Modalizing commitment space semantics[1,2]

Marvin SCHMITT — *ZAS & Humboldt-University Berlin*

Abstract. This article discusses short-comings of commitment space semantics as introduced in Krifka (2015) with respect to assertions, confirmations, and contradictions. It proposes to transform commitment space semantics to a modal logical framework which allows not only for an elegant solution of these issues, but also for the accommodation of other phenomena central to any comprehensive model of human communication (this really being the main point behind going modal).

Keywords: assertion, confirmation, contradiction, commitment space semantics, modal logic, speech acts

1. Introduction

Commitment Space Semantics is a formal model of speech acts and was proposed in Krifka (2015). It is designed to model the dynamic effects of speech acts in conversation, i.e., how speech acts shape conversation. In his proposal, Krifka considers assertions and reactions to assertions, particularly confirmations with bare *yes* and contradictions with bare *no* (among others). As will be argued here, his proposal is inadequate and therefore a refinement is called for: contrary to Krifka (2015) the essential effect of an assertion comes about only if the relevant interlocutors share the commitment of the assertion's author, and responses to assertions are a means to promote or block this effect thus preceding it. However, instead of providing such a refinement in commitment space semantics itself[3], this paper proposes to go modal straight away. The reason for this is mainly because conversing is not only a matter of expressing commitments: knowledge, beliefs, intentions, preferences, and strategies (among possibly other things) also play a key role in our daily communicative life, and these notions all have found analyses in modal logical frameworks (for an overview of such frameworks see for instance van Benthem (2011)). Thus, going modal allows for accommodating many phenomena relevant for a comprehensive model of human communication.

The paper is structured as follows. Section 2 introduces commitment space semantics as presented in Krifka (2015) and contains the critique. In section 3 the new framework will be presented and it will be shown how to model assertions and responses to assertions more accurately.[4] Section 4 quickly introduces a simple epistemic extension of the framework presented in section 3. Section 5 concludes.

2. Commitment Space Semantics

In Krifka (2015) *commitment space semantics* has been introduced. It is a formal model for studying the dynamic effects of speech acts in conversation. By this is meant a formal model for studying *how* speech acts *shape* conversation. Conceptually, the model is based on two

[1]I gratefully acknowledge support by the European Research Council, ERC-2017-ADG 787929 SPAGAD, Speech Acts in Grammar and Discourse. Special thanks to Stephanie Rotter for proofreading. Remaining errors are mine.

[2]Logicians shouldn't take the title too serious.

[3]June 10, 2020: Such a refinement has been worked out by and large.

[4]The framework does not model all aspects of assertions and responses to them. Consequently, it is only more accurate than Krifka's proposal.

Proceedings of Sinn und Bedeutung 24, vol. 2, pp.238–252. Osnabrück University.

points. First, conversation has to do primarily with *commitments*. Secondly, speech acts shape conversation by changing commitments of interlocutors. In the following, we will focus on the formal model and neglect the philosophical and conceptual underpinnings of it. We will now serve an introduction of commitment space semantics, followed by a discussion of the analysis of assertion, confirmation with bare *yes*, and contradiction with bare *no* respectively.

2.1. Commitment space semantics

The basic unit of commitment space semantics are *commitment states* c. These are sets of propositions. A commitment state c is intended to represent the shared public commitments of the interlocutors (Krifka (2015): 328-329). In fact, they are treated as common grounds (Krifka (2015): 343). *Speech acts* $\mathfrak{A}$ are identified as functions mapping a commitment state c and a proposition ϕ to a new commitment state $c' = c \cup \{\phi\}$ (Krifka (2015): 329). The basics in place, we can consider the object that gives the model its name: the commitment space. A *commitment space* C is a set of commitment states satisfying two conditions: first, $\bigcap C \neq \varnothing$, and second, $\bigcap C \in C$ (Krifka (2015): 329). A commitment space C is interpreted as adding a temporal dimension: C is supposed to represent the current shared public commitments of a conversation ($\bigcap C$) plus the future developments (any $c \in C$ with $c \neq \bigcap C$) (Krifka (2015): 329). The two conditions ensure that there is a unique non-empty set of shared public commitments, so that we indeed can speak of *the* current shared public commitments or *the* current common ground. This allows then to model the future developments of the current common ground in conversation. The conversational effects of speech acts are also modeled on the level of commitment spaces: $C + \mathfrak{A}_\phi = \{c \in C : \bigcap C \cup \{\phi\} \subseteq c\}$ for updating a commitment space with a speech act (Krifka (2015): 329). This means: the dynamic effect of a speech act $\mathfrak{A}$ with content ϕ (a proposition) on the commitment space C consists in taking all commitment states c in C such that they contain ϕ. The rest you throw away. Below an illustration:

