared commitments, this canonicity strongly resembles cooperativity (with the exception that Farkas and Bruce (2010) only define canonicity for reactions to speech acts, whereas cooperativity is a feature pertaining to *any* speech act). The model is able to reflect how a cooperative reaction, such as agreeing to an assertion, is more expected than an uncooperative reaction, such as disagreeing to an assertion. However, the model does not capture admissibility, since, as mentioned in the previous section, the authors assume Consistency. As a result, although the model can show how cooperative speech acts are more expected than uncooperative yet admissible speech acts, it cannot show how the latter type of speech act is still more in line with the projected discourse than inadmissible speech acts.

The second theory deserving attention in the context of this chapter is CSS. As discussed in section 1.3.1, CSS represents the projected discourse by commitment states branching out from the root commitment state (representing the current discourse), all within a commitment space. Speech acts are modeled as transitions to these branching commitment states. If a speech act does not represent a transition to any of the commitment states in the commitment space, then a retraction move is required to nevertheless model its performance. Because of this structural modelling choice, I believe that CSS does not satisfy Desideratum 2, nor that it is eligible to an interpretation on which it does. After all, we have identified three ways in which a speech act fits the prior discourse (possibility, admissibility, cooperativity) and CSS has only two ways to model a speech act performance (direct transition or retraction first). The framework thus faces a difficult decision: it either has to represent cooperativity within the commitment space and treat all uncooperative speech acts (admissible or not) as requiring a retraction, or it has to represent all admissible speech acts in the commitment space (cooperative or not) and require only a retraction for inadmissible speech acts.

In fact, both options have been explored in the work that has been done on CSS. In early versions of CSS (Cohen & Krifka, 2014; Kamali & Krifka, 2020; Krifka, 2015, 2017), only cooperative moves can be modeled directly without the use of any retraction. Thus, after making an assertion, only the commitment state is projected in which the addressee confirms the assertion (i.e. commits to its content). When the addressee decides to disagree with the speaker, a retraction is first required before this move can be made, reflecting the fact that disagreement goes against the communicative goal of the assertion. However, when some discourse participant decides to contradict herself, the same retraction is required in order for the participant's commitments to stay consistent (after all, CSS assumes Consistency, similarly to Farkas and Bruce (2010)). This version of CSS hence does not represent how inadmissible speech acts are somehow less expected and more socially costly than voicing disagreement with someone else's commitments.

Krifka (2021, 2022), perhaps for this reason, makes modifications to CSS such that retraction is now only required for (what would otherwise be) inadmissible moves. Admissible yet uncooperative moves, such as disagreeing with an assertion in a consistent way, now get projected in the commitment space, next

to cooperative moves. The unavoidable downside of this approach, of course, is that now all admissible moves stand on equal footing: both the commitment state in which the addressee agrees to an assertion and that in which she disagrees are projected from the root, not representing the fact that the former is more in line with discourse expectations than the latter. Moreover, as I will discuss in the next chapter, this version of CSS rather problematically still seems to model the cooperativity rather than admissibility of future discourse moves for some speech act types other than assertion, such as speech act denegations. It thus gives a mixed representation of notions pertaining to the projected discourse, making it unclear exactly what effect of speech acts on the projected discourse it models.

An additional aspect of CSS one can be critical of is that inadmissible discourse moves are not actually represented in the model. Since CSS assumes Consistency, it deals with commitment violation by requiring the agent who is about to instigate the violation to retract one of her previously made commitments first, such that the violation is avoided. This, I believe, gives the conceptually wrong representation of inadmissibility, since in reality violations can and sometimes do occur. Avoiding to make inconsistent commitments is not the same as actually *making* inconsistent commitments. A model of commitment withdrawal can be very valuable (as I will also argue for in Chapter 5), but it crucially captures an aspect of conversation that is different from the inadmissible moves that make up the lowest level of our three-level hierarchy.

In the next section, I will give my own formal account of the projected discourse in **SAL**, including the notions of possibility, admissibility and cooperativity. It will extend the framework I started the construction of in Chapter 2, by adopting the notion of a commitment space from CSS. However, whereas CSS fails to model all of possibility, admissibility and cooperativity, **SAL** will avoid this pitfall by formally interpreting speech acts differently: these are not transitions *within* one commitment space (as in CSS) but *between* two commitment spaces. This will allow us to exploit the commitment space structure for a representation of cooperativity without thereby losing the important contrast between admissible and inadmissible discourse moves.

3.3   Commitment spaces

In this section, I attempt to let **SAL** meet Desideratum 2 by extending it to a model of the projected discourse that accounts for the possibility, admissibility and cooperativity of future speech acts. These speech acts will for now only be simple assertions - in the next chapter, I show how the work done in this chapter can be extended to other speech acts as well.

I will express cooperativity similarly to Farkas and Bruce (2010), namely by determining the possible future expansions of the set of shared commitments on the basis of the sets of agents' individual commitments. Although this could in principle be achieved already with the formal tools introduced so far (given that we already have a way of expressing agents' sets of individual commitments), I nevertheless choose to expand **SAL**'s machinery. This is to already plan ahead for Chapter 4, where I will introduce speech acts that I believe are best modeled with the additional formal tools I here present. Thus, although the added richness is probably not strictly needed for modeling the effect of assertion on the cooperativity of future moves, it will nevertheless enable us to represent this effect.

We already have at our disposal a model of the current state of the discourse. This is what we called, similarly to CSS, a *commitment state*, introduced in section 2.2. To not only model the current discourse but its future developments as well, I also adopt from CSS the general idea of a *commitment space* - although I will strongly diverge from CSS in the application of this concept. A commitment space, as defined in CSS, is a rooted set of commitment states (though, as mentioned, commitment states in CSS are just sets of propositions, not the rich possible-world models the term refers to in **SAL**). The root,

which is a single commitment state, represents the current discourse. Other states in the commitment space, branching out from this root, represent the projected future discourse states (Cohen & Krifka, 2014).

We thus have yet another kind of accessibility relations in the framework, this time not between possible worlds but between complete commitment states (i.e. sets of possible worlds). To reflect the idea that projected discourse states embody an increase of commitments, I will denote this relation with $\sqsubset$ (Zhang, 2024). When should we say that two commitment states in a commitment space are $\sqsubset$-related? Roughly in line with early versions of CSS, I will take $\sqsubset$ to express cooperativity. Thus, for two commitment states c, c', if $c \sqsubset c'$, then c' can be achieved by performing a cooperative speech act in c. We might also say that c' is a *cooperative continuation* of c. Formally:

> **Definition 7.** (Cooperative continuations) Let c, c' be two commitment states over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$. c' is a *cooperative continuation* of c (notation: $c \sqsubset c'$) if:
>
> - $W^{c'} = W^c$
>
> - $B_a^{c'} \subseteq B_a^c$ for all $a \in \mathcal{A}$
>
> - $O_{a,b}^{c'} \subseteq O_{a,b}^c$ and $O_{a,b}^{c'} \neq \emptyset$ for all $(a, b) \in \mathcal{A}^2$ and $O_{d,e}^{c'} \subset O_{d,e}^c$ for some $(d, e) \in \mathcal{A}^2$
>
> - $I^{c'} = I^c$
>
> We write $\sqsubseteq$ to denote the improper version of $\sqsubset$, i.e. $c \sqsubseteq c'$ iff $c = c'$ or $c \sqsubset c'$.

Observe from Definition 7 that $c \sqsubset c'$ implies that c' is an admissible continuation of c but not necessarily a cooperative one. After all, c' only needs to contain new (consistent) commitments compared to c, which is in itself not enough to ensure cooperativity. However, as we discussed in section 3.1, admissibility and cooperativity only come apart after the performance of speech acts: when no commitments are pending and all issues are settled, any admissible speech act introducing new commitments is cooperative. Thus, instead of further constraining the $\sqsubset$-relation, speech act updates will be defined as updates restricting the set of $\sqsubset$-accessible commitment states to those that represent cooperative continuations. Before I turn to these updates, however, the notion of a commitment space needs to be formally introduced first (inspired by Cohen & Krifka, 2014, p. 50):

> **Definition 8.** (Commitment spaces) A *commitment space* C over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ is a set of commitment states over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ such that:
>
> $$\exists c \in C \, \forall c' \in C : c \sqsubseteq c'$$
>
> - This unique commitment state will be called the *root* of C (notation: $\sqrt{C}$).
>
> We write $C, c, w \models \phi$ to say that ϕ is true in a commitment space C, commitment state $c \in C$ and world $w \in W^c$. We write $C, c \models \phi$ to say that $C, c, w \models \phi$ for all $w \in W^c$ and $C \models \phi$ to say that $C, c \models \phi$ for all $c \in C$.
>
> We write $\models_{\mathbf{SAL}} \phi$ to say that $C \models \phi$ for all commitment spaces C.

Although truth conditions stay the same for the fragment of the language introduced so far, Definition 8 tells us that propositions should now be evaluated not only against a world and a commitment state but against the commitment space containing the commitment state as well. This is because the truth

of certain formulas not yet introduced, such as those involving the $\mathfrak{C}$-operator expressing that an action is cooperative to perform, depends on the commitment space in which it is evaluated.

In Definition 4, we spelled out the concept of updating a commitment state with a proposition from our basic language $\mathcal{L}^{\mathcal{A}}_{\text{Atom}}$, which we used to represent the effect of assertion on the current discourse. We are now ready to extend this idea to commitment spaces. Since speech acts not only affect the current but also the projected discourse, they will formally be modeled as updates not on commitment states but on commitment spaces. For this, I will make use of Propositional Dynamic Logic, where modal operators are used to define a class of actions (Fischer & Ladner, 1979). These actions are not propositions but rather define transitions from an input model to an updated one. In our case, the models will be commitment spaces and the actions available to the agents will all be speech acts (cf. Dignum & van Linder, 1997; Dignum & Weigand, 1995; Yamada, 2012). We therefore introduce a new language of actions $\mathcal{L}_{\text{Act}}$, which will be constructed in three separate steps. Since we only look at assertions in this chapter, I first introduce the fragment $\mathcal{L}^{\text{a}}_{\text{Act}}$ of $\mathcal{L}_{\text{Act}}$, standing for the language for assertion. Although I only specify $\mathcal{L}_{\text{Act}}$ in full in Definition 20, I will already refer to it in sections preceding this definition. For now, however, one could think of $\mathcal{L}_{\text{Act}}$ as only containing $\mathcal{L}^{\text{a}}_{\text{Act}}$.

> **Definition 9.** (Language for assertion) Let the basic language $\mathcal{L}^{\mathcal{A}}_{\text{Atom}}$ be given. For propositions $\pi \in \mathcal{L}^{\mathcal{A}}_{\text{Atom}}$, the language for assertion $\mathcal{L}^{\text{a}}_{\text{Act}}$ is then given by the following BNF:
> $$\alpha ::= \{\text{assert}_{a,b}(\pi)\}_{(a,b)\in\mathcal{A}^2}$$

For now then, $\mathcal{L}_{\text{Act}}$ only consists of actions of the form $\text{assert}_{a,b}(\pi)$, expressing that a asserts π to b (following Cohen & Krifka, 2014). Having a way of expressing assertions formally, we can now specify the process of updating a commitment space with this type of speech act. Since the root of the updated commitment space reflects the current state of the discourse after the assertion has been performed, the root update should mirror the commitment state update outlined in Definition 4. The rest of the updated commitment space represents the cooperative continuations of the discourse after the assertion has been performed. Given our discussion above, this should mean that all projected commitment states capture the result of confirming the assertion (since this is the only cooperative move after an assertion has been made). Thus, in the updated root the speaker should be committed to the assertion's content, and in all updated projected states the speaker *and* the addressee should be committed to that same content.

> **Definition 10.** (Assertion update) Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\text{Atom}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\text{Atom}}$ a proposition. Updating C with $\text{assert}_{a,b}(\pi)$ results in the commitment space:
> $$C[\text{assert}_{a,b}(\pi)] := \{\sqrt{C}\lceil\pi\rceil_{a,b}\} \cup \{c \in C \mid \sqrt{C}\lceil\pi\rceil_{a,b}\lceil\pi\rceil_{b,a} \sqsubseteq c\}$$

It is easy to show that $C[\text{assert}_{a,b}(\pi)]$ is well-defined, i.e. that it is indeed a commitment space, with $\sqrt{C}\lceil\pi\rceil_{a,b}$ at its root.

The basic idea of Definition 10 is that the root of the input model gets updated with the speaker's commitment to the assertion's content, and a copy of the root gets updated with both the speaker's and addressee's commitment to that content. Then, all commitment states in the input commitment space that represent either this updated copy or one of its cooperative continuations are transferred over to the new commitment space. Figure 3.1 gives a visual impression of this process.
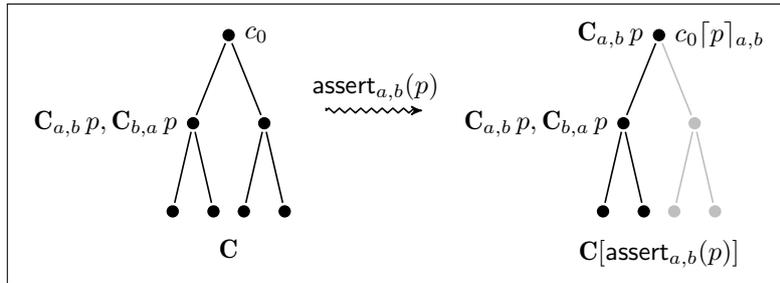
Figure 3.1: A visual representation of asserting p on the level of commitment spaces: the root c_0 of the input commitment space C gets updated with the information that the speaker a is committed towards b to p. Below this root are all commitment states in C in which both a and b are committed to p.

Note how this formal interpretation of speech acts is different from that in CSS. In CSS, the commitment space resulting from an assertion is a subset of the input commitment space. In **SAL**, however, the root of the updated commitment space does not have to be part of the input commitment space. This seemingly inconspicuous difference in the modeling of speech acts has significant consequences for **SAL**'s ability to represent all three levels of the possibility, admissibility and cooperativity hierarchy.

We now also see why $\sqsubset$ should be interpreted in terms of cooperativity: after asserting, the only commitment states $\sqsubset$-reachable from the root - that is, the only commitment states present in the commitment space excluding the root - are those in which the addressee is committed to the assertion's content. Moreover, the root of the updated commitment space is a commitment state in which the speaker (but not necessarily the addressee) is committed to the assertion's content, mirroring the commitment state update specified in Definition 4.

Now that we have a formal understanding of how speech acts realise transitions from the discourse at one point in time to the discourse at a future point in time, we could also give a tentative representation of the discourse itself. Krifka (2015) proposes to understand the full discourse as a sequence of commitment spaces $\langle C_0, C_1, ..., C_n \rangle$, succeeding each other via the performance of speech acts, i.e. $C_{m+1} = C_m[\alpha]$ for some $\alpha \in \mathcal{L}_{\mathsf{Act}}$. We could moreover require that C_0, the first commitment space in the sequence, is 'initial', meaning that no discourse participant is committed to anything yet and that all admissible discourse moves are still cooperative to perform (also see Yamada, 2012, p. 197). That is, $O_{a,b}^{\sqrt{C_0}} = W^{\sqrt{C_0}} \times W^{\sqrt{C_0}}$ for all $(a,b) \in \mathcal{A}^2$ and $C_0 = \{c' \mid \sqrt{C_0} \sqsubseteq c'\}$.

Given the prior discourse, it is possible that a speech act α does not constitute any change to the input commitment space C, i.e. $C[\alpha] = C$. This is what we can formally interpret as a *redundant* speech act, which, as we will see shortly, is by definition made uncooperative in **SAL**. More generally, we could say that a sequence of speech acts, constituting a commitment space sequence (not necessarily originating from an initial commitment space) can *entail* a speech act.

> **Definition 11.** (Speech act entailment) Let $\alpha_1, ..., \alpha_n, \beta \in \mathcal{L}_{\mathsf{Act}}$ all be speech acts. We say that $\alpha_1, ..., \alpha_n$ *entails* β (notation: $\alpha_1, ..., \alpha_n \rhd \beta$) iff $C[\alpha_1][...][\alpha_n] = C[\alpha_1][...][\alpha_n][\beta]$, for any commitment space C.

As we would expect, if a proposition τ is a logical consequence of a proposition π (for which I will write $\{\pi\} \models \tau$[23]), then asserting π entails asserting τ.

---

[23]Given a set of formulas Γ and a formula ϕ, it holds that $\Gamma \models \phi$ iff, for any commitment space C, commitment state $c \in C$ and world $w \in W^c$, $C, c, w \models \gamma$ for all $\gamma \in \Gamma$ implies $C, c, w \models \phi$.

> **Theorem 29.** *(Assertion entailment) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi, \tau \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ two $\mathbf{C}_{a,b}$-free and $\mathbf{C}_{b,a}$-free propositions such that $\{\pi\} \models \tau$. Then $\mathsf{assert}_{a,b}(\pi) \rhd \mathsf{assert}_{a,b}(\tau)$.*

To formally express the possibility, admissibility and cooperativity of speech acts, we will need to extend our basic propositional language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ to the language $\mathcal{L}_{\mathbf{SAL}}$ of **SAL**. $\mathcal{L}_{\mathbf{SAL}}$ is based on $\mathcal{L}_{\mathsf{Act}}$, which, again, can for now be taken to include assertions only.

> **Definition 12.** (Language of **SAL**) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and actions $\alpha \in \mathcal{L}_{\mathsf{Act}}$, the language $\mathcal{L}_{\mathbf{SAL}}$ of **SAL** is then given by the following BNF:
>
> $$\phi ::= \pi \mid \{V_{a,b}\}_{(a,b) \in \mathcal{A}^2} \mid \neg\phi \mid \phi \wedge \phi \mid [\alpha]\phi \mid \mathfrak{C}\alpha$$

The language of **SAL** includes three new operators, in addition to those already contained in the basic language.[24] First, propositions of the form $V_{a,b}$ express that a has violated a commitment towards b (Dignum & Weigand, 1995; Meyer, 1988), which will later on be useful in modeling inadmissibility. Propositions of the form $[\alpha]\phi$ express that ϕ is true in the commitment state updated with speech act α. Lastly, propositions of the form $\mathfrak{C}\alpha$ express that it is cooperative to perform α.

Since formulas get evaluated with respect to both a commitment space and a commitment state, we need to ensure that the commitment state in which the formula $[\alpha]\phi$ is evaluated is preserved in the updated commitment space. This is only guaranteed if this commitment state forms the root of the input commitment space. Hence, given any arbitrary commitment space C and commitment state c, we create a new commitment space C_c having c at its root as follows:

> **Definition 13.** (Commitment space restriction) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and $c \in C$ a commitment state. Then:
>
> $$C_c := \{c' \in C \mid c \sqsubseteq c'\}$$

The semantics for $[\alpha]\phi$ can then be formulated as the truth of ϕ in the root of the commitment space updated with α after restricting it to the commitment state in which $[\alpha]\phi$ is evaluated.

> **Definition 14.** (Truth after update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c, $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act and $\phi \in \mathcal{L}_{\mathbf{SAL}}$ a proposition. Then:
>
> $$C, c, w \models [\alpha]\phi \text{ iff } C_c[\alpha], \sqrt{C_c[\alpha]}, w \models \phi$$

Given that we now have a way of expressing the result of speech act updates in our language, we can formalise an important rule that was discussed before, namely that commitments can never disappear again after they are made (cf. Geurts, 2019). This rule, which I call Persistence, is rendered valid in **SAL** (also see Zhang, 2024, p. 1070):

---

[24]Note that only propositions of the basic language can function as arguments for belief, commitment or assertion (and later on for questions and speech act conditionals as well). Although this restriction is not strictly needed, it arguably makes for an easier showcase of **SAL** and its results. Another plausible reason for the restriction is that it is unclear what type of utterances, beliefs or commitments some of these unrestricted formulas (e.g. $\mathsf{assert}_{a,b}(V_{a,b})$) would represent and hence what the added value of their formal incorporation would be.