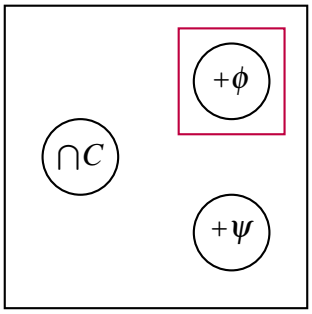


Figure 1: The least element of the space is denoted by $\bigcap C$. $+\phi = \bigcap C \cup \{\phi\}$ and similarly for $+\psi$. The purple square indicates the new commitment space after updating with ϕ, whereas the black square represents the original space.

With this in place, we are prepared for the discussion of assertions, confirmations, and contradictions in Krifka (2015).

2.2. Assertion, confirmation, and contradiction in Krifka (2015)

We will proceed as follows: we start with the formal analysis of assertions. We then consider confirmations with bare *yes*. Last, we do the same for contradictions with bare *no*.

Krifka assumes that a speaker S in asserting ϕ does two things: first, S expresses a commitment to the truth of ϕ (again, this is a proposition), formally written as $S \vdash \phi$, and second, S intends that S's addressee accepts ϕ, thereby making ϕ common ground (Krifka (2015): 332-333).

Formally, Krifka sees the first part of an assertion to be an update of a commitment space C by a speech act $\mathfrak{A}_{S\vdash\phi}$, and the second point by another update of the commitment space $C + \mathfrak{A}_{S\vdash\phi}$ with $\mathfrak{A}_{\phi}$. Importantly, though, it is said that the second update can be cancelled (Krifka (2015): 332-333), meaning it doesn't have to occur.[5]

Confirmations with *yes* are treated as reassertions of prior content (Krifka (2015): 334). So for instance, if speaker S asserted ϕ and speaker S' responses with *yes*, i.e., confirms S's claim, then S' in fact asserts ϕ too. Formally: let C' be the commitment space after S's assertion of ϕ including its essential effect. Then, $C' + \mathfrak{A}_{S'\vdash\phi}$ represents the confirmation of ϕ with *yes* by S'. Or as a sequence of updates:

$$\underbrace{\overbrace{C + \mathfrak{A}_{S_1\vdash\phi} + \mathfrak{A}_{\phi}}^{=C'}}_{\text{assertion by } S_1 \text{ plus grounding of } \phi} + \underbrace{\mathfrak{A}_{S_2\vdash\phi}}_{\text{confirmation by } S_2} \qquad \text{(Krifka (2015): 334)}$$

Contradictions with *no* are similar to their counterparts. According to Krifka, they consist of asserting the contrary, hence in contradicting ϕ S asserts $\neg\phi$ (see Krifka (2015): 334). Formally, they are more complex then the above suggests. This has to do with Krifka including an extra step: a retraction.[6] The reason for this being that he treats a contradiction to appear after a successful assertion, by which is meant an assertion that brought about its essential effect (grounding of its content). So, formally then, Krifka proposes the following:

$$\overbrace{\underbrace{\underbrace{C + \mathfrak{A}_{S_1\vdash\phi} + \mathfrak{A}_{\phi}}_{=\bigcap C\cup\{S_1\vdash\phi,\phi\}}}_{=\bigcap C\cup\{S_1\vdash\phi\}} + \mathfrak{R} + \overbrace{\mathfrak{A}_{S_2\vdash\neg\phi}}^{\text{Assertion by } S_2}}^{\text{Retraction}} \qquad \text{(Krifka (2015): 334)}$$

Retraction undoes the update preceding it (see Krifka (2015): 331).

2.3. Conversing is more interactive

We will now assess Krifka's proposals starting with confirmations, followed by contradictions, and ending with assertions.

As we saw above, confirming for Krifka boils down to reasserting a proposition. Importantly, on Krifka's account such happens after the essential effect of an assertion has come about. Here

[5] In fact, Krifka argues that the second update, which is what is commonly called *the essential effect of an assertion* comes about via implicature. The argumentation below will shed doubts on this.

[6] According to Krifka (2015): 335 retraction is not a feature of the semantics of *no* itself, but is triggered whenever necessary for maintaining a blatantly coherent common ground (for the latter see Krifka (2015): 329), which we ignored since it is of no interest to us).

again the sequence of updates:

$$C + \mathfrak{A}_{S_1 \vdash \phi} + \mathfrak{A}_{\phi} + \mathfrak{A}_{S_2 \vdash \phi}$$

Formally speaking the confirmation makes common ground that the addressee, here S_2, commits to the truth of the proposition, here ϕ, too. *Conceptually* though, there is an issue with the sequencing. We have prior to this commitment that ϕ itself becomes common ground, for it becomes an element of the least commitment state of that commitment space. The intended meaning of this is that all conversationalists share a public commitment to that proposition, thus making the confirmation by S_2 redundant.[7] In fact, other researcher on assertions and reactions to assertions argue that confirmations must precede the grounding of the content of the assertion (see for instance Farkas and Bruce (2010): 82, 92). Having confirmations preceding the essential effect of an assertion makes them conceptually non-redundant.

Contradictions, like confirmations, are treated as a sequence of updates:

$$C + \mathfrak{A}_{S_1 \vdash \phi} + \mathfrak{A}_{\phi} + \mathfrak{R} + \mathfrak{A}_{S_2 \vdash \neg \phi}$$

Again, we see that grounding the content of an assertion precedes the act of contradiction. This is problematic for at least two reasons. First, according to Krifka, grounding comes about as a conversational implicature (see below). This raises the very issue of why a contradiction then does not act as a cancelling-operation. Secondly, it makes contradictions look like revisions, and in fact they are modeled as such in Krifka (2015). However, revising one's stance on a matter is in general a different matter from contradicting someone else on a matter. The latter does not have to involve any revision. So, Krifka's formal proposal is not a good take on contradictions in general (and ones with bare *no* in particular). And this is precisely because he takes them to happen *after* the essential effect of an assertion came about which makes it necessary to involve a retraction (a kind of revision). Having contradictions blocking the essential effect of an assertion and thus potentially preceding it is the way to go (see also Farkas and Bruce (2010): 82, Incurvati and Schlöder (2017): 746 among others again).

Let us now turn to assertions. We have already touched upon the claim that grounding the content of an assertion is a conversational implicature. The above suggests that this is unlikely. Instead, grounding seems dependent on the presence of specific reactions. We will now take a closer look at this.

Krifka argues that the intention to make the addressee *believe* the content of one's proposition is a conversational implicature on the bases of cases such as the ones below (cf. Krifka (2015): 332):

(1) Believe it or not, I won the race.

(2) I don't care whether you believe me, but I won the race.

He then concludes that this is true also of the intention of making the addressee *accept* the proposition. Again, this is argued to follow from the cases above (Krifka (2015): 333). However, on a commitment-based account, acceptance is only reasonably understood as a matter of

[7] Someone may object: "Well, but S_2 commits to the *truth of* ϕ in one case and to ϕ in the other. These are different things." Usually, propositions are seen as things either true or false. So committing to them seems to make sense only when committing to them being true or false. Since committing to the falsity of ϕ is the same as committing to the truth of $\neg\phi$ we can conclude that in ϕ being common ground, everyone commits to its truth, and not to its falsity.

commitment. Clearly, the above cases are not about commitments. So, the argumentation falls short.[8]

Nevertheless, it is worth entertaining the thought that grounding comes about by means of a conversational implicature for little longer. What are the consequences of that thought? First, if the essential effect of an assertion – grounding its content – is a matter of conversational implicature, this means that it will come about whenever no one cancels it. This is what we have in Stalnaker (1999) as an requirement. Second, it follows that acceptance does not have to be displayed. In the following, we will see that these consequences do not hold and so the proposal offered by Krifka cannot hold either.

Conversation analysis provides empirical evidence against Krifka's account. To this end we will consider some general insights from the field as well as an example. We follow here the presentation in Sidnell (2010).

Conversation analysis found that conversation is organized by a couple of mechanisms, among which are so-called *preferences*. This term is used here not to describe psychological states, but to describe orderings on reaction choices which are institutionalized, i.e., orderings that are part of conversing (cf. Sidnell (2010): 77). In general: a response that promotes the achievement of whatever it is the conversational participants are trying to undertake is preferred (cf. Sidnell (2010): 81, 90-91).

Another important insight of conversation analysis is the import of delay in rejections. Rejections come in a bunch of ways, but a common pattern observed is that they come with a *delay*. One kind of delay consists in the rejection being preceded by other material. Another kind of delay is silence between turns (cf. Sidnell (2010): 78). This already makes clear that silence is not a good indicator of agreement.

Last but not least, assertions and assessments display a preference for aligning responses, hence for agreements (cf. Sidnell (2010): 85).