> **Theorem 30.** *(Persistence) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents, $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a $\mathbf{C}_{a,b}$-free proposition. Then:*
>
> $$\models_{\mathbf{SAL}} \mathbf{C}_{a,b}\, \pi \rightarrow [\alpha]\, \mathbf{C}_{a,b}\, \pi$$

Persistence makes the system relatively easy to work with, yet also limits its possible applications and effectiveness. I will discuss Persistence more elaborately in section 5.2.

Another new entry in $\mathcal{L}_{\mathbf{SAL}}$ is the formula $\mathfrak{C}\alpha$, expressing that α is cooperative to perform. If α is an assertion, $\mathfrak{C}\alpha$ should of course be true in a commitment space C and commitment state $c \in C$ when there is some $c' \in C$ such that $c \sqsubset c'$ and applying α to c yields c'. Given that I will introduce speech acts that preserve the root of the input commitment space in the next chapter, this would however be too strong a requirement, since then such speech acts would never be cooperative. It should therefore also be possible that c' still is a cooperative continuation in the updated commitment space. Moreover, a speech act α is only cooperative to perform in a commitment space C when it is not redundant (as argued for in section 3.1), which can be formally captured by requiring that $C[\alpha] \neq C$. All in all, we arrive at the following definition:

> **Definition 15.** (Cooperativity) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. Then:
>
> $$C, c, w \models \mathfrak{C}\alpha \quad \text{iff} \quad C_c[\alpha] \neq C_c \text{ and there is some } c' \in C$$
> $$\text{such that } c \sqsubset c' \text{ and } \sqrt{C_c[\alpha]} \sqsubseteq c'$$

Definition 15 ensures that confirmation but not denial is a cooperative move to perform after an agent has made an assertion (assuming that the move is not redundant and that the commitment state representing this move is in the input commitment space), which is precisely what we wanted.

> **Theorem 31.** *Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a $\mathbf{C}_{a,b}$-free and $\mathbf{C}_{b,a}$-free proposition. Suppose that $\sqrt{C}\lceil \pi \rceil_{a,b} \lceil \pi \rceil_{b,a} \in C$. Then:*
>
> *1) $C, \sqrt{C} \models \neg\, \mathbf{C}_{b,a}\, \pi \rightarrow [\mathsf{assert}_{a,b}(\pi)]\mathfrak{C}\, \mathsf{assert}_{b,a}(\pi)$*
>
> *2) $C, \sqrt{C} \models \neg[\mathsf{assert}_{a,b}(\pi)]\mathfrak{C}\, \mathsf{assert}_{b,a}(\neg\pi)$*

Note, however, that if the addressee decides to disagree and hence makes a commitment opposite to that of the speaker, the updated commitment space will be a singleton and so will not contain any projected commitment states. As a result, the system will wrongly conclude that all moves performed from that point in the discourse are uncooperative. I will return to this issue in Chapter 5, where I will propose incorporating an 'agree-to-disagree' move (cf. Farkas & Bruce, 2010) that can be used to settle the disagreement and generate new cooperative continuations again.

It remains for me to demonstrate how possibility and admissibility are captured in this new framework. Possibility follows from the fact that all updates (those already defined as well as those to be introduced in Chapter 4) are well-defined. Thus, performing any speech act always results in a commitment space. In other words, it holds in **SAL** that all speech acts are always possible to perform, even if these speech acts are redundant or result in inconsistent commitments (cf. Asher & Lascarides, 2008b). That is:

> **Theorem 32.** *(Possibility) Let $\alpha \in \mathcal{L}_{\mathsf{Act}}$ be a speech act. Then:*
>
> $$\models_{\mathbf{SAL}} [\alpha]\top$$

For admissibility, we use the newly added propositions of the form $V_{a,b}$, expressing that agent a has violated her commitments towards agent b (Dignum & Weigand, 1995; Meyer, 1988). Given that we interpreted commitment as a modality, an agent a having inconsistent commitments towards some agent b is simply represented by a complete absence of $O_{a,b}$-accessible worlds (notice that this was the reason for not making the $O_{a,b}$-relation serial in Chapter 2, in contrast to the B_a-relation for beliefs; see Definition 2). If there are no $O_{a,b}$-accessible worlds, a is committed towards b to *everything*. We could express this as $\mathbf{C}_{a,b} \perp$ and say that a speech act is admissible as long as it does not make this formula true. Taking the predicate $V_{a,b}$ as primitive is thus not necessary, but makes for an easier presentation of the framework.

> **Definition 16.** (Violation) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c and $a, b \in \mathcal{A}$ two agents. Then:
>
> $$C, c, w \models V_{a,b} \text{ iff } O_{a,b}^c = \emptyset$$

Note that I have defined violation on a global level, since in what follows it will not be of much importance in which world in the commitment state a violation took place. If one wishes, the definition could be made local by stating that only the set of $O_{a,b}^c$-accessible worlds from a specific world w should be empty (or, alternatively, that $c, w \models V_{a,b}$ iff $c, w \models \mathbf{C}_{a,b} \perp$).

Using our notion of violation, we can formalise admissibility as follows:

> **Definition 17.** (Admissibility) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. α is admissible w.r.t. C, c if $C, c \not\models [\alpha]V_{x,y}$, for all $x, y \in \mathcal{A}$.

This definition, together with Persistence (Theorem 30), entails that, once a violation occurs in discourse, *any* speech act thereafter is inadmissible. This is problematic, since certainly it is not the case in a typical conversation that making an inconsistent claim marks the end of the story. Ideally, we would have a way of modeling commitment retraction, similarly to CSS. Then, once a violation takes place, the violating agent should first retract one of her inconsistent commitments (this would then be the only admissible and perhaps also cooperative move to make), so that her total set of commitments becomes consistent again. I will not model retraction in this thesis, but I will discuss its possible integration in the framework more extensively in section 5.2. For present purposes, however, Definition 16 gives us what we want. For instance, it can be proven within **SAL** that making an inconsistent claim is always inadmissible:

> **Theorem 33.** *(Inadmissibility of contradictions) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition. Then:*
>
> $$\models_{\mathbf{SAL}} [\mathsf{assert}_{a,b}(\pi \wedge \neg\pi)]V_{a,b}$$

Crucially, **SAL** not only offers distinct representations of cooperativity, admissibility and possibility, but it also reflects our argument that these notions are hierarchically ordered. That is, cooperativity implies admissibility, and admissibility implies possibility, but neither implication holds the other way around as well.

Theorem 34. *Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Then for any two agents $a, b \in \mathcal{A}$ and speech act $\alpha \in \mathcal{L}_{\mathsf{Act}}$:*

1) $\models_{\mathbf{SAL}} \mathfrak{C}\alpha \to \neg[\alpha]V_{a,b}$

2) $\models_{\mathbf{SAL}} \neg[\alpha]V_{a,b} \to [\alpha]\top$

3) $\not\models_{\mathbf{SAL}} \neg[\alpha]V_{a,b} \to \mathfrak{C}\alpha$

4) $\not\models_{\mathbf{SAL}} [\alpha]\top \to \neg[\alpha]V_{a,b}$

Now that we have a formal understanding of the projected discourse and the way in which speech acts can manipulate it, we can turn to other speech act types and their integration in **SAL**. This is what I will do in the next chapter, which centers around polar questions and speech act embeddings. By discussing some natural language instances of these speech act types, Chapter 4 will also give additional support to the idea that we need both admissibility and cooperativity to be able to give an accurate representation of speech acts and the effects they bring about in the discourse.

Chapter 4

Questions and meta speech acts

4.1 Desideratum 3

Assertions are usually the typical case study for a formal theory of discourse. The current thesis forms no exception: we have focused on assertions for modeling the performative and informative update in Chapter 2 and also for dynamically capturing the projected discourse in Chapter 3. Reason for its popularity in the literature is perhaps the fact that assertions usually provide the primary method for conveying truths, knowledge and beliefs (Geurts, 2019; Goldberg, 2015; Pagin, 2014). However, a formal system only acknowledging assertions stands in obvious contrast to everyday conversations, where we use many other kinds of speech acts as well - think of questions, promises, orders, declarations, warnings, advices, greetings, exclamations, and so forth. All of these speech acts should eventually deserve a place in a formal representation of communication. It would be quite an undertaking to implement all existing types of speech acts into one framework - let alone to do this in one thesis. However, I do believe it is feasible and essential to structure one's formal framework in a way that allows for an extension to a model of these different kinds of speech acts. This will be our third and final desideratum.

> **Desideratum 3.** A good model of discourse should be generalisable to different types of speech acts.

In this chapter, I hope to demonstrate **SAL**'s generalisability by applying it to the case of polar questions and so-called 'meta speech acts'. Although these types of speech acts are conceptually quite different from one another, both strongly motivate the structure underlying **SAL**. In particular, both questions and meta speech acts show the usefulness of representing a discourse as a commitment space. Following Cohen and Krifka (2014) and Krifka (2015), I interpret polar questions and a type of meta speech act called speech act denegation as solely affecting the projected discourse while leaving the current state of the discourse untouched. Under this interpretation, these types of speech acts are different from assertion, a speech act that clearly also affects the current state of the discourse (see Chapter 2). Polar questions and speech act denegations thus show the usefulness of having a model of the projected discourse next to a model of the current discourse. An incorporation of these speech acts in the framework should therefore count as evidence for the fact that **SAL** can in theory be used as a model for many different speech act types.

It moreover follows from this interpretation of questions and speech act denegations that discourse models that do not give an extensive representation of the projected discourse consequently lack the ability to capture these types of speech acts - this includes the accounts of Asher and Lascarides (2008a) and Lauer (2013). Moreover, I will argue that the accounts that do have a way of representing the projected

discourse still fail to accurately depict the effect of polar questions and meta speech acts. In particular, both Commitment Space Semantics (CSS; Cohen & Krifka, 2014; Krifka, 2015) and the model of Farkas and Bruce (2010) lay too much emphasis on either the admissibility or cooperativity of future discourse moves. However, as I will show, we need a formal representation of both notions to give the right predictions in contexts featuring questions or meta speech acts.

Given the two types of speech acts, I will focus first on what I believe to be the most natural extension of the framework after having discussed assertions in previous chapters, namely polar questions. The formal integration of questions in the framework in section 4.3 will be preceded by an informal discussion of this type of speech act in section 4.2. I will then move on to an introduction of meta speech acts in section 4.4, which I will formalise in section 4.5. I will also show in this last section how, with the implementation of meta speech acts in the framework, some more complex question types (such as tag interrogatives or high-negation questions) can be accounted for. Throughout the chapter, I will compare my approach to that of Farkas and Bruce (2010) and most importantly to that of CSS. I will illustrate **SAL**'s higher explanatory power compared to these other approaches by investigating a number of natural language examples.

4.2 Questions

There are many types of questions, both in function and in form. Among the many categories are, for instance, *wh*-questions as in (18a), multiple alternative questions as in (18b) and polar questions, to which a simple polar particle ('yes' or 'no') suffices as an answer, as in (18c).[25]

(18)   a.   *Wh-questions*
           Who did Mary see?
    b.   *Multiple alternatives*
           Did Tom eat rice or did he eat pasta?
    c.   *Polar questions*
           Does John like books?

In this chapter, I focus on polar questions, a category even within which exists great variation, as the (non-exhaustive) list in (19) exhibits.

(19)   a.   *Default polar questions*
           Does John like books?
    b.   *Rising declaratives*
           John likes to read books?
    c.   *Reverse-polarity tag questions*
           John likes books, doesn't he?
    d.   *Same-polarity tag questions*
           John likes books, does he?
    e.   *High-negation questions*
           Doesn't John like books?

All of these different types of polar questions are claimed to have a slight difference in conversational effect

---

[25]I of course understand the term 'question' in a *pragmatic* sense, namely as the speech act of asking a question. Groenendijk and Stokhof (1994) call this the 'interrogative act', which is to be distinguished from the interpretation of questions as semantic or syntactic objects.

(see e.g. Farkas & Roelofsen, 2017; Northrup, 2014). One important distinction to be drawn between the pragmatic meanings of different types of polar questions is whether the question signals a *bias* towards one of the alternatives raised by the question. Default polar questions are often said to be non-biasing (Farkas & Bruce, 2010): the speaker signals in (19a) that she sees the two alternatives raised by the question (viz. that John likes books or that he does not like books) as equally likely.[26] In contrast, the constructions in (19b) to (19e) *do* express some kind of bias (Bolinger, 1957; Farkas & Roelofsen, 2017). For instance, when using the rising declarative in (19b) the speaker seems to signal that she believes it is more likely that John likes to read books than that he does not (Gunlogson, 2008; Šafařová, 2005).

In the commitment-based approach, a typical question can be seen as a request from the speaker to the addressee to commit to one of the alternatives by answering the question (Geurts, 2019; Gunlogson, 2008; Krifka, 2015).[27] This answer is usually given in the form of an assertion (Krifka, 2022). Thus, when asking a question the speaker hopes for an assertoric move from the addressee with one of the question's alternatives as content. After the addressee answers the question and hence commits to her answer, the speaker is expected to follow suit and make the same commitment (Farkas & Roelofsen, 2011). The speaker's commitment will hence be 'dependent' on the addressee's response (Gunlogson, 2008).

An important point of discussion here is that some scholars take the speaker to automatically commit to the answer's content after this answer gets provided by the addressee (e.g. Gunlogson, 2008; Krifka, 2015). This usually seems to be a reasonable assumption, since if someone asks a question they normally indicate that they are not certain what alternative to commit to. Accordingly, Malamud and Stephenson (2015) interpret asking a question as making a 'projected commitment', i.e. as signaling that one's commitment to any of the alternatives is a live possibility (or, in case of a biased question, to the alternative the question is biased towards). However, although it is presumably the norm that questions signal the speaker's willingness to commit to the addressee's answer, there are certainly contexts in which a question should not be interpreted this way. Consider, for instance, the following exchange between a teacher (A) and a student (B):

(20)   A: What is the square root of 25?
         B: Four.

Clearly, it would be wrong to think that the teacher is now committed to the truth of the student's answer or was prepared to make this commitment when asking the question. I am not claiming that the aforementioned theories intend to account for these rather special cases of question functions, but I do take their existence and our wish for a framework that is as general as possible as good reasons to not let the speaker's commitment to the addressee's answer follow instantly. Instead, I will interpret the speaker's commitment as a cooperative continuation. Since the addressee's answer normally comes in the form of an assertion, this approach is consistent with the idea that only confirming (but not denying) an assertion is a cooperative reaction. It is also in line with how questions are modeled in Farkas and Bruce (2010): if the addressee answers a question, this answer is first put on the Table. The speaker is then still free to decide whether to commit to the answer as well or express doubt. However, only confirmation gets represented in the projected set, reflecting the cooperativity of agreeing with the answer to the question asked.

From what we have discussed so far, we can already conclude that questions are strongly forward-looking speech acts. By requesting the addressee to commit to one of possibly multiple alternatives, they generate

---

[26]It can be argued that even default polar questions still contain a weak bias towards the alternative matching the polarity of the question (Büring & Gunlogson, 2000; Trinh, 2014; van Rooy & Šafařová, 2003). The consideration of importance here, however, is that any such bias is clearly much weaker than the ones existing in non-default polar questions.

[27]Not all questions intend for the addressee to commit to one of its alternatives - think of rhetorical or deliberative questions (Farkas & Roelofsen, 2017).

a variety of ways in which the normative status of the addressee (and subsequently that of the speaker) could change. At the same time, they leave the current commitments of both participants untouched: after asking a question, no direct change in either participant's commitments is registered.[28] While it may well be clear that questions have an effect on the development of the discourse, what remains to be discussed is *exactly what* this effect is. In the previous chapter, we have seen that there are at least two ways in which a speech act might affect the projected discourse, namely by changing the admissibility and cooperativity of future discourse moves. Assertions, we concluded, do both. Since an assertion commits its author to its content, it makes any future contradictory assertion inadmissible. Moreover, since confirming but not denying the assertion contributes directly to a new shared commitment, it makes the former but not the latter move cooperative. Given that questions, unlike assertions, do not manipulate any discourse participant's commitments directly, they cannot influence what is admissible to say next in conversation. However, I argue that, similarly to assertions, questions *do* relate to the cooperativity of future speech acts.

In the last chapter, it was stated that a speech act is cooperative only if it contributes to an increased set of shared commitments in a direct fashion. With 'direct', I meant that it should target any pending proposals for making a shared commitment, if such a proposal is present. After all, it should not be deemed cooperative to make a new assertion if one has not first addressed any outstanding assertions made by other discourse participants, even if making this new assertion is admissible and could lead to a new shared commitment. Similarly, I take it that the only cooperative move to make after being asked a question is to answer it, since only such an answering move directly addresses the questioner's proposal for creating a new shared commitment. Any other move would only possibly indirectly contribute to an increased set of shared commitments and hence not be cooperative.

For illustration, suppose a speaker asks a default polar question with content p in a discourse in which no one is committed to p or $\neg p$ yet. The addressee can answer this question by stating that p or $\neg p$, or she can assert some unrelated proposition q that does not play a role in any of the discourse participants' commitments. Both answering the question and asserting q are admissible moves to make. Furthermore, they both contribute to a new shared commitment: answering the question paves the way for a shared commitment to p or $\neg p$, while asserting q does so for a shared commitment to q. However, only the former constitutes this contribution directly, since the speaker, by asking the question, has already initiated a shared commitment to p or $\neg p$. Thus, only asserting p and asserting $\neg p$ constitute cooperative moves; asserting q does not.

The idea that only answering a question is cooperative explains, I believe, why such a move seems 'better' or more expected than ignoring the question. Consider, for instance, the following two possible responses from B after being a question asked by A:

(21)   A: Did Bell invent the telephone?
        B: Yes, Bell invented the telephone.
        B: ? I am hungry.

Clearly, B answering A's question by asserting "Yes, Bell invented the telephone" is more appropriate than making the irrelevant assertion "I am hungry". Although both responses may be admissible, only the former is also cooperative. After all, answering the question addresses A's proposal to make one of her question's alternatives a shared commitment. Broaching a new topic obviously does not further A's existing proposal and instead introduces a new one. Since cooperative moves are more expected to

---

[28]Some scholars oppose and argue that by asking a question the speaker commits to the fact that (at least) one of the alternatives is true (e.g. Beyssade & Marandin, 2006; Northrup, 2014). Since such a commitment is obviously trivial in the case of polar questions (Farkas & Roelofsen, 2011), I neglect this argument here.

be performed in conversation than uncooperative moves, this analysis explains why, intuitively, B's first assertion constitutes a more preferred response than her second assertion.

Moreover recall that, under our analysis, cooperativity is placed higher in the hierarchical order than admissibility. That is, performing an uncooperative yet admissible discourse move still fits the projected discourse better than performing an inadmissible move. Consequentially, our analysis would predict that the irrelevant assertion "I am hungry" in (21) is still a more acceptable reaction to A's question than any response that is certainly inadmissible, such as an outright contradiction. This, I believe, aligns well with our intuition. For illustration, compare B's uncooperative reaction in (21) - repeated in (22) as B's first response - with B's second, inadmissible response:

(22)   A: Did Bell invent the telephone?
         B: ? I am hungry.
         B: # Bell invented the telephone and he did not invent the telephone.