The following example illustrates these findings nicely. It is taken from Sidnell (2010): 2.

> Context: Ann and Jeff are married. They had two friends with their child visiting who stayed overnight, for breakfast and into the early afternoon. After saying goodbye, the couple returns to their house and the following happens:
>
> (1) Visit – FN
> 01 Ann: That was fun,
> 02 (0.4)
> 03 Jeff: mm
> 04 Ann: ish.

What is going on here? In line 01 we have a positive assessment of the event by Ann. Note that this is an assertion: Ann commits herself to the truth of the proposition that the stay by their friends with their child was fun. Also, she intends Jeff to agree on this as the rest of the conversation makes clear. Moreover, at that moment of the conversation Ann reaches a point

[8] Of course, one might be able to construe cases like the above ones addressing commitments. We will not consider this possibility here.

where she can finalize her turn and she does so – at least from the perspective of that very moment of the conversation. Now, it is supposedly Jeff's turn, for Ann started an assessment and she cannot conclude such without Jeff's contribution. But, silence follows (line 02 with 0.4 seconds of silence). This is an inter-turn gap. In line 03, Jeff finally response "mm" which can be taken to signal agreement. Importantly though, line 04 makes clear that Ann does not take Jeff to agree with her assessment, because she initiates a repair and downgrades her assessment by adding "ish". This is because she takes Jeff to not agree with her on the matter which is presumably because of the delay of Jeff's response, which is as we said earlier a common strategy to express disagreement. This suggests then to Ann that Jeff is in fact not aligning with her view of things and thus rejecting it. This then leads her to adjust her initial claim so that Jeff can align with her (see Sidnell (2010): 2-4 for the description of that case).

Thus: first, *silence* is not a means for signaling agreement in general. Second, people are seeking for positive responses on such occasions in general; hence, third, they are not satisfied with the absence of negative responses in general.[9] All this goes against Krifka's position, for according to that people are satisfied with the absence of negative responses on such occasions and silence is a way to positively assess assertions.

Another line of research shading doubt on Krifka's account is that of Herbert Clark and his colleagues. Without going into an extensive discussion of Clark's account[10] we want to note the following: in Clark (1996) it is argued on empirical grounds for a hierarchical architecture of human communication. This architecture consists of four levels (cf. Clark (1996): 149-150):

1. Level 1: Executing behavior and attending
2. Level 2: Presenting and identifying
3. Level 3: Signaling and understanding
4. Level 4: Projecting and considering

The issue we are considering here is situated on level 4. Importantly, Clark argues that any communicative action needs grounding on all levels (Clark (1996): 221ff). Furthermore, he argues that such is achieved by providing *positive evidence* that the action underway was successful and that conversationalists are looking for such evidence in conversing (cf. Clark (1996): 222, 226; see also Clark and Schaeffer (1989) making this point already for level 3 and lower). On level 4 this means that speaker *S* is looking for acceptance of *S*'s intention. And it is for the addressee to either accept or reject that intention upon uptake. It is not for the addressee to simply not reject it or do nothing. Again, this is contrary to Krifka's position. Therefore, we find that the essential effect of an assertion is in need of a reaction by the addressee and in particular a positive one.

In sum, we find that assertions demand positive assessments before grounding can occur. Otherwise, no grounding takes place. Confirmations and contradictions are means to react to assertions and either promote the essential effect of an assertion or block it. In the following section, we will introduce a formalism that allows us to model assertions in a way consistent with our

[9] *In general*, because some conversational practices deviate from the norm in some way, e.g. reacting to compliments; see Sidnell (2010) chapter 5. In any case, the argument against Krifka (2015) stays untouched.
[10] The reader is invited to consult Clark (1996).

conclusion. It will make sense of responses to assertions and will treat grounding (at level 4) as a function of the assertion and the reaction to it. What the formalism will not deliver is a formal take on the proposal nature of assertions, i.e., we will not be able to model the intentions of the author of an assertion to make the assertion's content common ground. This is not a short-coming over Krifka (2015) for there it is not modeled either. We will also not attempt to model grounding on levels lower than level 4. Overall, then, the proposal in the next section is a real improvement over the account offered in Krifka (2015) with respect to assertions and responses to assertions.

3. Modalizing Commitment Space Semantics

In this section we will introduce *modal commitment space semantics* which takes commitment space semantics and transforms it into a modal logical framework. We will serve the motivation first. Then, we will proceed classical: we introduce the language, the frames, and finally the models. Last, we cast an analysis of assertion and responses to assertion in the new framework and show that it fits the bill of the criticism.