Although not addressing a question by saying "I am hungry" is certainly not the preferred reaction, it is still more reasonable than a contradictory response such as "Bell invented the telephone and he did not invent the telephone". I therefore take it that examples (21) and (22) support the hierarchical distinction between admissibility and cooperativity made in Chapter 3.

If we assume that the function of questions is to affect the cooperativity of future discourse moves, **SAL** should have no trouble to accommodate for them. After all, the system centers around the notion of a commitment space, in which $\sqsubseteq$-accessibility is interpreted as cooperativity. In the next section, I will integrate questions into commitment spaces, mirroring the way in which this is done in CSS, most notably in the work of Krifka (2015, 2017, 2021). In CSS, a question preserves the root of the commitment space and restricts the set of projected commitment states to those in which the addressee is committed to one of the alternatives. Although I take questions to restrict commitment spaces using a method similar to the one employed in CSS, the ramifications of this restriction move are *not* the same in the two frameworks, since crucially commitment spaces play a different role in **SAL** than in CSS. This has consequences for the predictions these frameworks make, in particular with regards to examples (21) and (22) discussed above. I will return to these examples at the end of next section and show how **SAL** but not CSS is able to capture the differences in acceptability between the three kinds of responses.

4.3   Questions in the commitment space

To model default polar questions in **SAL**, we should first of all extend the language of assertions $\mathcal{L}^{\mathsf{a}}_{\mathsf{Act}}$ to one in which the act of asking a polar question is expressible, bringing us one step closer to the full language of actions $\mathcal{L}_{\mathsf{Act}}$ that will be introduced in section 4.5. This extended language I will call the language for questions $\mathcal{L}^{\mathsf{q}}_{\mathsf{Act}}$, consisting of the following formulae:

> **Definition 18.** (Language for questions) Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. For propositions $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, the language for questions $\mathcal{L}^{\mathsf{q}}_{\mathsf{Act}}$ is then given by the following BNF:
>
> $$\alpha ::= \{\mathsf{assert}_{a,b}(\pi)\}_{(a,b)\in\mathcal{A}^2} \mid \{\mathsf{question}_{a,b}(\pi)\}_{(a,b)\in\mathcal{A}^2}$$

Next to actions of the form $\mathsf{assert}_{a,b}(\pi)$, the language for questions also contains actions of the form $\mathsf{question}_{a,b}(\pi)$, expressing that a asks a default polar question with content π towards b. Similarly to assertions, we can define what a commitment space updated with a question looks like. I here follow CSS,

particularly the work of Krifka (2015, 2017, 2021), where questions are taken to only affect the projected states in the commitment space and leave its root intact.

Definition 19. (Question update) Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a proposition from the basic language. Updating C with $\mathsf{question}_{a,b}(\pi)$ results in the commitment space:

$$C[\mathsf{question}_{a,b}(\pi)] := \{\sqrt{C}\} \cup$$
$$\{c \in C \mid \sqrt{C}\lceil \pi \rceil_{b,a} \sqsubseteq c\} \cup$$
$$\{c \in C \mid \sqrt{C}\lceil \neg\pi \rceil_{b,a} \sqsubseteq c'\}$$

Again, we can easily show that $C[\mathsf{question}_{a,b}(\pi)]$ is well-defined, such that it has $\sqrt{C}$ as its root. Note that, different from assertions, a commitment space updated with a question always forms a subset of the input commitment space. More specifically, a question takes the root of the commitment space and removes all continuations in which it is not the case that the addressee is committed to an alternative posed by the question (see Figure 4.1).
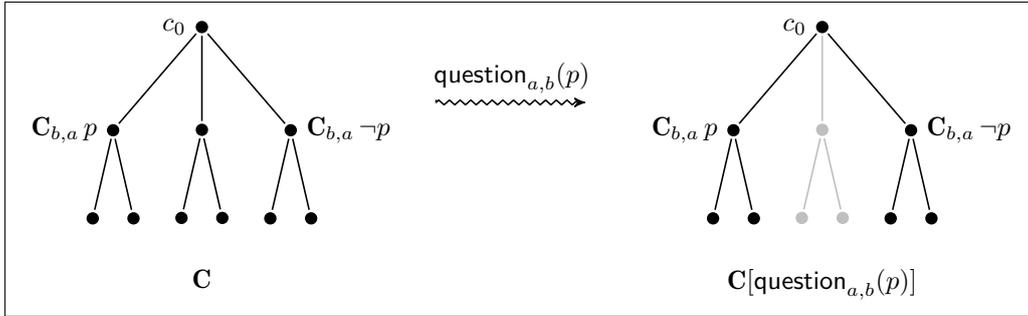

Figure 4.1:   A visual representation of asking a default polar question with content p: the root c_0 of the input commitment space C and the updated commitment space $C[\mathsf{question}_{a,b}(p)]$ are identical. Projected from the root in the updated commitment space are those commitment states in which the addressee b is either committed to p or to $\neg p$.

With Definition 19, we obtain a result that resembles a feature of inquisitive semantics (Ciardelli et al., 2013), namely that a proposition - or, in our case, an assertion of that proposition - counts as an answer to a question if it entails the question.

Theorem 35. *(Answerhood) Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a proposition. Then $\mathsf{assert}_{b,a}(\pi) \rhd \mathsf{question}_{a,b}(\pi)$ and $\mathsf{assert}_{b,a}(\neg\pi) \rhd \mathsf{question}_{a,b}(\pi)$.*

Moreover, since **SAL** (in contrast to inquisitive semantics) is a multi-agent system, it can represent the fact that it seems relevant *who* answers the question. More specifically, **SAL** produces the result that it is cooperative for the addressee to answer the question (again, on the assumption that the resulting commitment state is included in the input commitment space), but uncooperative to do so for the speaker.

> **Theorem 36.** *Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\text{Atom}}$, $c \in C$ a commitment state, $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\text{Atom}}$ a proposition. Suppose that $\sqrt{C}\lceil\pi\rceil_{b,a}, \sqrt{C}\lceil\neg\pi\rceil_{b,a} \in C$. Then:*
>
> *1)* $C, \sqrt{C} \models [\text{question}_{a,b}(\pi)](\mathfrak{C} \text{ assert}_{b,a}(\pi) \wedge \mathfrak{C} \text{ assert}_{b,a}(\neg\pi))$
>
> *2)* $C, \sqrt{C} \models (\neg \mathbf{C}_{a,b}\,\pi \wedge \neg \mathbf{C}_{a,b}\,\neg\pi) \rightarrow [\text{question}_{a,b}(\pi)]\neg(\mathfrak{C} \text{ assert}_{a,b}(\pi) \vee \mathfrak{C} \text{ assert}_{a,b}(\neg\pi))$

Although I confine myself here to a formal analysis of default polar questions, it is worthwhile to also briefly mention what a normative treatment of other types of questions could look like. Questions that raise multiple alternatives, such as (18b), could be seen as updating the commitment space in a way that creates one cooperative continuation for each of the question's alternatives. Rising declaratives, such as (19b), can be modeled as only generating one cooperative continuation, namely the commitment state supporting the question's bias (Krifka, 2015, 2017). A treatment of rising declaratives would open the door to modeling other types of non-default polar questions as well, such as reverse-polarity and same-polarity tag questions. For these question types, as well as for high-negation questions, we will require a formal account of meta speech acts, which is why I will return to the integration of other question types at the end of section 4.5.

Unfortunately, since **SAL** is a propositional logic, there is no way currently to accommodate for *wh*-questions such as (18a). An extension to first-order logic could change this fact. Next to assertion entailment and entailment between a question and an assertion (see Theorem 29 and 35, respectively), we could then also start investigating entailment between questions (e.g. the *wh*-question "Who came?" entailing the polar question "Did Mary come?"). A formal integration of *wh*-questions would moreover enable us to analyse embedded questions, which can also entail each other (e.g. "John knows who came" entails "John knows whether Mary came"). To arrive at such an analysis, however, we not only need a representation of *wh*-questions but also of attitude verbs (such as "know"). Given that we are interpreting questions on the illocutionary level, perhaps such attitude verbs should be modeled as taking speech acts (rather than propositions) as arguments. Alternatively, since commitment states are based on sets of possible worlds, attitude verbs could be modeled as inducing a set of downward closed subsets on commitment states, similarly to the approach taken by inquisitive epistemic logic (Ciardelli & Roelofsen, 2015). However, in what way such a process is realised and how these attitude verbs relate to those included in the framework already (viz. commitment and belief) are questions that will need to be addressed in future work.

Although the amount of question types **SAL** accounts for is limited, I believe the general approach it takes in modeling questions comes with significant advantages. In **SAL**, questions manipulate the set of cooperative continuations, yet do not require that the addressee's reaction fits one of these continuations. If the addressee's response is unrelated to the question, then the commitment space can nevertheless be updated with this response without any accessory predicaments. As a result, it can simulate our explanation for the differences in acceptability between B's three different responses to A's question in examples (21) and (22), combined and repeated below.

(23)   A: Did Bell invent the telephone?
       B: Yes, Bell invented the telephone.
       B: ? I am hungry.
       B: # Bell invented the telephone and he did not invent the telephone.

Answering the question (B's first response) is the best response because it is the only cooperative response of the three. This is reflected in **SAL**, where only states in which B commits to an alternative raised

by A's question are projected as cooperative continuations after updating the commitment space with that question. Second, we would like to explain why asserting some irrelevant proposition (B's second response) is more acceptable than stating a contradiction (B's third response). Although neither is projected as a cooperative continuation, **SAL** shows how the former move can nevertheless result in a consistent commitment state, while after updating a commitment space with the latter move a violation ($V_{B,A}$) always takes place. Thus, **SAL** not only represents that B's first response is cooperative and her other responses are not, but also that B's first and second response can be admissible whereas her third response is not. We can thus explain, using **SAL**, why the acceptability of B's responses seems to decrease gradually.

CSS cannot give these predictions with the same ease. Previously, we discussed how CSS cannot simultaneously capture both admissibility and cooperativity, since it models all speech acts as transitions *within* a commitment space: the performance of a speech act can only be modeled in CSS if the speech act results in one of the commitment states projected from the root. If projection represents cooperativity, as in e.g. Cohen and Krifka (2014) and Krifka (2015), then this means that uncooperative moves always require a retraction and no formal difference is made between uncooperative admissible and uncooperative inadmissible moves. If projection represents admissibility, as Krifka (2021, 2022) later opted for, then we lose the important distinction between admissibility and cooperativity. In both approaches, questions restrict the projected states to those in which the addressee commits to one of the raised alternatives. Not answering the question (e.g. asserting some irrelevant proposition) is therefore treated either identically to inadmissible speech acts (the first approach) or identically to cooperative speech acts (the second approach). Neither seems conceptually likely: we need cooperativity to explain why answering a question is preferred over not answering it, i.e. the difference between "Yes, Bell invented the telephone" and "I am hungry". Yet we also need admissibility to explain why intuitively not answering a question is still more preferred over making an inconsistent claim. Thus, neither version of CSS can account for the difference between B's responses in (21) while simultaneously also accounting for the difference between her responses in (22).

Since Farkas and Bruce (2010) do not formally consider the difference between admissible and inadmissible speech acts, they face the same issue. Although their model will be able to account for the acceptability differences between answering and ignoring a question (as in (21)), it cannot give the additional prediction that ignoring a question is more legitimate than contradicting oneself (as in (22)). **SAL** avoids these problems by not requiring that performing a speech act results in one of the projected commitment states. Rather, it is always possible for a speech act to update a commitment space, even if this speech act is uncooperative and/or inadmissible. This enables **SAL** to use commitment spaces for a model of cooperativity, without losing the important distinction between admissible and inadmissible discourse moves as a result.

4.4   Meta speech acts

SAL does not only provide fertile ground for a model of polar questions. Another phenomenon that can be elegantly captured in the current framework is the embedding of elementary speech acts under natural language connectives, resulting in a class Cohen and Krifka (2014) coin 'meta speech acts'. Traditionally, it was assumed that connectives are truth conditional and can only play a role in semantic recursion (see Hooper & Thompson, 1973; Lewis, 1972; Stenius, 1969). However, connectives arguably can sometimes also take illocutionary forces as arguments (Green, 2000; Krifka, 2001, 2002, 2014). Consider conjunction: the word 'and' can be used to concatenate two different speech acts of the same type, such as two questions in (24a) or two commands in (24b). It even seems possible to conjoin two speech acts of different types, such as an assertion with a question in (25).

(24)   a.   Which dish did Al make? And which dish did Bill make?
         b.   Eat the chicken soup! And drink the hot tea!

(Krifka, 2001, p. 13)

(25)   We invited Bill, but did he accept?

(Roelofsen & Farkas, 2015, p. 368)

Similarly, disjunction can be used on the illocutionary level: (26) shows a disjunction of two commands, (27) a command disjoined with a warning, and in (28) the disjunction appears to take scope over an assertion together with a question.

(26)   Either shut the door or open the window.

(Dummett, 1973, p. 303)

(27)   Get out of here or I will call the police.

(Krifka, 2001, p. 16)

(28)   Anne went to London, or did she go to Berlin?

Next to conjunctions and disjunctions, speech acts can appear in the consequents of conditional sentences as well, resulting in what is sometimes called a 'speech act conditional' (Sweetser, 1990; also see Van der Auwera, 1986). For instance, in (29), a famous example by Austin (1979), the consequent contains an offer towards the addressee.

(29)   There are biscuits on the sideboard, if you want them.

(Austin, 1979, p. 210)

Similarly, it seems possible to conditionalise orders and questions, as shown in (30a) and (30b), respectively.

(30)   a.   Go home, if you're feeling sick.
         b.   If I do my homework now, can I go shopping later?

(Asher, 2007, p. 214)

Next to speech act conjunction, disjunction and conditionalisation, it appears that negation can also take wide scope over illocutionary forces. This is sometimes called 'speech act denegation' (e.g. Cohen & Krifka, 2014; Searle & Vanderveken, 1985) and is perhaps the most discussed and theoretically intriguing case of meta speech acts. Whether negation takes wide or narrow scope is important for the meaning of a speech act, as observed by Searle (1969):

(31)   a.   I don't promise to come.
         b.   I promise not to come.

(Searle, 1969, p. 32)

The possibility to raise negation not only seems to apply to promises, but also to assertions (as in (32)) and questions (as in (33) or (19e) previously). Note again the differences in meaning between low and high negation in each case.

(32) a. I don't say that Bill is to be blamed.
      b. I say that Bill isn't to be blamed.

(Zhang, n.d. p. 6)

(33) a. Isn't there a vegetarian restaurant around here?
      b. Is there no vegetarian restaurant around here?

(Krifka, 2015, p. 341)

The supposition that high negation takes scope over the illocutionary force rather than the propositional content of the speech act is further evidenced by the observation that positive polarity items can occur in high-negation environments, such as 'too' in (34).

(34) Isn't John coming too?

(Asher, 2007, p. 217)

Similarly to our discussion on questions, we can theorise how meta speech acts could receive a commitment-based interpretation. Speech act conjunction seems quite straightforward: if two speech acts get conjoined, then the commitments they invoke (and project) both come into effect. In contrast, speech act disjunction seems to give the addressee a choice to bring about the commitments pertaining to either the first or the second disjunct. The semantics of speech act conditionals are perhaps already a little more difficult to understand, since in most cases the speech act presented in the consequent does not conditionally depend upon the truth of the antecedent: in "There are biscuits on the sideboard, if you want them" (example (29)), the addressee is always offered biscuits, irrespective of whether she wants them or not (Austin, 1979). An explanation for nevertheless formulating the speech act conditionally probably relates to contextual or social factors (see e.g. Dancygier, 1999; DeRose & Grandy, 1999). Leaving aside this (very interesting) pragmatic debate, we could roughly interpret the semantics of speech act conditionals as follows: if the antecedent of the conditional is true, the speech act in the consequent gets performed and its normative consequences come into effect, while if the antecedent is false, the consequent may but need not get performed.[29]

Of specific interest to this project is the meaning of speech act denegation. Cohen and Krifka (2014) analyse denegation as refraining from making a commitment in future discourse (also see Hare, 1970; Lyons, 1977): by uttering (31a), the speaker signals that she does not intend to commit to coming. This is different from (31b), where the speaker commits to not coming. Thus, importantly, whereas (31b) results in a commitment to a negated proposition, the denegation in (31a) does not result in any commitment. Rather, denegation preserves the current state of the discourse and only affects the discourse development, similarly to questions. Evidence for the fact that denegation does not induce commitments is given by the following example from Cohen and Krifka, showing that it seems acceptable to contradict a speech act denegation but not a regular speech act:

(35) a. I don't promise to come to your party, but I might do so later when I've had a look at my calendar.
      b. # I promise to come to your party, but I might say that I won't come after I've had a look in my calendar.

(Cohen & Krifka, 2014, p. 51)

---

[29]The debate on how to correctly analyse different types of speech act conditionals is ongoing. For instance, Bassi and Trinh (2024) propose to analyse conditional questions such as (30b) as questions taking scope over a conditional rather than the other way around. For reasons of brevity, I leave it for future work to consider and compare such competing analyses, and I for now simply keep to the general approach to speech act conditionals outlined above.

Since no commitment is made by performing the speech act denegation in (35a), it is permissible to announce that you might change your mind later - in our terminology, we could say that such an announcement would be *admissible*. However, the promise in (35b) *does* create a commitment, namely one for the speaker to come to the party. Later saying that you won't come to the party violates this commitment, i.e. is inadmissible. If we plausibly assume that it is unacceptable to announce during conversation that you might violate your commitments, then this explains the felicity differences between (35a) and (35b).

Since speech act denegations do not generate new commitments, they cannot influence the admissibility of future speech acts. However, speech act denegations do relate to discourse development, as we discussed: they signal the speaker's intention to not undertake a certain commitment in the future. If they do not relate to admissibility, then how do they affect discourse development? Similarly to questions, I argue that denegation manipulates the projected discourse in terms of the *cooperativity* of moves possible to perform next. That is, a denegation renders all (and only) discourse moves cooperative that do not result in the commitments that the speech act without negation would induce. Staying close to our informal definition of cooperativity in section 3.1.3, we could see denegation as a proposal to not add these commitments to the set of shared commitments. Any discourse move not supporting this proposal (i.e. any discourse move resulting in these commitments) hence does not make the 'direct' contribution to shared commitment expansion and is therefore uncooperative.

This is why I will model denegation in the next section as removing from a commitment space all those commitment states that are the result of performing the speech act that gets denegated. This approach is taken directly from CSS. However, just like questions, the fact that commitment spaces have a different function in **SAL** compared to CSS entails that this modeling choice has different consequences for the two frameworks. Specifically, I will argue that **SAL** but not CSS provides an accurate depiction of scenarios such as (35). This is what I will do in the following section, where I also formalise other types of meta speech acts and show how they relate to the formal representation of polar questions outlined in section 4.3.

4.5   Meta speech acts in the commitment space

In **SAL**, speech acts are represented as actions that can be used to update commitment spaces with. Since meta speech acts are generated by applying natural language connectives to speech acts of lower complexity, the natural approach seems to define a set of operators on our class of actions imitating these connectives.[30] With the introduction of these operators, the language of actions $\mathcal{L}_{\mathsf{Act}}$ gets extended one final time.

> **Definition 20.** (Language of actions) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, the language of actions $\mathcal{L}_{\mathsf{Act}}$ is then given by the following BNF:
>
> $$\alpha ::= \{\mathsf{assert}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2} \mid \{\mathsf{question}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2} \mid \ominus \mid \sim\beta \mid \beta; \gamma \mid \pi \hookrightarrow \beta/\gamma$$

Firstly, the empty action $\ominus$ represents the act of doing nothing, which will be needed to model speech act conditionals. The action operator $\sim$ represents speech act denegation: $\sim\beta$ expresses the denegation of β (Cohen & Krifka, 2014). Speech act conjunction is accounted for by the action composition operator (;):

---

[30]I leave it an open question how to translate natural language connectives operating on the illocutionary level to action operators in the logical language. For instance, the sentence "Come home by 5pm and we can go to the hardware store before it closes" (Lascarides & Asher, 2008, p. 16) arguably receives a conditional interpretation and hence should not be formally represented by the operator resembling speech act conjunction.

$\beta; \gamma$ means that the performance of β is succeeded by the performance of γ.[31] Finally, $\pi \hookrightarrow \beta/\gamma$ captures speech act conditionals, expressing that if π is true β is performed, else γ is performed (Meyer, 1988).

As we did with assertions and questions, we need to specify how a commitment space is updated when these different types of meta speech acts are performed. Updating with the empty action, obviously, has no effect:

> **Definition 21.** (Empty update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$. Updating C with $\ominus$ results in the commitment space:
>
> $$C[\ominus] := C$$

Updating a commitment space with a speech act denegation gives us the following (cf. Cohen & Krifka, 2014, p. 51):

> **Definition 22.** (Denegation update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. Updating C with $\sim \alpha$ results in the commitment space:
>
> $$C[\sim \alpha] := (C \setminus C[\alpha]) \cup \{\sqrt{C}\}$$

Similarly to Cohen and Krifka (2014), a speech act denegation narrows a commitment space down to the commitment states that do not show the result of performing the speech act the negation scopes over. Different from CSS, however, the root is always preserved in the updated model, ensuring that denegation update is well-defined. See Figure 4.2 for a visual representation.
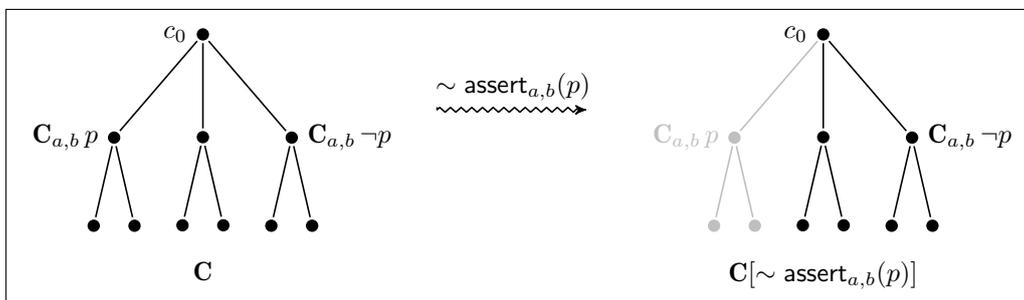

Figure 4.2: Speech act denegation visualised: after a performs the assertion denegation with content p to b, all commitment states in which the result of the assertion is registered (viz. all states in which a is committed towards b to p) are removed (except for possibly the root c_0).

Going back to our discussion on the status of speech act degenation in relation to discourse development, recall the differences in acceptability between contradicting a speech act denegation and contradicting a regular speech act:

(35)   a.   I don't promise to come to your party, but I might do so later when I've had a look at my calendar.

     b.   # I promise to come to your party, but I might say that I won't come after I've had a look in my calendar.

<div align="right">(Cohen & Krifka, 2014, p. 51)</div>

---

[31] In CSS, speech act conjunction is typically modeled as the intersection of two commitment spaces (Cohen & Krifka, 2014; Krifka, 2015). Conjunction in natural language, however, is arguably not commutative (think e.g. of the importance of conjunct order in resolving anaphoric references). I hence follow Kamali and Krifka (2020) and Krifka (2022) in defining a dynamic version of speech act conjunction.

From these examples, we concluded that the effect speech act denegation has on the projected discourse is related to cooperativity, not to admissibility. This explains why (35a), where no commitment violation is announced, is acceptable, whereas (35b), in which the speaker reveals that she might violate her commitments, is unacceptable.

I would now like to show how we can apply **SAL** to arrive at the same predictions. Since **SAL** currently has no way to represent the act of promising, I will look at a similar example involving assertions, which I claim show the same felicity results. Consider a detective (a) uttering either of the following sentences to her colleague (b):

(36)   a.   I don't say that mr. X certainly is the culprit, but I might say so later when the DNA results are in.
       b.   # I say that mr. X certainly is the culprit, but I might say that he is not the culprit after the DNA results are in.

The argument that (36b) is infelicitous because it announces the performance of an inadmissible speech act can obviously be replicated in **SAL**: saying that mr. X is the culprit (p) commits the detective to the truth of the statement ($\mathbf{C}_{a,b}\, p$). Later saying that he is not the culprit commits her to the falsity of the same statement ($\mathbf{C}_{a,b}\, \neg p$), hence causing a violation ($V_{a,b}$). However, this violation does not necessarily take place in (36a): by performing the speech act denegation $\sim \mathsf{assert}_{a,b}(p)$, the commitment space gets updated such that $\mathsf{assert}_{a,b}(p)$ cannot result in a cooperative continuation. Thus, when $\mathsf{assert}_{a,b}(p)$ nevertheless gets performed, this is an uncooperative move. However, the commitment space still gets updated. In the root of this updated commitment space, it will be true that $\mathbf{C}_{a,b}\, p$. Assuming that it was not true before that $\mathbf{C}_{a,b}\, \neg p$, this means that no violation took place. This formally reflects our explanation for the felicity of (36a).

Denegation is modeled in **SAL** in a similar way as in CSS, namely as restricting the commitment space to those commitment states in which the speech act the negation scopes over is not performed. If we take commitment spaces to represent cooperativity as Cohen and Krifka (2014) appear to do, CSS is able to show how denegation impacts the cooperativity of future discourse moves. However, since all uncooperative futures get deleted from the commitment space, counteracting a speech act denegation (as in (35a) or (36a)) requires a retraction. In fact, Krifka (2017) argues that (35a) is felicitous precisely because this possibility to retract exists. However, if announcing a possible retraction of a speech act that you make in the same move is felicitous, this does not explain why doing the same with elementary speech acts, such as in (35b) or (36b), all of a sudden is infelicitous. This problem does not get solved in Krifka (2021, 2022), who refines the CSS model so that retraction is not required any longer for disagreeing with an assertion or not answering a question as intended. This is because, unlike assertions and questions, denegations are still modeled by removing the commitment states that represent the performance of a certain speech act. Confusingly then, although Krifka (2021, 2022) uses commitment spaces to represent the *admissibility* of reactions to assertions and questions, he simultaneously uses them to represent the *cooperativity* of moves following from denegations. This also means that, although uncooperative reactions to assertions and questions may not require retraction, uncooperative moves following a denegation (such as contradicting it) still do. Krifka's (2021, 2022) more recent rendition of CSS therefore still cannot account for the difference between (35a) and (35b), or between (36a) and (36b).

The case of speech act denegation, I believe, lays bare once more the fundamental flaw of CSS, which is to treat speech act performance as always removing information from the commitment space without also having the ability to change the content of its commitment states. Since a speech act can only be performed if it realises a transition to a projected state, by choosing to project either admissibility or cooperativity CSS is forced to lose a representation of the other concept. As a result, basic felicity

differences between discourse moves such as in (35) or (36) cannot be accurately accounted for. What we need to do is treat speech acts not as changes occurring *within* a commitment space but as transformations *between* commitment spaces, which is precisely what **SAL** does.

Next to speech act denegation, we define the update following from action composition:

> **Definition 23.** (Composition update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and $\alpha, \beta \in \mathcal{L}_{\mathsf{Act}}$ two speech acts. Updating C with $\alpha; \beta$ results in the commitment space:
> $$C[\alpha; \beta] := C[\alpha][\beta]$$

Action composition is nothing more than performing the composed actions in immediate succession. Assuming that updating the commitment space with each action individually is well-defined, the eventual commitment space is well-defined as well.

Lastly, we specify conditional update as follows:

> **Definition 24.** (Conditional update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition and $\alpha, \beta \in \mathcal{L}_{\mathsf{Act}}$ two speech acts. Updating C with $\psi \hookrightarrow \alpha/\beta$ results in the commitment space:
> $$C[\pi \hookrightarrow \alpha/\beta] := \begin{cases} C[\alpha] & \text{if } C, \sqrt{C} \models \pi; \\ C[\beta] & \text{otherwise.} \end{cases}$$

This update works as expected, evaluating ψ with respect to the root of the input commitment space. Again, assuming the operations α and β both generate a commitment space, $C[\psi \hookrightarrow \alpha/\beta]$ will also result in a commitment space, making conditional update well-defined.

Usually, β will either be the empty action $\ominus$ or the speech act α contained in the consequent. Taking β to be $\ominus$ can be helpful in interpreting speech act conditionals such as (30a) repeated below, where the illocutionary act in the consequent only seems to be performed in case the antecedent is true:

(30)   a.   Go home, if you're feeling sick.

(Asher, 2007, p. 214)

For speech act conditionals in which the performance of the speech act in the consequent does not depend on the truth of the antecedent, such as (29) repeated below, an analysis where $\beta = \alpha$ seems more in place.

(29)   There are biscuits on the sideboard, if you want them.

(Austin, 1979, p. 210)

What is clearly missing in the current framework is a way of representing speech act disjunction. Disjunction could be incorporated by introducing yet another action operator $\sqcup$, expressing action choice. $\alpha \sqcup \beta$ would then mean that either α or β gets performed (cf. Meyer, 1988). The reason I do not formally implement action choice here is because it significantly increases the complexity of the framework, since choice update may result in a commitment space with multiple roots. One solution to this problem would be to say that choice update simply is not always well-defined (cf. Cohen & Krifka, 2014). However, this would go against our desire to formally allow any speech to be performed at any time. Another solution would be to redefine the notion of a commitment space, such that it can contain more than one root (as

in Kamali & Krifka, 2020; Krifka, 2021, 2022). Given the intricacy of working this out formally, I choose to leave such an extension of the framework to future work.

To close off the chapter, I would like to briefly mention the possibilities arising from the integration of meta speech acts for capturing different kinds of non-default polar questions in the framework. Krifka (2015) argues that same-polarity tag questions can be seen as a conjunction of an assertion with a rising declarative and that reverse-polarity tag questions are disjunctions of an assertion with a rising declarative containing a negation (also see Reese & Asher, 2009). Krifka (2015, 2017, 2021) moreover analyses high-negation questions as denegations of default polar questions, i.e. as requests for the addressee's non-commitment to their content. High-negation questions can thus already be modeled with the machinery introduced in this chapter. If we moreover extend the framework to rising declaratives (which I believe would not be too challenging), then same-polarity tag questions are accounted for as well.

Chapter 5

Discussion

The goal of this thesis was to develop a formal account of speech acts in dialogue, capturing their effect on the commitments, beliefs and expectations of discourse participants. This last chapter is meant to evaluate to what extent this goal has been achieved (section 5.1) and to discuss what can be done in future work to improve upon the work that this thesis presents. Section 5.2 examines two significant limitations that I believe hamper the way in which the system currently functions. In the spirit of extending the framework even further, in section 5.3 I will also briefly mention some less relevant yet interesting directions to potentially take **SAL** in. The chapter closes off with a final conclusion in section 5.4.

5.1 Returning to the desiderata

In this thesis, three desiderata for **SAL** were presented and individually motivated, based on the theoretical debate surrounding commitments and speech acts. I will here recap how **SAL** handles these desiderata in comparison to the other three approaches that were taken as a benchmark. By returning to some natural languages examples alluded to in previous chapters, I will illustrate the differences between the frameworks and motivate why I think **SAL** outperforms its competition.

Desideratum 1. A good model of discourse should capture both the performative and informative effect of speech acts.

> In Chapter 2, we discussed how speech acts update the current state of the discourse in at least two distinct ways. First, speech acts invoke commitments, giving rise to a performative update. Second, speech acts change the information states of the interlocutors: after the performance of a speech act, all interlocutors will believe that the performance took place and hence that commitments were created (the informative update 1), and in some cases they will also come to believe in the speech act's content itself (the informative update 2). The desire to separate between the performative and informative effect of speech acts was one of the main motivations to formally distinguish discourse participants' commitments from their beliefs.
>
> From the three competing commitment-based frameworks discussed throughout the thesis, only Asher and Lascarides's (2008a, 2008b) rendition of the cognitive approach seems to meet Desideratum 1. As we have seen, Lauer's (2013) Dynamic Pragmatics account only represents discourse participants' information states, leaving the performative effect of speech acts unmodeled (although Lauer does acknowledge that some commitment operator could get implemented to formally account

for the performative update as well). The same is true for some versions of Commitment Space Semantics (CSS; Cohen & Krifka, 2014; Krifka, 2015), in which commitment states fulfill the role of information states (Krifka, 2015, 2021, 2022; also see Krifka, 2024a). In other work on CSS (Cohen & Krifka, 2014; Krifka, 2017), commitment states are arguably best seen as representing the performative update instead, which has the result that the account loses a representation of the informative update. Moreover, although Farkas and Bruce's (2010) model accounts for both commitments and (background) beliefs, it mistakenly unites their content in the same set (the common ground), thereby losing the functional distinction between commitments and beliefs (Bary, 2025). Consequentially, it cannot explain why (4a) is an intuitively stronger statement than (4b). The same seems to be true for Lauer (who treats propositional commitment as having to act as if you believe in a proposition) and CSS, in which a belief operator is not formally specified.

(4)   a.   The buffet is open.
       b.   I believe that the buffet is open.

<div align="right">(Krifka, 2024b, p. 52)</div>

In **SAL**, commitments are expressible independent of mental attitudes and relate to the truth of rather than a belief in propositions. The system adopts the notion of a commitment state from CSS but interprets it as a possible-world model in which commitments and belief are defined as distinct modal operators, which is more resemblant of the work of Asher and Lascarides (2008a, 2008b) than of its use in CSS. Commitment states in **SAL** can get updated with propositions indexed to a specific speaker and addressee, after which this speaker is committed towards the addressee to that proposition. This formally represents the performative update of assertion. Since these commitments hold in each world in the commitment state, all discourse participants now necessarily also believe in these commitments, representing the first informative update. Moreover, by modeling Sincerity and Competenence as additional constraints on the accessibility relations in a commitment state, **SAL** is able to model how the discourse participants can but need not come to believe the proposition that the commitment state was updated with, representing the second informative update. If we assume Sincerity, then **SAL** is able to predict the difference in strength between (4a) and (4b), since in such a context a commitment to (4a) entails a commitment to (4b), but not vice versa.

Desideratum 2. A good model of discourse should capture a speech act's effect on both the current and the projected state of the discourse, in which a distinction is made between the possibility, admissibility and cooperativity of future speech acts.

In Chapter 3, we moved from an investigation of the effect of assertion on the current state of the discourse to discussing its influence on the expectations interlocutors hold regarding discourse development. Based on the literature, a hierarchical distinction was presented between the possibility, admissibility and cooperativity of future speech acts: whereas all speech acts are possible to perform next, only those that introduce consistent commitments are admissible, and from the set of admissible speech acts only those directly contributing to a new shared commitment are cooperative. Speech acts can manipulate the future discourse along the axes of admissibility and cooperativity: after performing a speech act, what was admissible or cooperative to do before may not be from that point on.

Both CSS and the model of Farkas and Bruce (2010) provide a dynamic representation of the projected discourse. However, neither framework succeeds in distinctively capturing the possibility, admissibility and cooperativity of speech acts. Since both frameworks assume Consistency, they lose the ability to account for the fact that even speech acts leading to inconsistent commitments are

possible to perform. Moreover, in early versions of CSS (e.g. Cohen & Krifka, 2014; Krifka, 2015, 2017), only cooperative continuations are projected. CSS is hence not able to represent how the expectation that an agent will not disagree with herself is more fundamental than the expectation that she will not disagree with some other discourse participant. Krifka's (2021, 2022) more recent version of CSS gives a way of distinguishing between future inadmissible and admissible moves, yet it introduces an equally significant problem of not being able to distinguish between future uncooperative and cooperative moves. It can thus not explain anymore, for instance, how agreeing to an assertion is preferred over disagreement.

SAL adopts the commitment space model from CSS but avoids the framework's fundamental flaw by treating speech acts as transformations rather than restrictions of commitment spaces. This way, the commitment states that stand projected in the commitment space can be used to represent the cooperative continuations without this making it impossible to distinguish between possible and admissible moves: any move resulting in a commitment space that has a consistent commitment state at its root is an admissible move, whereas all other moves are inadmissible yet still possible to perform. Not only does **SAL** represent each of the three concepts possibility, admissibility and cooperativity, it moreover represents how these concepts form an hierarchical structure: cooperativity entails admissibility, and admissibility entails possibility, but not vice versa.

Desideratum 3. A good model of discourse should be generalisable to different types of speech acts.

To attune **SAL** to the last desideratum, which was formulated in Chapter 4, we focused on the formal integration of default polar questions and different types of meta speech acts. This was not only because I take it that these speech acts demonstrate the need for **SAL**'s structure if one wants to meet Desideratum 3, but also because both types have been treated in CSS as well, thereby giving us a direct comparison between the two frameworks.

Since the formalisation of polar questions and meta speech acts seems to rely on a representation of the projected discourse, the cognitive approach already seems to be on the back foot when it comes to satisfying Desideratum 3. For the other two approaches (viz. Farkas and Bruce's (2010) model and CSS), the critical comments that apply to them regarding Desideratum 2 translate to Desideratum 3. More specifically, since Farkas and Bruce do not model the possibility of inadmissible speech acts, they cannot explain why in (23) B's (uncooperative yet admissible) first answer to A's polar question seems still more acceptable than B's (inadmissible) second answer.

(23)   A: Did Bell invent the telephone?
        B: ? I am hungry.
        B: # Bell invented the telephone and he did not invent the telephone.

Similarly, since CSS requires retraction both for the uncooperative yet admissible move of contradicting a denegation and for the inadmissible move of contradicting an elementary speech act, it cannot explain why (35a) seems acceptable but (35b) does not.

(35)   a.   I don't promise to come to your party, but I might do so later when I've had a look at my calendar.
        b.   # I promise to come to your party, but I might say that I won't come after I've had a look in my calendar.

(Cohen & Krifka, 2014, p. 51)

This is different for **SAL**: since it can hierarchically model possible, admissible and cooperative continuations of speech acts such as polar questions and denegations, it is able to predict the felicity

differences in (23), and in cases similar to (35). With regards to discourse expectations, questions and meta speech acts thus seem to be modeled better in **SAL** compared to its competitors, baring witness to the fact that **SAL** is generalisable to speech acts of varying types.

In conclusion, **SAL** appears to pose quite serious competition to existing approaches, or at least to the three popular frameworks payed special attention to in this thesis. With the exception of Desideratum 1, which also seems to be accounted for by Asher and Lascarides (2008a, 2008b), **SAL** shows more success in approximating the desiderata described, which has positive results for the number of natural discourse phenomena it is able to explain. Moreover, **SAL** is unique in that it combines all three desiderata in one system. I dare therefore say that this thesis has made a new and valuable contribution to the modeling of speech acts in dialogue.

5.2 Limitations

While I do believe **SAL** to have achieved its goals, it is not perfect. As with all formal theories, it has its shortcomings, some more serious than others. The serious ones - problems that truly limit the effectiveness of the framework and should get solved - will be discussed in the current section. In section 5.3, I will then more loosely theorise on what might be other interesting avenues worth pursuing.