3.1. Motivation

The issues addressed so far are not about the architecture of commitment space semantics, so why proposing to move to another architecture? The reason is methodological: even though it seems possible that commitment space semantics in its current state can account for the criticism put forward here[11], moving to a proper modal logical setting brings advantages: we can readily accommodate many other phenomena which are of importance to human communication. This is not to say, that commitment space semantics cannot accommodate phenomena such as knowledge or belief, but this is more complicated. As we will see, transforming commitment space semantics into a modal logical framework is not hard at all (when giving up on a few formalities). Moreover, dynamic modal logics can account for conversational dynamics by the same means they account for other dynamic processes (for an overview see van Benthem (2011)) which provides us with a lean formalism.

3.2. Modal language, frames, and models

Definition 1 (Modal language)**.** Given a set of proposition letters Φ and a set of agents $\mathcal{A}$ (which we fix for specific applications), our modal language $\mathfrak{L}$ is given by the following inductive syntax:

$$\phi :: p \mid \neg\phi \mid \phi \wedge \phi \mid C_a\phi, \textit{ where } a \in \mathcal{A}$$

For simplicity we will take $\mathcal{A}$ to be finite in the following.[12]

Definition 2 (Commitment frame)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. A *commitment frame* is a tuple $(W, \{\Sigma^a_w : a \in \mathcal{A}, w \in W\}, \{R_a : a \in \mathcal{A}\})$, where W is a non-empty set of worlds, for each $a \in \mathcal{A}$ and $w \in W$, Σ^a_w is a set of formulas ϕ from $\mathfrak{L}$. We call Σ^a_w the *commitment set of a in w*. Last, for each $a \in \mathcal{A}$, R_a is a binary accessibility relation on W. We set: wR_av if and only if $\Sigma^a_w \subseteq \Sigma^a_v$.

[11] June 10, 2020: In fact, the proposal made below seems to work for commitment space semantics too.

[12] The definitions below can be easily adjusted to fit the general case with $\mathcal{A}$ finite or denumerable. However, for exposition the finite case is more suited and is anyway of more interest to our specific applications.

We define frames in such a way because in commitment space semantics commitment states are seen as representing the shared public commitments of conversationalists. Obviously, these are a function of the individual commitments of the participants, so we take the latter as starting point. Making commitments a matter of frames is a contingent choice, however, it resembles the architecture of commitment space semantics more closely. In a sense, what we did here is taking commitment states and glue them to points which we related with lines. Last, the accessibility relations are reflexive not because we want to claim that commitments are veridical (like knowledge), but because the commitments of an agent in a world depend on the commitment set associated with that world.

Definition 3 (Commitment model)**.** A *commitment model* $\mathfrak{M}$ is a tuple $(\mathfrak{F}, V)$ consisting of a commitment frame $\mathfrak{F}$ and a valuation function V, with $V : \Phi \to \mathcal{P}(W)$ ($\mathcal{P}(W)$ denotes the powerset of W).

With the models in place, we can turn to the semantics.

Definition 4 (Semantics)**.**

1. $\mathfrak{M}, w \Vdash p$ iff $w \in V(p)$
2. $\mathfrak{M}, w \Vdash \neg\phi$ iff not $\mathfrak{M}, w \Vdash \phi$
3. $\mathfrak{M}, w \Vdash \phi \wedge \psi$ iff $\mathfrak{M}, w \Vdash \phi$ and $\mathfrak{M}, w \Vdash \psi$
4. $\mathfrak{M}, w \Vdash C_a\phi$ iff for all $v \in W$, wR_av implies $\phi \in \Sigma_v^a$

Clauses 1 to 3 are standard and do not need any further explanation. Regarding clause 4, we treat commitments as universal modalities.

Having the basics in place, we can next consider how to model assertion and responses to assertions.

3.3. Assertion and responses to assertion in modal commitment space semantics

Above we said (following Krifka (2015) among others) that in asserting ϕ a speaker S does the following: first S publicizes a commitment to ϕ, and second, S proposes to make ϕ common ground. The latter we argued is a function of the addressee's public opinion on ϕ: if the addressee also publicly commits to ϕ, then ϕ eventually will become common ground, otherwise ϕ will not become common ground.

Assertions will be modeled similar as in Krifka (2015). They involve an update. However, unlike in Krifka (2015), we will not treat the second effect of an assertion as an update. Instead, it will come about as a global property of frames that derives from the actual individual commitments of interlocutors. To this end, we further complicate our formal architecture.