What I believe to be the main limitation of the framework relates to an observation shortly touched upon in section 3.3. It was noted there that, once a violation takes place, no speech act thereafter is admissible - let alone cooperative - to perform, not even when performed by an agent that plays no role in this violation. This result is clearly undesirable - if I contradict myself once, then this should of course not mean that anything I do next is equally mischievous. Reason for this unwanted result is the following theorem, which we called Persistence:

> **Theorem 30.** *(Persistence) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents, $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a $\mathbf{C}_{a,b}$-free proposition. Then:*
> $$\models_{\mathbf{SAL}} \mathbf{C}_{a,b}\,\pi \to [\alpha]\,\mathbf{C}_{a,b}\,\pi$$

Persistence tells us that commitments, once made, will always keep existing. Thus, when an agent a makes inconsistent commitments towards some other agent b, these inconsistent commitments will never disappear again. The fact that a violated her commitments towards b ($V_{a,b}$) will remain true, which means that no subsequent speech act will be admissible anymore, given Definition 17 repeated here below.

> **Definition 17.** (Admissibility) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. α is admissible w.r.t. C, c if $C, c \not\models [\alpha]V_{x,y}$, for all $x, y \in \mathcal{A}$.

One way out of this problem would be to implement some kind of formal device with which commitments can be removed again, making Persistence true by default (as Geurts (2019) argues for) rather than by design (as in **SAL**). By giving agents the option to undo their commitments, they can resolve violations when these occur or even avoid an impending violation, preventing future speech acts from always being inadmissible. As discussed, CSS already allows for this move and suitably gives it the name of *retraction* (Krifka, 2015, 2022; see also MacFarlane, 2011).[32] We have seen how CSS uses retraction to prevent

---

[32]Retraction does not yet offer a complete solution to our problem: if I violate my commitments once but refuse to retract, it is still better from the viewpoint of the conversation to from now on make consistent (and preferably even cooperative) claims rather than producing even more violations. Even with retraction included in **SAL**, all my future moves would in this case still be predicted to be equally problematic.

agents from undertaking inconsistent commitments: by retracting a previous commitment, making a commitment that is contradictory to it is not inadmissible any longer. Since **SAL**, in contrast to CSS, allows agents to violate commitments, retraction could play a dual role in the framework: it could help prevent an agent from violating her commitments (as in CSS), and it could also be used recessively as a way to nullify a violation that has already taken place.

Motivation for formally assigning such a dual role to retraction comes from the fact that both types of commitment-retractions seem to exist in everyday life. It could be that I am aware that what I want to say next is incompatible with my previous stance and therefore first announce that I want to revise that stance. Consider the act of conceding in arguments: suppose that two discourse participants *a* and *b* are in disagreement and *a* manages to convince *b*, who subsequently chooses to yield. This seems to happen all the time in real life, witnessed by such expressions as 'I concede' or 'Okay, you may be right' (Krifka, 2022). Alternatively, it could be that someone has (perhaps unbeknowingly) violated their commitments already and is pointed to this fact by others, after which a retraction seems to be in place. It does not even need to be the case that retraction is used to prevent or repair a violation: sometimes we require each other to 'take back' a statement that we feel is unacceptable to make (Smits et al., 2022), or retracting one's commitments is simply the rational thing to do in light of new information (Heath, 2001).[33],[34]

From a formal standpoint, having some sort of retraction move would moreover open the door to a model of speech acts involving 'telic' commitments. As discussed in Chapter 1, telic commitments are commitments with an end point, brought about by, for instance, promises and orders (Geurts, 2019). Retraction could be used for the process of removing a commitment after a promise has been fulfilled or an order obeyed. Similarly, we discussed that even propositional or 'atelic' commitments, such as those arising from assertions, may not persist indefinitely. A clear case was the actor in Hamlet committing himself to the truth of the proposition that he is the king of Denmark: such a commitment should obviously be revoked once the play has ended (Geurts, 2017). Of course, retracting an ongoing commitment is conceptually different from removing commitments that have been fulfilled. Nevertheless, one all-encompassing formal device for commitment deletion could perhaps be used to represent both types of processes.

Implementing retraction in **SAL** would thus not only offer a tentative solution to the problem regarding commitment violation outlined above, but would also be of help in representing new types of discourse moves. Unfortunately however, retraction seems challenging to formalise (Yamada, 2012). Many choices concerning its mechanisms must be considered, since retraction is not an operation determined by logic alone (Hamblin, 1970; Kibble, 2006). What to do, for instance, with the commitments that the to-be-retracted commitment logically entails, or that it is logically entailed by (Hamblin, 1970)? Whereas CSS seems to assume that these commitments should be retracted as well (although this assumption stays rather implicit; Krifka, 2015, 2022), authors such as Yamada (2012) believe that they should not.

To make matters even more complicated, it is not at all clear what impact retraction should have on commitment spaces in which Sincerity and possibly also Competence is assumed. Since in sincere contexts, speakers adjust their beliefs on the basis of what they commit to, it seems obvious that somehow these adjustments should be undone when their commitments get retracted. The problem of entailment repeats itself here: exactly what beliefs should be revised given a retraction? And do we hold the same requirements for commitment retraction as for belief revision?[35]

Another important consideration given **SAL**'s aim is the relation between retraction and cooperativity. We

---

[33]We should not think too lightly of retraction either. Given that it is the norm that one keeps their commitments once made (otherwise others cannot count on you living by your commitments), retraction should of course still come at a social penalty, yet perhaps one that is less severe than the social penalty resulting from (keeping) a violation (Krifka, 2022).

[34]On a secondary yet interesting note, MacFarlane (2011, p. 84) uses the existence of retraction in everyday life as an argument against intentionalism, since it seems hard to imagine how one would 'take back' an expression of a mental attitude.

[35]For an overview on belief revision and its formalisation, see Hansson (1999).

could for instance argue that, in case of disagreement, it is cooperative to retract one of the commitments causing the disagreement. After all, if a is committed to p, b to $\neg p$ and b decides to retract her commitment to $\neg p$, then she is contributing to making p a shared commitment: whereas it was not possible before to (admissibly) have a shared commitment to p, after the retraction this becomes a live possibility again. Moreover, b makes this contribution 'directly': it addresses a's already existing proposal to add p to the set of shared commitments. Not only should we consider the cooperativity of retraction itself, we should moreover investigate what the effect of retraction is on the cooperativity of other moves. If b retracts her commitment to $\neg p$, then it seems that it should become cooperative again for b to commit to p (given that a is still committed to p). Retracting a commitment hence might make moves cooperative again that were rendered uncooperative by that commitment. Thus, whereas all speech acts formalised so far only have the ability to restrict the set of cooperative continuations, perhaps there is something to say for allowing retraction to possibly add cooperative continuations to a commitment space as well.

Although it is plausible that this makes formalising retraction even more challenging, this interpretation of retraction would come with the added benefit of solving a second problem discussed in section 3.3: not only does **SAL** currently render all future moves inadmissible after a violation takes place, it also renders all future moves uncooperative once a singleton commitment space has been reached. After all, speech acts only possibly remove cooperative continuations, and cooperativity requires that the set of cooperative continuations is non-empty (Definition 15). Moreover note that such a singleton commitment space can come about relatively easily, for instance as a result of disagreement: if a asserts p in a commitment space C, $\mathbf{C}_{a,b}\, p$ will hold in all commitment states in C. Then if b asserts $\neg p$, C gets restricted to those cooperative continuations in which $\mathbf{C}_{a,b}\, \neg p$ is true. Since C does not contain any such states, we are left with only the root.

Conceptually, one could argue that it is not a completely wrong conclusion that continuing the discourse without addressing a disagreement first is uncooperative. Rather, the issue that the disagreement causes has to be settled first. This could be achieved by retracting one of the commitments causing the disagreement (as we have just seen). A more typical approach, however, would be for the discourse participants to 'agree to disagree' (Farkas & Bruce, 2010). We could formalise this move by adding again all admissible continuations as cooperative continuations, similarly to the construction of an 'initial' commitment space talked about in section 3.3 (also see Yamada, 2012). That is, updating a commitment space C with an agree-to-disagree move AtD could be defined as the commitment space $C[\mathsf{AtD}] := \{c \mid \sqrt{C} \sqsubseteq c\}$. Similarly to retraction, agreeing to disagree would thus formally give us a way to increase the amount of cooperative continuations again, solving the problem outlined above.

The last limitation that I would like to discuss in this section concerns the application of **SAL** to Moore sentences such as (9), analysed in section 2.6.

(9)   # It is raining, but I don't believe it is raining.

The infelicity of Moore sentences can only be replicated in **SAL** when Sincerity is assumed, since in sincere (but not insincere) commitment states a commitment to a proposition entails a commitment to a belief in that proposition. This result was specified in Theorem 27:

> **Theorem 27.** *Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $p \in \mathsf{Atom}$ a proposition. Moreover let $\models^{\mathrm{sin}} \pi$ denote that $c^{\mathrm{sin}} \models \pi$ for all commitment states c^{sin} satisfying Sincerity. Then:*
>
> *1)* $\models^{\mathrm{sin}} \mathbf{C}_{a,b}\, p \rightarrow \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$
>
> *2)* $\not\models^{\mathrm{sin}} \mathbf{C}_{a,b}\, \mathbf{B}_a\, p \rightarrow \mathbf{C}_{a,b}\, p$

In section 2.6, we concluded that this forms a problem, since Moore sentences should ideally be deemed infelicitous in all contexts, not just those satisfying Sincerity. For example, the Moore construction (17) featuring the (openly) insincere actor asserting he is the king of Denmark seems equally infelicitous as (9):

(17)   # I am the king of Denmark, but I don't believe I am the king of Denmark.

One solution to this problem could be to revise our conception of Sincerity, as already hinted at in section 2.6. What examples such as (17) perhaps suggest is that Sincerity can come in layers: although the actor in (17) is certainly insincere given the larger context in which we see the actor as an actor, he is in some sense sincere within the context of the play, in which we see the actor not as an actor but as Claudius (viz. the role he is playing). As long as the play is ongoing, we expect the actor to act as if he is sincere, although everyone obviously knows that he is not. From a philosophical standpoint, works of fiction seem to create their own small reality within a much larger reality, within which the same normative conventions hold. This is somewhat resemblant of Geurts's (2017) argument that propositional commitments made during fictional play should only be revoked once the play ends and are effective until that time.

It thus seems plausible that, during the play, the audience expects the actor to act as if he believes what he asserts: if they did not have this expectation, then probably (17) would not be infelicitous at all. Note how this expectation (which mirrors Theorem 27) is only a byproduct of Sincerity and does not require that it holds. We do not expect the actor to truly believe what he asserts, only that he commits himself to this belief. Rather than assuming full or 'strong' Sincerity, many contexts including the one above seem to require that this 'weaker' form of Sincerity is at least respected. We could formalise this distinction between Strong and Weak Sincerity in **SAL** by specifying the constraints on the accessibility relations that represent them:

1. *Strong Sincerity.* $\mathbf{C}_{x,y}\,\phi$ entails $\mathbf{B}_x\,\phi$, for all agents $x, y \in \mathcal{A}$.

    A commitment state c satisfies Strong Sincerity iff:
        For all $v, w \in W^c$ and $x, y \in \mathcal{A}$: $vB_x^c w$ entails $vO_{x,y}^c w$.

2. *Weak Sincerity.* $\mathbf{C}_{x,y}\,\phi$ entails $\mathbf{C}_{x,y}\,\mathbf{B}_x\,\phi$, for all agents $x, y \in \mathcal{A}$.

    A commitment state c satisfies Weak Sincerity iff:
        For all $u, v, w \in W^c$ and $x, y \in \mathcal{A}$: $uO_{x,y}^c vB_x^c w$ entails $uO_{x,y}^c w$.

Importantly, Strong Sincerity implies Weak Sincerity (see Theorem 27) but not vice versa.[36] We could hence assume Weak Sincerity for analysing sentences such as (17) without thereby being forced to interpret the speaker as being truly sincere. Weak Sincerity seems to be easily implemented in **SAL**, although it will remain a question for future research in exactly what circumstances to assume Weak Sincerity and what the ramifications of this assumption are on the rest of the system.

5.3   Further extensions

Now that we have discussed some of the more demanding limitations of the framework, I would also like to spend some time on briefly outlining a number of exciting ways I think the framework could be further

---

[36]Consider the following countermodel: $W^c = \{w\}$, $B_a^c = \{(w,w)\}$ and $O_{a,b}^c = \emptyset$. Then c trivially satisfies Weak Sincerity, yet since $wB_a^c w$ but not $wO_{a,b}^c w$ it does not satisfy Strong Sincerity.

developed. Although I do not believe the proposals made here are essential to **SAL**'s functioning in its current form, they could be valuable suggestions to future work hoping to increase its applicability.

The most obvious next extension of **SAL** seems to be the integration of more types of speech acts. Three types that were not covered in this thesis but are discussed at quite some length in the literature and have been formalised before are promises, orders and declarations (see e.g. Dignum & Weigand, 1995; Singh, 2000). As discussed in the introduction, promises and orders appear to induce a different type of commitment than assertions and questions, namely a commitment to carrying out some action (rather than acting in accordance with the truth of some proposition). Whereas promises induce such a commitment for the speaker, orders do so for the addressee. Since **SAL** now treats commitment as an operator that takes propositional formulas as input, some modifications will have to be made. Perhaps the most feasible strategy will be to enrich the language with propositions of the form $\mathbb{S}_x\alpha$, expressing that agent x sees to it that action α is carried out (cf. Asher & Lascarides, 2003; Meyer, 1988), representing the objects of the commitments introduced by promises and orders.

Declarations, as quickly touched upon in section 2.3.1, differ from ordinary types of speech acts by primarily manipulating the truth values of propositions rather than the interlocutors' commitments: after a declaration is successfully performed, its content simply becomes true in the world. This could be represented in **SAL** by some kind of update that does not change the commitment relations (as with e.g. assertions) but instead modifies the valuation function of the commitment state. Special care is in place for the word 'successfully' here. Only when someone is authorised to make a declaration should the update come into effect. The same applies to orders: only when the one giving the order has the right to do so, should the order result in a commitment for the addressee. We will thus have to take into account what Searle and Vanderveken (1985) term the *felicity conditions* of speech acts.[37] To model them, we could for instance adopt the authorisation relation $\mathsf{auth}(x, \alpha)$ from Dignum and Weigand (1995), expressing that agent x is authorised to perform action α.

Next to extending the framework to other speech act types, more work could also be done on the formalisation of conversational implicatures and other forms of implicit content. Recall Camp's (2018) example of a driver insinuating a bribe to the police officer pulling him over:

(11)   I'm in a bit of a hurry. Is there any way we can settle this right now?

<div align="right">(Camp, 2018, p. 43)</div>

In Chapter 2 we have seen how formulating a proposition implicitly (such as the proposition that the driver is proposing a bribe) makes it possible to avoid commitment to that proposition while nevertheless creating a shared belief (Bary, 2025). This gave us a motivation for separating between commitments and beliefs in **SAL**. However, we have not stood to consider what an update of a commitment space with a speech act such as (11) would look like. Insinuations such as (11) indeed seem to require some belief update that does not affect the discourse participants' commitments. However, we should be careful in jumping to the conclusion that implicit expressions are never employable as the object of commitments. Scalar implicatures, for instance, seem to behave (at least in certain contexts) less like pragmatic inferences and more as arising from the grammar (Chemla & Spector, 2011; Chierchia et al., 2012). Unlike more typical cases of implicated meaning, scalar implicatures are not always cancelable (e.g. Magri, 2009), which was one of the main reasons presented in Chapter 2 for upholding that implicit content is non-committal (cf. Mazzarella et al., 2018). Perhaps then, one does sometimes commit to content one implies.

---

[37]Note that, if we were to implement retraction in the framework as well, then this move should also come with a (very obvious) felicity condition, stating that the commitment that is to be retracted has to exist in the first place (Meyer et al., 1999). Some authors moreover require that agents can only retract their own commitments (e.g. Krifka, 2022; Yamada, 2012), but in reality I think it is possible to retract commitments of others as well, e.g. by withdrawing an order.

Not only does implicit meaning, by affecting beliefs and possibly even commitments, relate to the future of the discourse. Not-at-issue content, for instance, seems to refer to what has previously been established as well. We discussed how "The king of the Netherlands is at a ceremony" presupposes that it is common ground who "the king of the Netherlands" refers to. For other presuppositions, however, it seems that mutual belief is not enough. Saying "My shirt is red, too" comes over odd when no one has asserted, say, that fire trucks are red, even when this is commonly believed. Perhaps, then, such presuppositions imply that someone is committed to their content. Formally, one idea could be to model speech acts containing presuppositions or other not-at-issue content as coming with a felicity condition (viz. that the presupposed information is mutually believed or that it is committed to), similarly to how we might introduce felicity conditions for orders and declarations discussed above.

Moreover, implicit meaning can function as a reaction to a previously performed speech act. A question, for instance, can get answered implicitly:

(37)   A: Do you think the student is a suitable candidate?
        B: The student has nice handwriting.

What observations like (37) seem to suggest is that cooperativity is not only understood in terms of undertaking commitments. In (37), B seems to answer A's question with the use of a conversational implicature (viz. that the student is not a suitable candidate). Given that this conversational implicature is arguably non-committal (as discussed in section 2.1), this would mean that one can give an answer to a question (and hence be cooperative) without needing to commit to that answer. Much more work would of course have to be done to capture the behaviour of implicit content and its relation to discourse participants' attitudes formally. Be that as it may, I believe that **SAL**, providing a dynamic model not only of agents' commitments but their beliefs and expectations as well, offers a strong foundation on which this work can be realised.

The last extension that I would like to suggest here is one relating to the notion of commitment itself. In this thesis, as in most of the literature, commitment is regarded as a binary notion: a commitment is something you either have or do not have. One could argue, however, that this is an oversimplification. It might be the case that, in real life, someone can be less or more committed to a certain course of action (Boulat & Maillat, 2017; Cornillie, 2018; Katriel & Dascal, 1989; Kissine, 2008). This would also mean that the normative consequences of undertaking a commitment are sensitive to its height: the greater the commitment, the higher the social risk that comes with undertaking it (Krifka, 2024b; Vullioud et al., 2017).

While this conception of gradable commitments may seem novel, it quite strongly relates to the more established idea that assertability comes in degrees (Jackson, 1979; Lewis, 1981). Think, for instance, of the function of adverbials such as 'truly', 'certainly', 'apparently' and 'frankly'. By using them in an assertion, speakers seem to be able to modulate the commitments that result from performing the assertion (Bary & Maier, 2021; Elder, 2021; Faller, 2019; Krifka, 2014, 2023, 2024b). This theory entertains empirical support as well. In a second fEMG study, Yates et al. (forthcoming) measured the frowning muscle activity of participants presented with inconsistent claims featuring a reportative evidential, such as (38b) below. Results show that this activity, with which the authors stipulate norm violations can be measured (Yates et al., 2025, see also section 3.1.2), lies between the measured activity in the condition where no inconsistency occurs (as in (38a)) and the condition in which a plain inconsistent claim is made (as in (38c)).

(38)   a.   Tom is a bachelor. He is single.
        b.   I heard that Tom is a bachelor. But he is married.

      c.   Tom is a bachelor. But he is married.

<div align="right">(Yates et al., forthcoming)</div>

If increased frowning muscle activity can indeed be used as an indication of perceived commitment violation, then these results seem to suggest that it is possible to violate a commitment to a lower or higher degree, which could find its explanation in a gradable notion of commitments.

Not only assertions seem to involve commitments of various degrees. Farkas and Bruce (2010) entertain the possibility that the addressee makes a stronger commitment to a speaker's assertion if she explicitly confirms or denies it (e.g. by responding with 'yes' or 'no') than when she merely acknowledges the speaker's move (think of staying silent, nodding, saying 'oh', 'hmmm' and the like; also see Bary, 2025; Clark, 1996; Gunlogson, 2008; Heritage, 1985; Kibble, 2006; Krifka, 2015, 2022). Moreover, Moeschler (2013) hypothesises that one is more committed to meaning conveyed explicitly than to meaning only conveyed implicitly (see Mazzarella et al. (2018) for empirical support). Additionally, Northrup (2014) believes that biased questions entail a weak commitment for the questioner to the biased alternative, which can be turned into a 'strong' commitment upon ratification from the addressee.