Definition 5 (Agent indexed commitment frame)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. An *agent indexed commitment frame* is a tuple $(\mathfrak{F}, a_i)$ where $\mathfrak{F}$ is based on $\mathfrak{L}$.

Notation. An agent indexed commitment frame can also be denoted by $\mathfrak{F}_{a_i}$.

Definition 6 (Agent indexed commitment model)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. An *agent indexed commitment model* is a

tuple $(\mathfrak{F}_{a_i}, V)$ where $\mathfrak{M}$ is based on $\mathfrak{L}$.

Notation. An agent indexed commitment model can also be denoted by $\mathfrak{M}_{a_i}$.

Definition 7 (Discourse model)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. A *discourse model* $\mathfrak{D}$ is an n-tuple $(\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n})$ of agent indexed commitment models based on $\mathfrak{L}$.

Convention. For modelling conversation we will assume that at the beginning of the conversation, we have a discourse model $\mathfrak{D}$ with the following property: for $i, j \leq n$ and $i \neq j$: $\mathfrak{M}_{a_i} = ((\mathfrak{F}, a_i), V)$ and $\mathfrak{M}_{a_j} = ((\mathfrak{F}, a_j), V)$. So, we will be assuming that the individual agent indexed commitment models are the alike at the start. We will say they are the same for simplicity.

We install this convention as a means of simplification. It gives us that conversationalists conceive the same possible outcomes of a conversation. This is an idealization since people seem not to do so in general. Importantly, though, nothing hinges on this convention – it only makes exposition easier.

Before turning to the definition of an update-operator, that we will use for modelling the first effect of assertions and the like, we will provide definitions of actual commitments and common ground.

Notation. Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. By a_i^{-1} we denote the agent that has been assigned the number i by the enumeration.

Definition 8 (Actual commitments of agent a)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $\mathfrak{M}_{a_i}$ with domain W be an agent indexed commitment model based on $\mathfrak{L}$. Let ϕ be a formula from $\mathfrak{L}$. Let $a_i^{-1} = a$. Then, we say that ϕ is an actual commitment of a iff $\phi \in \{\psi : \psi \in \Sigma_w^a \textit{ for each } w \in W\}$.

Definition 9 (Common ground)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $\mathfrak{D} = (\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n})$ be a discourse model with $\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n}$ based on $\mathfrak{L}$. Further, let ϕ be a formula from $\mathfrak{L}$. We say that ϕ is common ground if and only if we have for each $\mathfrak{M}_{a_i} (i = 1, \ldots, n)$ that $\phi \in \{\psi : \psi \in \Sigma_w^a \textit{ for each } w \in W\}$ with $a = a_i^{-1}$.

So, an actual commitment of an agent a is a formula to which the agent is committed in every world in a's commitment model. And, a formula is common ground if and only if it is an actual commitment of every agent. The notion of common ground here is related to the notion of a context set then (see Stalnaker (1999) and Stalnaker (2002)).

Finally, we will define an update operator which will give us a good formal representation of assertion, confirmation, and contradiction.

Definition 10 (+-operator)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $\mathfrak{D}$ be a discourse model $(\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n})$ with $\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n}$ based on $\mathfrak{L}$. Further, let ϕ be a formula from $\mathfrak{L}$ and a_i be an indexed agent ($i \leq n$). Last, let $W^{\mathfrak{M}_{a_i}}$ be the domain of $\mathfrak{M}_{a_i}$. Then:

$$+_{a_i}^{\mathfrak{D}} \phi := \begin{cases} \mathfrak{D}' & \textit{if } \exists w \in W^{\mathfrak{M}_{a_i}} : \phi \in \Sigma_w^a \textit{ with } a = a_i^{-1} \\ \textit{undefined} & \text{else} \end{cases}$$

where $\mathfrak{D}'$ is defined as follows:

$$\mathfrak{D}' = (\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}'_{a_i}, \ldots \mathfrak{M}_{a_n}), \textit{ with}$$

$$\mathfrak{M}'_{a_i} := (W' := \{w \in W^{\mathfrak{M}_{a_i}} : \phi \in \Sigma_w^{a_i^{-1}}\}, \{R_a \bigcap (W' \times W')\}_{a \in \mathcal{A}}, V'), \textit{ with } V'(p) := V(p) \bigcap W',$$

for each $p \in \Phi$.