Given the many mentions of gradable commitments in different places in the literature, the consequences of allowing commitments to come in degrees may be very interesting to study more thoroughly in a formal setting. Modeling degrees of commitments is most obviously achieved by resorting to probabilistic logic, which already finds many applications in formal epistemology (see e.g. Titelbaum, 2022). It is not straightforward how **SAL** could be combined with a probabilistic approach, and there will certainly be different ways to achieve this quite ambitious goal. I leave this challenging yet highly interesting expedition, then, to future work.

5.4 Conclusion

This thesis presented a formalisation of the commitment-based approach on speech acts. Specifically, it presented a Speech Act Logic **SAL** that centered around three distinct effects of speech acts arising from the assumption that speech acts generate commitments. First and foremost, since commitments are social obligations, speech acts manipulate the *normative* attitudes of discourse participants (the performative update). Second, they relate to *belief*: discourse participants normally will believe that commitments are made when a speech act is performed (the informative update 1), and, under the additional assumption that the speaker is sincere and competent, they moreover come to believe in the commitment's content (the informative update 2). Lastly, the commitments invoked by speech acts affect interlocutors' *expectations* regarding the future of the dialogue, inducing an hierarchical order on future discourse moves. The bulk of the thesis has focused on modeling these three effects for assertion, which is widely regarded in the philosophy of language as the most basic yet important type of speech act. In Chapter 4, however, the formal framework that needed to be set up for assertions was applied and extended to default polar questions and different kinds of embedded speech acts, also called 'meta speech acts' (Cohen & Krifka, 2014). Both questions and meta speech acts show the need for modeling not only the current but also the projected state of the discourse.

To this end, **SAL** draws inspiration from Commitment Space Semantics (CSS; Cohen & Krifka, 2014; Krifka, 2015) by interpreting the discourse as a commitment space. In the basis, a commitment space is made up of a tree, its root representing the current state of the discourse and its branches the states that are projected. Speech acts are formally defined as updates on these commitment spaces: they not only change what the discourse participants are committed to and what information they have available to them, but also how they expect the discourse to develop. Different from CSS, **SAL** captures the fact

that these expectations are categorised in order of importance: whereas it is most expected that someone will not violate any of their own commitments (what we have called admissibility), a less pressing yet still constitutive expectation is that future speech acts address what has been said before by contributing to new shared commitments (what we have called cooperativity). **SAL** is able to model both dimensions while preserving the fact that all speech acts, whether expected or not, are possible to perform. As a result, the system is able to give accurate predictions regarding the felicity of different reactive discourse moves.

SAL offers a model that accounts for discourse participants' commitments, their beliefs as well as their expectations regarding the future of the discourse, all with respect to a multitude of speech act types. As a result, **SAL** constitutes a valuable addition to the formal work on commitments and speech acts. Whereas previous theories limit themselves to the relation either between commitments and beliefs or between commitments and expectations, **SAL** unites all three concepts in one model of discourse. Moreover, existing formalisation attempts arguably fail to fully capture the concepts they are designed for. Most prominently, I have argued that CSS is lacking the means to make a proper distinction between possibility, admissibility and cooperativity, thereby not being able to explain some natural language examples that **SAL** does account for. With the fundaments of the framework covered in this thesis, **SAL** can be extended further in a number of valuable ways. Most notably, incorporating a 'retraction' move allowing agents to undo the commitments they have made would contribute to solving some undesirable results and increasing the range of discourse phenomena that **SAL** is able to capture. The result of the work done in this thesis is a broadly applicable model of discourse that is apt to explain many different conversational phenomena and that allows for exciting extensions covering an even wider variety of discourse moves.

Bibliography

Alston, W. P. (2000). *Illocutionary acts and sentence meaning.* Cornell University Press.

Asher, N. (2007). Dynamic discourse semantics for embedded speech acts. In S. L. Tsohatzidis (Ed.), *John Searle's philosophy of language: Force, meaning and mind* (pp. 211–243). Cambridge University Press.

Asher, N., & Lascarides, A. (2003). *Logics of conversation.* Cambridge University Press.

Asher, N., & Lascarides, A. (2008a). Commitments, beliefs and intentions in dialogue. *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (Londial)*, 35–42. https://www.research.ed.ac.uk/en/publications/commitments-beliefs-and-intentions-in-dialogue/

Asher, N., & Lascarides, A. (2008b). Making the right commitments in dialogue. *Fall 2008 Workshop in Philosophy and Linguistics.*

Austin, J. L. (1975). *How to do things with words.* Clarendon Press.

Austin, J. L. (1979). Ifs and cans. In J. O. Urmson & G. Warnock (Eds.), *Philosophical papers* (3rd ed., pp. 205–232). Oxford University Press.

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts.* MIT press.

Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, *12*(3), 197–204. https://doi.org/10.1111/1467-9280.00336

Bary, C. (2025). Speech acts, common ground and commitments. *Linguistics and Philosophy*, *48*(3), 505–526. https://doi.org/10.1007/s10988-025-09434-y

Bary, C., & Maier, E. (2021). The landscape of speech reporting. *Semantics and Pragmatics*, *14*(8), 1–54. https://doi.org/10.3765/sp.14.8

Bassi, I., & Trinh, T. (2024). Conditional Questions as Matrix Questions with syntactic reconstruction [Presentation at IATL workshop, Tel Aviv University].

Beyssade, C., & Marandin, J.-M. (2006). The speech act assignment problem revisited: Disentangling speaker's commitment from speaker's call on addressee. *Empirical issues in syntax and semantics*, *6*(37-68). https://shs.hal.science/halshs-01450299v1

Bolinger, D. L. M. (1957). Interrogative structures of American English: The direct question. *American Dialect Society N° 28.*

Boulat, K., & Maillat, D. (2017). She said you said I saw it with my own eyes: A pragmatic account of commitment. In J. Blochowiak, C. Grisot, S. Durrleman, & C. Laenzlinger (Eds.), *Formal models in the study of language: Applications in interdisciplinary contexts* (pp. 261–279). Springer International Publishing.

Brandom, R. (1983). Asserting. *Noûs*, *17*(4), 637–650. https://doi.org/10.2307/2215086

Büring, D., & Gunlogson, C. (2000). Aren't positive and negative polar questions the same? http://hdl.handle.net/1802/1432

Camp, E. (2018). Insinuation, common ground, and the conversational record. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New work on speech acts* (pp. 40–66). Oxford University Press.

Chemla, E., & Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of Semantics*, *28*(3), 359–400. https://doi.org/10.1093/jos/ffq023

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In C. Maienborn, K. v. Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 2297–2332, Vol. 3). De Gruyter Mouton.

Ciardelli, I. A., Groenendijk, J., & Roelofsen, F. (2013). Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass*, *7*(9), 459–476. https://doi.org/10.1111/lnc3.12037

Ciardelli, I. A., & Roelofsen, F. (2015). Inquisitive dynamic epistemic logic. *Synthese*, *192*(6), 1643–1687. https://doi.org/10.1007/s11229-014-0404-7

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Cohen, A., & Krifka, M. (2014). Superlative quantifiers and meta-speech acts. *Linguistics and Philosophy*, *37*(1), 41–90. https://doi.org/10.1007/s10988-014-9144-x

Cohen, P. R., & Levesque, H. J. (1990). Persistence, intention and commitment. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 33–69). MIT Press.

Condoravdi, C., & Lauer, S. (2011). Performative verbs and performative acts. *Proceedings of Sinn und Bedeutung*, *15*, 149–164. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/369

Cornillie, B. (2018). On speaker commitment and speaker involvement. Evidence from evidentials in Spanish talk-in-interaction. *Journal of Pragmatics*, *128*, 161–170. https://doi.org/10.1016/j.pragma.2017.11.014

Dancygier, B. (1999). *Conditionals and prediction: Time, knowledge and causation in conditional constructions*. Cambridge University Press.

DeRose, K., & Grandy, R. E. (1999). Conditional assertions and "biscuit" conditionals. *Noûs*, *33*(3), 405–420. https://www.jstor.org/stable/2671993

Dignum, F., & van Linder, B. (1997). Modelling social agents: Communication as action. *Intelligent Agents III. Agent Theories, Architectures, and Languages*, 205–218. https://doi.org/10.1007/BFb0013587

Dignum, F., & Weigand, H. (1995). Communication and deontic logic. In R. Wieringa & R. B. Feenstra (Eds.), *Information systems, correctness and reusability* (pp. 242–260). World Scientific.

Dummett, M. (1973). *Frege: Philosophy of language*. Duckworth.

Elder, C.-H. (2021). Speaker meaning, commitment and accountability. In D. Z. Kádár, M. Terkourafi, & M. Haugh (Eds.), *The cambridge handbook of sociopragmatics* (pp. 48–68). Cambridge University Press.

Ettinger, A., & Malamud, S. A. (2015). Mandarin utterance-final particle *ba* (吧) in the conversational scoreboard. *Proceedings of Sinn und Bedeutung*, *19*, 232–251. https://doi.org/10.18148/sub/2015.v19i0.840

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. MIT Press.

Faller, M. T. (2019). The discourse commitments of illocutionary reportatives. *Semantics and Pragmatics*, *12*, 8:1–53. https://doi.org/10.3765/sp.12.8

Farkas, D. F. (2010). The grammar of polarity particles in Romanian. In A. M. Di Sciullo & V. Hill (Eds.), *Edges, heads, and projections: Interface properties* (pp. 87–124). John Benjamins Publishing Company.

Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, *27*(1), 81–118. https://doi.org/10.1093/jos/ffp010

Farkas, D. F., & Roelofsen, F. (2011). *Polar initiatives and polarity particle responses in an inquisitive discourse model*.

Farkas, D. F., & Roelofsen, F. (2017). Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics*, *34*(2), 237–289. https://doi.org/10.1093/jos/ffw012

Fischer, M. J., & Ladner, R. E. (1979). Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, *18*(2), 194–211. https://doi.org/10.1016/0022-0000(79)90046-1

Gaudou, B., Herzig, A., & Longin, D. (2006). Grounding and the expression of belief. *Principles of Knowledge Representation and Reasoning: Proceedings of the Tenth International Conference (KR2006)*, 221–229. https://aaai.org/papers/kr06-024-grounding-and-the-expression-of-belief/

Geurts, B. (2017). Fictional commitments. *Theoretical Linguistics*, *43*(1-2), 53–60. https://doi.org/10.1515/tl-2017-0003

Geurts, B. (2018). Making sense of self talk. *Review of Philosophy and Psychology*, *9*(2), 271–285. https://doi.org/10.1007/s13164-017-0375-y

Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*(1-2), 1–30. https://doi.org/10.1515/tl-2019-0001

Geurts, B. (2024). Common ground in pragmatics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2024). Metaphysics Research Lab, Stanford University.

Goldberg, S. (2015). *Assertion: On the philosophical significance of assertoric speech*. Oxford University Press.

Green, M. S. (2000). Illocutionary force and semantic content. *Linguistics and Philosophy*, *23*(5), 435–473. https://doi.org/10.1023/A:1005642421177

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Brill.

Groenendijk, J., & Stokhof, M. (1994). Questions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 1055–1124). Elsevier.

Gunlogson, C. (2001). *True to form: Rising and falling declaratives as questions in English* [Doctoral dissertation]. UC Santa Cruz.

Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics*, *22*(1), 101–136. https://doi.org/10.1075/bjl.22.06gun

Hamblin, C. L. (1971). Mathematical models of dialogue. *Theoria*, *37*(2), 130–155. https://doi.org/10.1111/j.1755-2567.1971.tb00065.x

Hamblin, C. L. (1970). *Fallacies*. Methuen.

Hansson, S. O. (1999). *A textbook of belief dynamics: Theory change and database updating*. Kluwer Academic Publishers.

Hare, R. M. (1970). Meaning and speech acts. *The Philosophical Review*, *79*(1), 3–24. https://doi.org/10.2307/2184066

Harris, D. W. (2019). Intention and commitment in speech acts. *Theoretical Linguistics*, *45*(1-2), 53–67. https://doi.org/10.1515/tl-2019-0004

Heath, J. (2001). *Communicative action and rational choice*. MIT Press.

Heritage, J. (1985). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson (Ed.), *Structures of social action* (pp. 299–345). Cambridge University Press.

Hogeweg, L., Ramachers, S. T. M. R., & Wottrich, V. (2011). *Doch, toch* and *wel* on the table. In R. Nouwen & M. Elenbaas (Eds.), *Linguistics in the Netherlands 2011* (pp. 50–60). John Benjamins.

Hooper, J. B., & Thompson, S. A. (1973). On the applicability of root transformations. *Linguistic Inquiry*, *4*(4), 465–497. https://www.jstor.org/stable/4177789

Jackson, F. (1979). On assertion and indicative conditionals. *The Philosophical Review*, *88*(4), 565–589. https://doi.org/10.2307/2184845

Jary, M. (2007). Are explicit performatives assertions? *Linguistics and Philosophy*, *30*(2), 207–234. https://doi.org/10.1007/s10988-007-9015-9

Kamali, B., & Krifka, M. (2020). Focus and contrastive topic in questions and answers, with particular reference to Turkish. *Theoretical Linguistics*, *46*(1-2), 1–71. https://doi.org/10.1515/tl-2020-0001

Katriel, T., & Dascal, M. (1989). Speaker's commitment and involvement in discourse. In Y. Tobin (Ed.), *From sign to text: A semiotic view of communication* (pp. 275–296). John Benjamins Publishing Company.

Kibble, R. (2006). Reasoning about propositional commitments in dialogue. *Research on Language and Computation*, *4*(2), 179–202. https://doi.org/10.1007/s11168-006-9003-1

Kissine, M. (2008). Assertoric commitments. *Belgian Journal of Linguistics*, *22*(1), 155–177. https://doi.org/10.1075/bjl.22.08kis

Krifka, M. (2001). Quantifying into question acts. *Natural Language Semantics*, *9*(1), 1–40. https://doi.org/10.1023/A:1017903702063

Krifka, M. (2002). Embedded speech acts.

Krifka, M. (2014). Embedding illocutionary acts. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in cognition* (pp. 59–87). Springer International Publishing.

Krifka, M. (2015). Bias in Commitment Space Semantics: Declarative questions, negated quetions, and question tags. *Semantics and Linguistic Theory*, *25*, 328–345. https://doi.org/10.3765/salt.v25i0.3078

Krifka, M. (2017). Negated polarity questions as denegations of assertions. In C. Lee, F. Kiefer, & M. Krifka (Eds.), *Contrastiveness in information structure, alternatives and scalar implicatures* (pp. 359–398). Springer International Publishing.

Krifka, M. (2021). Modeling questions in commitment spaces. In M. Cordes (Ed.), *Asking and answering: Rivalling approaches to interrogative methods* (pp. 63–95). Gunter Narr Verlag.

Krifka, M. (2022). Adjacency pairs in common ground update: Assertions, questions, greetings, offers, commands. *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue*, 94–105. https://www.semdial.org/anthology/papers/Z/Z22/Z22-3012/

Krifka, M. (2023). Layers of assertive clauses: Propositions, judgments, commitments, acts. In J. M. Hartmann & A. Wöllstein (Eds.), *Propositionale Argumente im Sprachvergleich / Propositional arguments in cross-linguistic research: Theorie und Empirie / Theoretical and empirical issues* (pp. 115–181). Gunter Narr Verlag.

Krifka, M. (2024a). Performative updates and the modeling of speech acts. *Synthese*, *203*(1), 31. https://doi.org/10.1007/s11229-023-04359-0

Krifka, M. (2024b). Structure and interpretation of declarative sentences. *Journal of Pragmatics*, *226*, 51–63. https://doi.org/10.1016/j.pragma.2024.04.002

Lance, M. N., & Kremer, P. (1994). The logical structure of linguistic commitment I: Four systems of nonrelevant commitment entailment. *Journal of Philosophical Logic*, *23*(4), 369–400. https://www.jstor.org/stable/30226526

Lascarides, A., & Asher, N. (2008). Imperatives in dialogue. In P. Kühnlein, H. Rieser, & H. Zeevat (Eds.), *Perspectives on dialogue in the new millennium* (pp. 1–24). John Benjamins Publishing Company.

Lascarides, A., & Asher, N. (2009). Agreement, disputes and commitments in dialogue. *Journal of Semantics*, *26*(2), 109–158. https://doi.org/10.1093/jos/ffn013

Lauer, S. (2013). *Towards a dynamic pragmatics* [Doctoral dissertation]. Stanford University.

Lewis, D. (1972). General semantics. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 169–218). Springer Netherlands.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359. https://doi.org/10.1007/BF00258436

Lewis, D. (1981). Probabilities of conditionals and conditional probabilities. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *IFS: Conditionals, belief, decision, chance and time* (pp. 129–147). Springer.

Liau, C.-J. (2001). A logical analysis of the relationship between commitment and obligation. *Journal of Logic, Language and Information*, *10*(2), 237–261. https://doi.org/10.1023/A:1008353310841

Lyons, J. (1977). *Semantics*. Cambridge University Press.

Lyons, J. (1995). *Linguistic semantics: An introduction*. Cambridge University Press.

MacFarlane, J. (2005). Making sense of relative truth. *Proceedings of the Aristotelian Society*, *105*(1), 305–323. https://doi.org/10.1111/j.0066-7373.2004.00116.x

MacFarlane, J. (2011). What is assertion? In J. Brown & H. Cappelen (Eds.), *Assertion: New philosophical essays* (pp. 79–96). Oxford University Press.

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*(3), 245–297. https://doi.org/10.1007/s11050-009-9042-x

Malamud, S. A., & Stephenson, T. (2015). Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics*, *32*(2), 275–311. https://doi.org/10.1093/jos/ffu002

Mastop, R. (2005). *What can you do? Imperative mood in semantic theory* [Doctoral dissertation]. University of Amsterdam.

Matheson, C., Poesio, M., & Traum, D. (2000). Modelling grounding and discourse obligations using update rules. *Proceedings of NAACL 2000*, 1–8. https://aclanthology.org/A00-2001/

Mazzarella, D., Reinecke, R., Noveck, I., & Mercier, H. (2018). Saying, presupposing and implicating: How pragmatics modulates commitment. *Journal of Pragmatics*, *133*, 15–27. https://doi.org/10.1016/j.pragma.2018.05.009

Meyer, J.-J. C. (1988). A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, *29*(1), 109–136. https://doi.org/10.1305/ndjfl/1093637776

Meyer, J.-J. C., van der Hoek, W., & van Linder, B. (1999). A logical approach to the dynamics of commitments. *Artificial Intelligence*, *113*(1), 1–40. https://doi.org/10.1016/S0004-3702(99)00061-2

Michael, J. (2021). *The philosophy and psychology of commitment*. Routledge.

Moeschler, J. (2013). Is a speaker-based pragmatics possible? Or how can a hearer infer a speaker's commitment? *Journal of Pragmatics*, *48*(1), 84–97. https://doi.org/10.1016/j.pragma.2012.11.019

Northrup, O. B. (2014). *Grounds for commitment* [Doctoral dissertation]. UC Santa Cruz.

Owens, D. (2006). Testimony and assertion. *Philosophical Studies*, *130*(1), 105–129. https://doi.org/10.1007/s11098-005-3237-x

Pagin, P. (2014). Assertion. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). Metaphysics Research Lab, Stanford University.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, *105*(3), 833–838. https://doi.org/10.1073/pnas.0707192105

Portner, P. (2004). The semantics of imperatives within a theory of clause types. *Semantics and Linguistic Theory*, *14*, 235–252. https://doi.org/10.3765/salt.v14i0.2907

Recanati, F. (1987). *Meaning and force: The pragmatics of performative utterances*. Cambridge University Press.

Reese, B., & Asher, N. (2009). Biased questions, intonation, and discourse. In M. Zimmermann & C. Féry (Eds.), *Information structure: Theoretical, typological, and experimental perspectives* (pp. 139–174). Oxford University Press.

Roelofsen, F., & Farkas, D. F. (2015). Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, *91*(2), 359–414. https://doi.org/10.1353/lan.2015.0017

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*(153). https://doi.org/10.3389/fpsyg.2016.00153

Sacks, H. (1992). *Lectures on conversation*. Blackwell.

Šafářová, M. (2005). The semantics of rising intonation in interrogatives and declaratives. *Proceedings of Sinn und Bedeutung*, *9*, 355–369. https://doi.org/10.18148/sub/2005.v9i0.774

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge University Press.

Shapiro, L. (2020). Commitment accounts of assertion. In S. C. Goldberg (Ed.), *The Oxford handbook of assertion* (pp. 74–97). Oxford University Press.

Singh, M. P. (1999). An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, *7*(1), 97–113. https://doi.org/10.1023/A:1008319631231

Singh, M. P. (2000). A social semantics for agent communication languages. In F. Dignum & M. Greaves (Eds.), *Issues in agent communication* (pp. 31–45). Springer.

Smits, A., Jol, G., & Stommel, W. (2022). "Ik neem dat niet terug" : Een conversatieanalytische studie van terugneemverzoeken in de Tweede Kamer en de rol van de voorzitter. *Tijdschrift voor Taalbeheersing*, *44*(3), 129–155. https://doi.org/10.5117/TVT2022.3.001.SMIT

Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, *89*(3), 425–440. https://www.jstor.org/stable/20116982

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (pp. 315–332). Academic Press.

Stenius, E. (1969). Mood and language-game. In J. W. Davis, D. J. Hockney, & W. K. Wilson (Eds.), *Philosophical Logic* (pp. 251–271). Springer Netherlands.

Strawson, P. F. (1964). Intention and convention in speech acts. *The Philosophical Review*, *73*(4), 439–460. https://doi.org/10.2307/2183301

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press.

Szabolcsi, A. (1982). Model theoretic semantics of performatives. In F. Kiefer (Ed.), *Hungarian General Linguistics* (pp. 515–535). John Benjamins Publishing Company.

Titelbaum, M. G. (2022). *Fundamentals of Bayesian epistemology: Introducing credences*. Oxford University Press.

Traum, D. R., & Allen, J. F. (1994). Discourse obligations in dialogue processing. *Proceedings of the 32nd meeting of the ACL*, 1–8. https://doi.org/10.48550/ARXIV.CMP-LG/9407011

Trinh, T. (2014). How to ask the obvious: A presuppositional account of evidential bias in English yes/no questions. In L. Crnič & U. Sauerland (Eds.), *The art and craft of semantics: A Festschrift for Irene Heim* (pp. 227–249). MIT Working Papers in Linguistics.

Tuzet, G. (2006). Responsible for truth? Peirce on judgment and assertion. *Cognitio*, *7*, 317–336. https://papers.ssrn.com/abstract=2888435

van Rooy, R., & Šafářová, M. (2003). On polar questions. In R. Young & Y. Zhou (Eds.), *Proceedings of SALT XIII*. Cornell University.

Van der Auwera, J. (1986). Conditionals and speech acts. In A. T. Meulen, C. A. Ferguson, E. C. Traugott, & J. S. Reilly (Eds.), *On conditionals* (pp. 197–214). Cambridge University Press.

Vanderveken, D. (1994). A complete formulation of a simple logic of elementary illocutionary acts. In S. L. Tsohatzidis (Ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives* (pp. 99–131). Taylor & Francis Routledge.

Vullioud, C., Clément, F., Scott-Phillips, T., & Mercier, H. (2017). Confidence as an expression of commitment: Why misplaced expressions of confidence backfire. *Evolution and Human Behavior*, *38*(1), 9–17. https://doi.org/10.1016/j.evolhumbehav.2016.06.002

Walton, D. (1992). Commitment, types of dialogue, and fallacies. *Informal Logic*, *14*(2). https://doi.org/10.22329/il.v14i2.2532

Watson, G. (2004). Asserting and promising. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *117*(1/2), 57–77. https://www.jstor.org/stable/4321436

Williams, B. (2002). *Truth and truthfulness: An essay in genealogy.* Princeton University Press.

Yamada, T. (2007). Acts of commanding and changing obligations. *Computational Logic in Multi-Agent Systems (CLIMA 2006)*, 1–19. https://doi.org/10.1007/978-3-540-69619-3_1

Yamada, T. (2012). Dynamic logic of propositional commitments. In M. Trobok, N. Miščević, & B. Žarnić (Eds.), *Between logic and reality* (pp. 183–200). Springer Netherlands.

Yates, H., Bary, C., de Swart, P., & van Tiel, B. (forthcoming). Evidentiality and speaker commitment: An fEMG study.

Yates, H., Bary, C., Swart, P. D., & Tiel, B. v. (2025). fEMG as a window into conversational commitments. *Proceedings of Sinn und Bedeutung*, *29*, 1765–1783. https://doi.org/10.18148/sub/2024.v29.1308

Zhang, Y. (Raven). (2024). Discourse consistency and dynamic modals in commitment space semantics. *Proceedings of Sinn und Bedeutung*, *28*, 1060–1078. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/1179

Zhang, Y. (Raven). (n.d.). *Denegational logic for epistemic modals.*

Appendix A

The logical system

A.1 Definitions

Definition 1. (Basic language) Let a finite set of agents $\mathcal{A}$ and a countable set of propositional atoms Atom be given. For $p \in \mathsf{Atom}$, the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ of **SAL** is given by the following BNF:

$$\pi ::= p \mid \neg\pi \mid \pi \wedge \pi \mid \{\mathbf{B}_a\, \pi\}_{a \in \mathcal{A}} \mid \{\mathbf{C}_{a,b}\, \pi\}_{(a,b) \in \mathcal{A}^2}$$

Definition 2. (Commitment states) A *commitment state* over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ is a tuple $c = \langle W^c, \{B_a^c\}_{a \in \mathcal{A}}, \{O_{a,b}^c\}_{(a,b) \in \mathcal{A}^2}, I^c \rangle$, such that:

- W^c is a non-empty set of possible worlds

- $\{B_a^c\}_{a \in \mathcal{A}} \subseteq W^c \times W^c$ is a belief relation (one for each agent) that is transitive, Euclidean and serial.

- $\{O_{a,b}^c\}_{(a,b) \in \mathcal{A}^2} \subseteq W^c \times W^c$ is a commitment relation (one for each agent pair) that is transitive and Euclidean.

- $I^c : \mathsf{Atom} \to \mathcal{P}(W^c)$ is a valuation function.

We write $c, w \models \pi$ to say that π is true in a world w in c. We write $c \models \pi$ to say that $c, w \models \pi$ for all $w \in W^c$.

Definition 3. (State satisfaction) Let c be a commitment state over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $w \in W^c$ a world in c, $a, b \in \mathcal{A}$ two agents, and $p \in \mathsf{Atom}$ and $\pi, \tau \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ all propositions. Then:

- *Atomic propositions.*

  $c, w \models p$ iff $w \in I^c(p)$

- *Connectives.*

  $c, w \models \neg\pi$ iff $c, w \not\models \pi$

  $c, w \models \pi \wedge \tau$ iff $c, w \models \pi$ and $c, w \models \tau$

- *Modal operators.*

  $c, w \models \mathbf{B}_a\, \pi$ iff $c, v \models \pi$ for all worlds $v \in W^c$ such that $wB_a^c v$

  $c, w \models \mathbf{C}_{a,b}\, \pi$ iff $c, v \models \pi$ for all worlds $v \in W^c$ such that $wO_{a,b}^c v$

Definition 4. (Commitment state update) Let c be a commitment state over $\mathcal{L}_{\text{Atom}}^{\mathcal{A}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}_{\text{Atom}}^{\mathcal{A}}$ a proposition. Updating c with π results in the commitment state:

$$c\lceil\pi\rceil_{a,b} = \langle W^c, \{B_x^c\}_{x\in\mathcal{A}}, \{O_{x,y}^{c\lceil\pi\rceil_{a,b}}\}_{(x,y)\in\mathcal{A}^2}, I^c\rangle$$

... such that

$$O_{a,b}^{c\lceil\pi\rceil_{a,b}} = O_{a,b}^c \cap \{(v,w) \in (W^c)^2 \mid c, w \models \pi\}$$

... and $O_{x,y}^{c\lceil\pi\rceil_{a,b}} = O_{x,y}^c$ for all $(x,y) \in \mathcal{A}^2$ such that $(x,y) \neq (a,b)$.

Definition 5. (Sincerity and Competence) Let c be a commitment state over $\mathcal{L}_{\text{Atom}}^{\mathcal{A}}$.

1) c satisfies Sincerity (notation: c^{sin}) iff:

For all $v, w \in W^c$ and $x, y \in \mathcal{A}$: $vB_x^c w$ entails $vO_{x,y}^c w$.

2) c satisfies Competence (notation: c^{comp}) iff:

For all $v, w \in W^c$ and $x, y \in \mathcal{A}$: $vB_y^c w$ entails $vB_x^c w$.

If c satisfies both Sincerity and Competence, we use the notation $c^{\text{sin,comp}}$.

Definition 6. (Sincere commitment state update) Let c^{sin} be a commitment state over $\mathcal{L}_{\text{Atom}}^{\mathcal{A}}$ satisfying Sincerity, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}_{\text{Atom}}^{\mathcal{A}}$ a proposition. Updating c^{sin} with π results in the commitment state:

$$c^{\text{sin}}\lceil\pi\rceil_{a,b}^{\text{sin}} = \langle W^c, \{B_x^{c^{\text{sin}}\lceil\pi\rceil_{a,b}^{\text{sin}}}\}_{x\in\mathcal{A}}, \{O_{x,y}^{c^{\text{sin}}\lceil\pi\rceil_{a,b}}\}_{(x,y)\in\mathcal{A}^2}, I^c\rangle$$

... such that
$$B_a^{c^{\text{sin}}\lceil\pi\rceil_{a,b}^{\text{sin}}} = B_a^{c^{\text{sin}}} \cap O_{a,b}^{c^{\text{sin}}\lceil\pi\rceil_{a,b}}$$

... and $B_x^{c^{\text{sin}}\lceil\pi\rceil_{a,b}^{\text{sin}}} = B_x^{c^{\text{sin}}}$ for all $x \in \mathcal{A}$ such that $x \neq a$.

Definition 7. (Cooperative continuations) Let c, c' be two commitment states over $\mathcal{L}_{\text{Atom}}^{\mathcal{A}}$. c' is a *cooperative continuation* of c (notation: $c \sqsubset c'$) if:

- $W^{c'} = W^c$

- $B_a^{c'} \subseteq B_a^c$ for all $a \in \mathcal{A}$

- $O_{a,b}^{c'} \subseteq O_{a,b}^c$ and $O_{a,b}^{c'} \neq \emptyset$ for all $(a,b) \in \mathcal{A}^2$ and $O_{d,e}^{c'} \subset O_{d,e}^c$ for some $(d,e) \in \mathcal{A}^2$

- $I^{c'} = I^c$

We write $\sqsubseteq$ to denote the improper version of $\sqsubset$, i.e. $c \sqsubseteq c'$ iff $c = c'$ or $c \sqsubset c'$.

Definition 8. (Commitment spaces) A *commitment space* C over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ is a set of commitment states over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ such that:

$$\exists c \in C \; \forall c' \in C : c \sqsubseteq c'$$

- This unique commitment state will be called the *root* of C
  (notation: $\sqrt{C}$).

We write $C, c, w \models \phi$ to say that ϕ is true in a commitment space C, commitment state $c \in C$ and world $w \in W^c$. We write $C, c \models \phi$ to say that $C, c, w \models \phi$ for all $w \in W^c$ and $C \models \phi$ to say that $C, c \models \phi$ for all $c \in C$.

We write $\models_{\mathbf{SAL}} \phi$ to say that $C \models \phi$ for all commitment spaces C.

Definition 9. (Language for assertion) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, the language for assertion $\mathcal{L}_{\mathsf{Act}}^{\mathsf{a}}$ is then given by the following BNF:

$$\alpha ::= \{\mathsf{assert}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2}$$

Definition 10. (Assertion update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition. Updating C with $\mathsf{assert}_{a,b}(\pi)$ results in the commitment space:

$$C[\mathsf{assert}_{a,b}(\pi)] := \{\sqrt{C}\lceil\pi\rceil_{a,b}\} \cup \{c \in C \mid \sqrt{C}\lceil\pi\rceil_{a,b}\lceil\pi\rceil_{b,a} \sqsubseteq c\}$$

Definition 11. (Speech act entailment) Let $\alpha_1, ..., \alpha_n, \beta \in \mathcal{L}_{\mathsf{Act}}$ all be speech acts. We say that $\alpha_1, ..., \alpha_n$ *entails* β (notation: $\alpha_1, ..., \alpha_n \rhd \beta$) iff $C[\alpha_1]...[\alpha_n] = C[\alpha_1]...[\alpha_n][\beta]$, for any commitment space C.

Definition 12. (Language of **SAL**) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and actions $\alpha \in \mathcal{L}_{\mathsf{Act}}$, the language $\mathcal{L}_{\mathbf{SAL}}$ of **SAL** is then given by the following BNF:

$$\phi ::= \pi \mid \{V_{a,b}\}_{(a,b) \in \mathcal{A}^2} \mid \neg\phi \mid \phi \wedge \phi \mid [\alpha]\phi \mid \mathfrak{C}\alpha$$

Definition 13. (Commitment space restriction) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and $c \in C$ a commitment state. Then:

$$C_c := \{c' \in C \mid c \sqsubseteq c'\}$$

Definition 14. (Truth after update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c, $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act and $\phi \in \mathcal{L}_{\mathbf{SAL}}$ a proposition. Then:

$$C, c, w \models [\alpha]\phi \text{ iff } C_c[\alpha], \sqrt{C_c[\alpha]}, w \models \phi$$

Definition 15. (Cooperativity) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. Then:

$$C, c, w \models \mathfrak{C}\alpha \quad \text{iff} \quad C_c[\alpha] \neq C_c \text{ and there is some } c' \in C$$
$$\text{such that } c \sqsubset c' \text{ and } \sqrt{C_c[\alpha]} \sqsubseteq c'$$

Definition 16. (Violation) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state, $w \in W^c$ a world in c and $a, b \in \mathcal{A}$ two agents. Then:

$$C, c, w \models V_{a,b} \text{ iff } O_{a,b}^c = \emptyset$$

Definition 17. (Admissibility) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $c \in C$ a commitment state and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. α is admissible w.r.t. C, c if $C, c \not\models [\alpha]V_{x,y}$, for all $x, y \in \mathcal{A}$.

Definition 18. (Language for questions) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, the language for questions $\mathcal{L}_{\mathsf{Act}}^{\mathsf{q}}$ is then given by the following BNF:

$$\alpha ::= \{\mathsf{assert}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2} \mid \{\mathsf{question}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2}$$

Definition 19. (Question update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, $a, b \in \mathcal{A}$ two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition from the basic language. Updating C with $\mathsf{question}_{a,b}(\pi)$ results in the commitment space:

$$C[\mathsf{question}_{a,b}(\pi)] := \{\sqrt{C}\} \cup$$
$$\{c \in C \mid \sqrt{C}\lceil \pi \rceil_{b,a} \sqsubseteq c\} \cup$$
$$\{c \in C \mid \sqrt{C}\lceil \neg\pi \rceil_{b,a} \sqsubseteq c'\}$$

Definition 20. (Language of actions) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. For propositions $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$, the language of actions $\mathcal{L}_{\mathsf{Act}}$ is then given by the following BNF:

$$\alpha ::= \{\mathsf{assert}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2} \mid \{\mathsf{question}_{a,b}(\pi)\}_{(a,b) \in \mathcal{A}^2} \mid \ominus \mid \sim\beta \mid \beta; \gamma \mid \pi \hookrightarrow \beta/\gamma$$

Definition 21. (Empty update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$. Updating C with $\ominus$ results in the commitment space:

$$C[\ominus] := C$$

Definition 22. (Denegation update) Let C be a commitment space over $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. Updating C with $\sim \alpha$ results in the commitment space:

$$C[\sim \alpha] := (C \setminus C[\alpha]) \cup \{\sqrt{C}\}$$

Definition 23. (Composition update) Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ and $\alpha, \beta \in \mathcal{L}_{\mathsf{Act}}$ two speech acts. Updating C with $\alpha; \beta$ results in the commitment space:

$$C[\alpha; \beta] := C[\alpha][\beta]$$

Definition 24. (Conditional update) Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a proposition and $\alpha, \beta \in \mathcal{L}_{\mathsf{Act}}$ two speech acts. Updating C with $\psi \hookrightarrow \alpha/\beta$ results in the commitment space:

$$C[\pi \hookrightarrow \alpha/\beta] := \begin{cases} C[\alpha] & \text{if } C, \sqrt{C} \models \pi; \\ C[\beta] & \text{otherwise.} \end{cases}$$

A.2 Theorems

Theorem 25. *Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a $\mathbf{C}_{a,b}$-free proposition.[a] Then:*

$$\models_{\mathbf{SAL}} [\mathsf{assert}_{a,b}(\pi)] \, \mathbf{C}_{a,b} \, \pi$$

---
[a] A proposition π is $\mathbf{C}_{a,b}$-free if $\mathbf{C}_{a,b}$ does not occur anywhere in π.