The +-operator (*read*: plus-operator) thus defines the usual update we saw already. Importantly, though, + operates *only* on the agent indexed model. This is crucial since it allows us to model the first effect of an assertion as an update independently from the second effect. The latter will come about as an overall emerging property of the discourse. An example will illustrate this:

Let us assume that we have two agents a and b and the discourse model $\mathfrak{D}$ displayed below. Now, a asserts ϕ which results in the following situation:

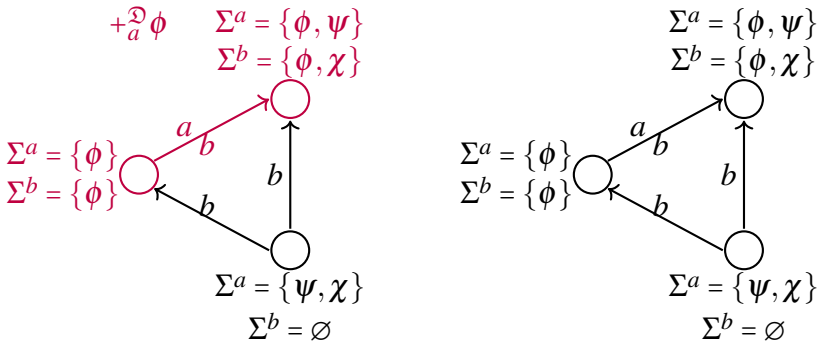


Figure 2: On the left, we have the agent indexed commitment model for agent a, on the right the one for b. Reflexive arrows have been omitted.

By convention, the initial discourse model has the same commitment model underlying its agent indexed commitment models. The purple part in the left model denotes the result of the assertion of ϕ by a. A simple calculation shows that this act results in adding ϕ to the actual commitments of a. Another simple calculation shows then that ϕ is not common ground relative to the discourse model $\mathfrak{D}'$, because it is not an actual commitment of b. Obviously, if we have that b asserts ϕ as well (an act of confirming), ϕ will become common ground relative to the discourse model $\mathfrak{D}''$. If such act does not occur, no grounding will occur.

As the above already indicated, we will follow Krifka (2015) in treating confirmations (with bare *yes*) simply as assertions. Similarly, we will follow Krifka (2015) in treating contradictions (with *no*) as assertions to the contrary.[13] Hence, the +-operator gives us all we need.

3.4. And now for something completely different: weak rejection

Weak rejections have not been discussed in Krifka (2015), however, they are of interest to anyone interested in assertion. While assertions are used in order to get a proposition grounded (their essential effect), weak rejections have a different job: they are used to block this effect.

[13] Obviously, we do not have the means to model *yes* or *no*. So, in fact, our treatment here is more general.

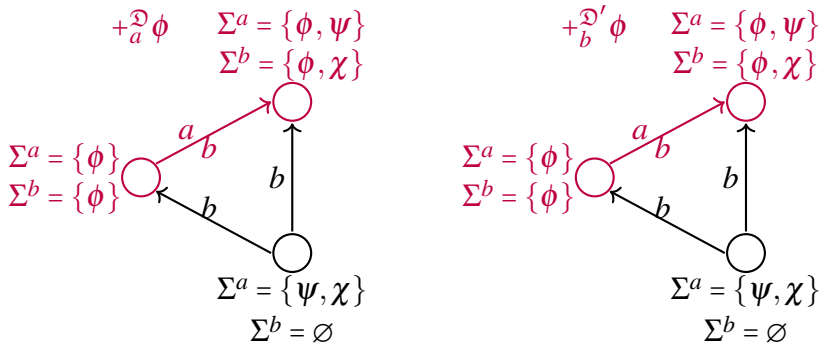


Figure 3: On the left, we have the agent indexed commitment model for agent a, on the right the one for b. Reflexive arrows have been omitted. On the right, in purple the result of confirmation. The update $+^{\mathfrak{D}'}_{b}\neg\phi$ is not defined, since b does not commit to $\neg\phi$ in any world. So, a contradiction by b cannot occur in that case.

And that is all there is to them. For an extensive discussion of weak rejections the reader is referred to Incurvati and Schlöder (2017).