Proof. Take any commitment space C, commitment state $c \in C$ and world $w \in W^c$. Denote $\alpha := \mathsf{assert}_{a,b}(\pi)$. First note that $\sqrt{C_c[\alpha]} = \sqrt{C_c}\lceil\pi\rceil_{a,b}$ (by Definition 10) and $\sqrt{C_c} = c$ (by Definition 13), hence $\sqrt{C_c[\alpha]} = c\lceil\pi\rceil_{a,b}$. Now take any $w' \in W^c$ such that $wO_{a,b}^{c\lceil\pi\rceil_{a,b}} w'$. By Definition 4, it must be that $C_c[\alpha], c\lceil\pi\rceil_{a,b}, w' \models \pi$ (since π is $\mathbf{C}_{a,b}$-free and hence its truth in w' is not affected by the update). Hence, $C_c[\alpha], c\lceil\pi\rceil_{a,b}, w \models \mathbf{C}_{a,b} \, \pi$. Therefore, $C, c, w \models [\alpha] \, \mathbf{C}_{a,b} \, \pi$. $\quad\square$

Theorem 26. *Let c^{sin} be a commitment state over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ satisfying Sincerity, c^{comp} a commitment state over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ satisfying Competence and $c^{\mathsf{sin,comp}}$ a commitment state over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ satisfying both Sincerity and Competence. Moreover let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a proposition. Then:*

1) $c^{\mathsf{sin}} \models \mathbf{C}_{a,b} \, \pi \rightarrow \mathbf{B}_a \, \pi$

2) $c^{\mathsf{comp}} \models \mathbf{B}_a \, \pi \rightarrow \mathbf{B}_b \, \pi$

3) $c^{\mathsf{sin,comp}} \models \mathbf{C}_{a,b} \, \pi \rightarrow \mathbf{B}_b \, \pi$

Proof. For 1), take any $w \in W^{c^{\mathsf{sin}}}$ and suppose $c^{\mathsf{sin}}, w \models \mathbf{C}_{a,b} \, \pi$. Now consider any $w' \in W^{c^{\mathsf{sin}}}$ such that $wB_a^{c^{\mathsf{sin}}} w'$. By Definition 5, it follows that $wO_{a,b}^{c^{\mathsf{sin}}} w'$ and so $c^{\mathsf{sin}}, w' \models \pi$ by our assumption. Therefore, $c^{\mathsf{sin}}, w \models \mathbf{B}_a \, \pi$. The proof for 2) is similar. Statement 3) follows from 1) and 2). $\quad\square$

Theorem 27. *Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $p \in \mathsf{Atom}$ a proposition. Moreover let $\models^{\mathrm{sin}} \pi$ denote that $c^{\mathrm{sin}} \models \pi$ for all commitment states c^{sin} satisfying Sincerity. Then:*

1) $\models^{\mathrm{sin}} \mathbf{C}_{a,b}\, p \to \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$

2) $\not\models^{\mathrm{sin}} \mathbf{C}_{a,b}\, \mathbf{B}_a\, p \to \mathbf{C}_{a,b}\, p$

Proof. For 1), take any commitment state c^{sin} satisfying Sincerity and world $w \in W^{c^{\mathrm{sin}}}$. Suppose that $c^{\mathrm{sin}}, w \models \mathbf{C}_{a,b}\, p$. Now take $w', w'' \in W^{c^{\mathrm{sin}}}$ such that $wO^{c^{\mathrm{sin}}}_{a,b} w' B^{c^{\mathrm{sin}}}_a w''$. By Definition 5, we get $wO^{c^{\mathrm{sin}}}_{a,b} w' O^{c^{\mathrm{sin}}}_{a,b} w''$, and so by transitivity (Definition 2) $wO^{c^{\mathrm{sin}}}_{a,b} w''$. By our assumption then, $c^{\mathrm{sin}}, w'' \models p$. Thus, $c^{\mathrm{sin}}, w \models \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$.

For 2), consider the commitment state c such that $W^c = \{w, w'\}$, $B^c_a = \{(w, w'), (w', w')\}$, $O^c_{a,b} = B^c_a \cup \{(w, w)\}$ and $I^c(p) = \{w'\}$. Clearly, c satisfies Sincerity and $c, w \models \mathbf{C}_{a,b}\, \mathbf{B}_a\, p$. However, since $wO^c_{a,b}w$, we get $c, w \not\models \mathbf{C}_{a,b}\, p$. $\qquad\square$

Lemma 28. *Let C be commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $c \in C$ a commitment state, $a, b \in \mathcal{A}$ two agents and $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act. Denote $c_\alpha := \sqrt{C_c[\alpha]}$. Then:*

$$O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$$

Proof. By induction on the complexity of α.

- $\alpha := \mathsf{assert}_{x,y}(\pi)$ for some agents $x, y \in \mathcal{A}$ and proposition $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$. Then $c_\alpha = \sqrt{C_c\lceil \pi \rceil_{x,y}} = c\lceil \pi \rceil_{x,y}$. It follows from Definition 4 that $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

- $\alpha := \mathsf{question}_{x,y}(\pi)$ for some agents $x, y \in \mathcal{A}$ and proposition $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$. Then $c_\alpha = \sqrt{C_c} = c$ by Definition 19, so clearly $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

- $\alpha := \ominus$. Then $c_\alpha = c$ by Definition 21, and so $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

- $\alpha := \, \sim \beta$ for some $\beta \in \mathcal{L}_{\mathsf{Act}}$. Then again $c_\alpha = \sqrt{C_c} = c$ by Definition 22, so clearly $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

- $\alpha := \beta; \gamma$ for some $\beta, \gamma \in \mathcal{L}_{\mathsf{Act}}$. By applying our Induction Hypothesis twice, we get $O^{\sqrt{C_c[\beta][\gamma]}}_{a,b} \subseteq O^{\sqrt{C_c[\beta]}}_{a,b} \subseteq O^c_{a,b}$, hence $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

- $\alpha := \pi \hookrightarrow \beta/\gamma$ for some $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ and $\alpha, \beta \in \mathcal{L}_{\mathsf{Act}}$. If $C, c \models \pi$, then $c_\alpha = c_\beta$ and by the Induction Hypothesis we get $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$. If $C, c \not\models \pi$, then $c_\alpha = c_\gamma$ and so again we get $O^{c_\alpha}_{a,b} \subseteq O^c_{a,b}$.

$\qquad\square$

Theorem 29. *(Assertion entailment) Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi, \tau \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ two $\mathbf{C}_{a,b}$-free and $\mathbf{C}_{b,a}$-free propositions such that $\{\pi\} \models \tau$.[a,b] Then $\mathsf{assert}_{a,b}(\pi) \rhd \mathsf{assert}_{a,b}(\tau)$.*

---

[a] A proposition π is $\mathbf{C}_{a,b}$-free if $\mathbf{C}_{a,b}$ does not occur anywhere in π.

[b] Given a set of formulas Γ and a formula ϕ, it holds that $\Gamma \models \phi$ iff, for any commitment space C, commitment state $c \in C$ and world $w \in W^c$, $C, c, w \models \gamma$ for all $\gamma \in \Gamma$ implies $C, c, w \models \phi$.

Proof. Let c be any commitment state. Since $\{\pi\} \models \tau$, we have $\{(v, w) \in (W^c)^2 \mid c, w \models \pi\} \subseteq \{(v, w) \in (W^c)^2 \mid c, w \models \tau\}$, and so $c\lceil \pi \rceil_{a,b} = c\lceil \pi \rceil_{a,b}\lceil \tau \rceil_{a,b}$ and $c\lceil \pi \rceil_{b,a} = c\lceil \pi \rceil_{b,a}\lceil \tau \rceil_{b,a}$ by Definition 4 (and the fact that π and τ are $\mathbf{C}_{a,b}$-free and $\mathbf{C}_{b,a}$-free). Thus, by Definition 10, for any commitment space C, $C[\mathsf{assert}_{a,b}(\pi)] = C[\mathsf{assert}_{a,b}(\pi)][\mathsf{assert}_{a,b}(\tau)]$, i.e. $\mathsf{assert}_{a,b}(\pi) \rhd \mathsf{assert}_{a,b}(\tau)$. $\square$

Theorem 30. *(Persistence) Let the basic language $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ be given. Let $a, b \in \mathcal{A}$ be two agents, $\alpha \in \mathcal{L}_{\mathsf{Act}}$ a speech act and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a $\mathbf{C}_{a,b}$-free proposition.[a] Then:*

$$\models_{\mathbf{SAL}} \mathbf{C}_{a,b}\, \pi \to [\alpha]\, \mathbf{C}_{a,b}\, \pi$$

---

[a] A proposition π is $\mathbf{C}_{a,b}$-free if $\mathbf{C}_{a,b}$ does not occur anywhere in π.

Proof. Take any commitment space C, commitment state $c \in C$ and world $w \in W^c$. Suppose that $C, c, w \models \mathbf{C}_{a,b}\, \pi$ and denote $c_\alpha := \sqrt{C_c[\alpha]}$. Take any $w' \in W^c$ such that $wO^{c_\alpha}_{a,b}w'$. By Lemma 28, $wO^c_{a,b}w'$. Since π is $\mathbf{C}_{a,b}$-free and moreover does not depend on $\sqsubset$, it follows that $C_c[\alpha], c_\alpha, w \models \mathbf{C}_{a,b}\, \pi$, i.e. $C, c, w \models [\alpha]\, \mathbf{C}_{a,b}\, \pi$. $\square$

Theorem 31. *Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a $\mathbf{C}_{a,b}$-free and $\mathbf{C}_{b,a}$-free proposition.[a] Suppose that $\sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a} \in C$. Then:*

1) $C, \sqrt{C} \models \neg\, \mathbf{C}_{b,a}\, \pi \to [\mathsf{assert}_{a,b}(\pi)]\mathfrak{C}\, \mathsf{assert}_{b,a}(\pi)$

2) $C, \sqrt{C} \models \neg[\mathsf{assert}_{a,b}(\pi)]\mathfrak{C}\, \mathsf{assert}_{b,a}(\neg\pi)$

---

[a] A proposition π is $\mathbf{C}_{a,b}$-free if $\mathbf{C}_{a,b}$ does not occur anywhere in π.

Proof. Denote $\alpha := \mathsf{assert}_{a,b}(\pi)$ and $\beta := \mathsf{assert}_{b,a}(\pi)$.

For 1), suppose that $C, \sqrt{C} \models \neg\, \mathbf{C}_{b,a}\, \pi$. Since $\sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a} \in C$, by Definition 10 $\sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a} \in C[\alpha]$. Moreover, $C[\alpha], \sqrt{C[\alpha]} \not\models \mathbf{C}_{b,a}\, \pi$, since $C, \sqrt{C} \models \neg\, \mathbf{C}_{b,a}\, \pi$ by assumption and since π is $\mathbf{C}_{a,b}$-free. By Theorem 25, however, $C[\alpha][\beta], \sqrt{C[\alpha][\beta]} \models \mathbf{C}_{b,a}\, \pi$ (since π is $\mathbf{C}_{b,a}$-free). Thus, $C[\alpha][\beta] \neq C[\alpha]$. For the same reason, $\sqrt{C[\alpha]} \neq \sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a}$ and so $\sqrt{C[\alpha]} \sqsubset \sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a}$. Since $\sqrt{C[\alpha][\beta]} = \sqrt{C[\alpha]}\lceil \pi \rceil_{b,a} = \sqrt{C}\lceil \pi \rceil_{a,b}\lceil \pi \rceil_{b,a}$, we get $C, \sqrt{C} \models [\alpha]\mathfrak{C}\beta$.

For 2), simply observe that $\sqrt{C}\lceil \pi \rceil_{a,b}\lceil \neg\pi \rceil_{b,a} \notin C[\alpha]$, and so $C[\alpha], \sqrt{C[\alpha]} \not\models \mathfrak{C}\, \mathsf{assert}_{b,a}(\neg\pi)$, i.e. $C, \sqrt{C} \models \neg[\alpha]\mathfrak{C}\, \mathsf{assert}_{b,a}(\neg\pi)$. $\square$

Theorem 32. *(Possibility) Let $\alpha \in \mathcal{L}_{\mathsf{Act}}$ be a speech act. Then:*

$$\models_{\mathbf{SAL}} [\alpha]\top$$

Proof. Take any commitment space C, commitment state $c \in C$ and world $w \in W^c$. By induction on α, we can show that $C_c[\alpha], \sqrt{C_c[\alpha]}, w \not\models \bot$, since any update that α represents is well-defined. Thus, $C, c, w \models [\alpha]\top$. $\square$

Theorem 33. *(Inadmissibility of contradictions) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition. Then:*

$$\models_{\mathbf{SAL}} [\mathsf{assert}_{a,b}(\pi \wedge \neg\pi)]V_{a,b}$$

Proof. Take any commitment space C, commitment state $c \in C$ and world $w \in W^c$ and denote $\alpha := \mathsf{assert}_{a,b}(\pi \wedge \neg\pi)$. Since $\{(v,w) \in (W^c)^2 \mid c, w \models \pi \wedge \neg\pi\} = \emptyset$, by Definition 4 we obtain $O_{a,b}^{\sqrt{C_c[\alpha]}} = \emptyset$. Thus, $C_c[\alpha], \sqrt{C_c[\alpha]}, w \models V_{a,b}$, i.e. $C, c, w \models [\alpha]V_{a,b}$. $\qquad\square$

Theorem 34. *Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Then for any two agents $a, b \in \mathcal{A}$ and speech act $\alpha \in \mathcal{L}_{\mathsf{Act}}$:*

1) $\models_{\mathbf{SAL}} \mathfrak{C}\alpha \rightarrow \neg[\alpha]V_{a,b}$

2) $\models_{\mathbf{SAL}} \neg[\alpha]V_{a,b} \rightarrow [\alpha]\top$

3) $\not\models_{\mathbf{SAL}} \neg[\alpha]V_{a,b} \rightarrow \mathfrak{C}\alpha$

4) $\not\models_{\mathbf{SAL}} [\alpha]\top \rightarrow \neg[\alpha]V_{a,b}$

Proof. For 1) and 2) and 4), take any commitment space C, commitment space $c \in C$ and world $w \in W^c$.

1) Take any agents $a, b \in \mathcal{A}$ and speech act $\alpha \in \mathcal{L}_{\mathsf{Act}}$. Suppose $C, c, w \models \mathfrak{C}\alpha$. Then by Definition 15 there is some $c' \in C$ such that $c \sqsubset c'$ and $\sqrt{C_c[\alpha]} \sqsubseteq c'$. Furthermore suppose for contradiction that $C, c, w \models [\alpha]V_{a,b}$. Then $C_c[\alpha], \sqrt{C_c[\alpha]}, w \models V_{a,b}$. Since $\sqrt{C_c[\alpha]} \sqsubseteq c'$, by Definition 7, $O_{a,b}^{c'} = \emptyset$ and so $c \not\sqsubset c'$. Contradiction. Hence, $C, c, w \not\models [\alpha]V_{a,b}$.

2) Trivially true by Theorem 32.

3) Consider $\alpha := \mathsf{assert}_{a,b}(p)$ for some $p \in \mathsf{Atom}$ and $C = \{c\}$ such that $W^c = \{v, w\}$, $O_{a,b}^c = W^c \times W^c$ and $I^c(p) = \{w\}$. Denote $\sqrt{C[\alpha]} = c^\alpha$. Then $C[\alpha], c^\alpha, w \not\models V_{a,b}$ and so $C, c, w \models \neg[\alpha]V_{a,b}$. However, since there is no $c' \in C$ such that $c \sqsubset c'$, $C, c, w \not\models \mathfrak{C}\alpha$.

4) Consider $\alpha := \mathsf{assert}_{a,b}(p \wedge \neg p)$ for some $p \in \mathsf{Atom}$. By Theorem 32, $C, c, w \models [\alpha]\top$. However, by Theorem 33, $C, c, w \not\models \neg[\alpha]V_{a,b}$.

$\qquad\square$

Theorem 35. *(Answerhood) Let the basic language $\mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ be given. Let $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}_{\mathsf{Atom}}^{\mathcal{A}}$ a proposition. Then $\mathsf{assert}_{b,a}(\pi) \rhd \mathsf{question}_{a,b}(\pi)$ and $\mathsf{assert}_{b,a}(\neg\pi) \rhd \mathsf{question}_{a,b}(\pi)$.*

Proof. Take any commitment space C. For all $c \in C[\mathsf{assert}_{b,a}(\pi)]$ we have $\sqrt{C}\lceil\pi\rceil_{b,a} \sqsubseteq c$ by Definition 10. So $C[\mathsf{assert}_{b,a}(\pi)] \subseteq C[\mathsf{assert}_{b,a}(\pi)][\mathsf{question}_{a,b}(\pi)]$. Moreover, $C[\mathsf{assert}_{b,a}(\pi)][\mathsf{question}_{a,b}(\pi)] \subseteq C[\mathsf{assert}_{b,a}(\pi)]$ by Definition 19. Hence, $\mathsf{assert}_{b,a}(\pi) \rhd \mathsf{question}_{a,b}(\pi)$. A very similar argument shows $\mathsf{assert}_{b,a}(\neg\pi) \rhd \mathsf{question}_{a,b}(\pi)$. $\qquad\square$

Theorem 36. *Let C be a commitment space over $\mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$, $c \in C$ a commitment state, $a, b \in \mathcal{A}$ be two agents and $\pi \in \mathcal{L}^{\mathcal{A}}_{\mathsf{Atom}}$ a proposition. Suppose that $\sqrt{C}\lceil \pi \rceil_{b,a}, \sqrt{C}\lceil \neg\pi \rceil_{b,a} \in C$. Then:*

1) $C, \sqrt{C} \models [\mathsf{question}_{a,b}(\pi)](\mathfrak{C}\,\mathsf{assert}_{b,a}(\pi) \wedge \mathfrak{C}\,\mathsf{assert}_{b,a}(\neg\pi))$

2) $C, \sqrt{C} \models (\neg\mathbf{C}_{a,b}\,\pi \wedge \neg\mathbf{C}_{a,b}\,\neg\pi) \rightarrow [\mathsf{question}_{a,b}(\pi)]\neg(\mathfrak{C}\,\mathsf{assert}_{a,b}(\pi) \vee \mathfrak{C}\,\mathsf{assert}_{a,b}(\neg\pi))$

Proof. Denote $\alpha := \mathsf{question}_{a,b}(\pi)$.

For 1), note that $\sqrt{C}\lceil\pi\rceil_{b,a} \not\sqsubseteq \sqrt{C}\lceil\neg\pi\rceil_{b,a}$ and $\sqrt{C}\lceil\neg\pi\rceil_{b,a} \not\sqsubseteq \sqrt{C}\lceil\pi\rceil_{b,a}$, hence $\sqrt{C} \neq \sqrt{C}\lceil\pi\rceil_{b,a} \neq \sqrt{C}\lceil\neg\pi\rceil_{b,a}$. By definition, $\sqrt{C}, \sqrt{C}\lceil\pi\rceil_{b,a}, \sqrt{C}\lceil\neg\pi\rceil_{b,a} \in C[\alpha]$. Since $\sqrt{C}\lceil\pi\rceil_{b,a} = \sqrt{C[\mathsf{assert}_{b,a}(\pi)]}$ and $\sqrt{C}\lceil\neg\pi\rceil_{b,a} = \sqrt{C[\mathsf{assert}_{b,a}(\neg\pi)]}$ by Definition 10, we get $C[\alpha], \sqrt{C[\alpha]} \models \mathfrak{C}\,\mathsf{assert}_{b,a}(\pi) \wedge \mathfrak{C}\,\mathsf{assert}_{b,a}(\neg\pi)$, i.e. $C, \sqrt{C} \models [\alpha](\mathfrak{C}\,\mathsf{assert}_{b,a}(\pi) \wedge \mathfrak{C}\,\mathsf{assert}_{b,a}(\neg\pi))$.

For 2), suppose that $C, \sqrt{C} \models \neg\mathbf{C}_{a,b}\,\pi \wedge \neg\mathbf{C}_{a,b}\,\neg\pi$. Then $C[\alpha], \sqrt{C[\alpha]} \models \neg\mathbf{C}_{a,b}\,\pi \wedge \neg\mathbf{C}_{a,b}\,\neg\pi$, since $\sqrt{C} = \sqrt{C[\alpha]}$ by definition 19. By the same definition, $\sqrt{C[\alpha][\mathsf{assert}_{a,b}(\pi)]}, \sqrt{C[\alpha][\mathsf{assert}_{a,b}(\neg\pi)]} \notin C[\alpha]$, since $\sqrt{C[\alpha][\mathsf{assert}_{a,b}(\pi)]} = \sqrt{C}\lceil\pi\rceil_{a,b}$ and $\sqrt{C[\alpha][\mathsf{assert}_{a,b}(\neg\pi)]} = \sqrt{C}\lceil\neg\pi\rceil_{a,b}$. Hence by Definition 15, $C[\alpha], \sqrt{C[\alpha]} \not\models \mathfrak{C}\,\mathsf{assert}_{a,b}(\pi) \vee \mathfrak{C}\,\mathsf{assert}_{a,b}(\neg\pi)$, i.e. $C, \sqrt{C} \models [\alpha]\neg(\mathfrak{C}\,\mathsf{assert}_{a,b}(\pi) \vee \mathfrak{C}\,\mathsf{assert}_{a,b}(\neg\pi))$. $\square$