We can do the following here:

Definition 11 (−-operator)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $\mathfrak{D}$ be a discourse model $(\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n})$ with $\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}_{a_n}$ based on $\mathfrak{L}$. Further, let ϕ be a formula from $\mathfrak{L}$ and a_i be an indexed agent ($i \leq n$). Last, let $W^{\mathfrak{M}_{a_i}}$ be the domain of $\mathfrak{M}_{a_i}$. Then:

$$-^{\mathfrak{D}}_{a_i}\phi := \begin{cases} \mathfrak{D}' & \text{if } \exists w \in W^{\mathfrak{M}_{a_i}} : \phi \notin \Sigma^a_w \text{ with } a = a_i^{-1} \\ \textit{undefined} & \text{else} \end{cases}$$

where $\mathfrak{D}'$ is defined as follows:

$$\mathfrak{D}' = (\mathfrak{M}_{a_1}, \ldots, \mathfrak{M}'_{a_i}, \ldots \mathfrak{M}_{a_n}), \text{ with}$$

$$\mathfrak{M}'_{a_i} := (W' := \{w \in W^{\mathfrak{M}_{a_i}} : \phi \notin \Sigma^{a_i^{-1}}_w\}, \{R_a \bigcap (W' \times W')\}_{a \in \mathcal{A}}, V'), \text{ with } V'(p) := V(p) \bigcap W',$$

for each $p \in \Phi$.

This excludes all worlds from the indexed agent model of a_i^{-1} where a_i^{-1} is committed to ϕ thereby blocking any commitment to ϕ, which is what we wanted. However, we must note that the above does not entirely match the proposal in Incurvati and Schlöder (2017) since they have a weak rejection involving a public commitment to not committing to a proposition (cf. Incurvati and Schlöder (2017): 756). So, we are missing out on an important point here. This is because of our static-base architecture and could be tackled by adding dynamic operations on commitment sets: we could define an operation which adds formulas to commitment sets and redefine the minus-operator by utilizing that operation plus the notion of consistency. Other possibilities exist.[14] We will not pursue such matters here. An example for the minus-operator can be found at the top of the next page.

[14]The speech act of weak rejection might be modelled in Krifka (2015) in terms of denegation. It would then be treated as a denegation of committing to the truth of the relevant proposition. The same issues would arise for Krifka (2015).

We will end this section with two propositions.

Proposition 1. *Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. Let $\mathfrak{D}$ be any discourse model based on $\mathfrak{L}$, and let ϕ be any formula of our language. Then it holds: it is not the case that whenever $-^D_{a_i}\phi$ we also have that $C_a\neg\phi \in \{\psi : \psi \in \Sigma^a_w$ for each $w \in W\}$ with $a = a_i^{-1}$.*

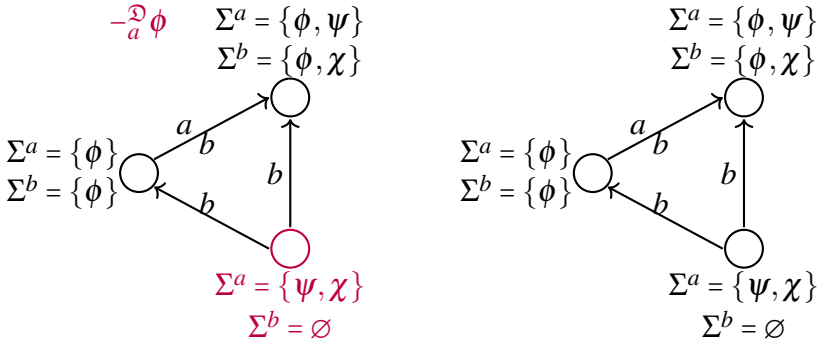
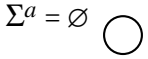


Figure 4: On the left, we have the agent indexed commitment model for agent a, on the right the one for b. Reflexive arrows have been omitted. In purple the effect of a's weakly rejecting ϕ.

Proof. Let $\mathfrak{L}$ be defined from $\Phi = \{p\}$ and $\mathfrak{A} = \{a\}$. Then, let $\mathfrak{D}$ be the discourse model graphically represented below, where we omitted reflexive arrows. Further, let $\phi = p$.



Performing a weak rejection $-^D_{a_i}p$ does not alter the model, and so we will not find that $C_a\neg p$ holds afterwards. □

This property is important, for a weak rejection is different from a contradiction. In weakly rejecting, I am not committing to the contrary of what I am rejecting. Again, we refer to Incurvati and Schlöder (2017) for discussion.

On the other hand Incurvati and Schlöder (2017): 749 suggest that a strong rejection (the speech act of contradicting) entails a weak rejection. This property does not hold in the current setting as the next proposition makes clear.

Proposition 2. *Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. Let $a_1, \ldots, a_n$ be an enumeration of $\mathcal{A}$. Let $\mathfrak{D}$ be any discourse model based on $\mathfrak{L}$, and let ϕ be any formula of our language. Then it holds: it is not the case that whenever we have $+^D_{a_i}\neg\phi$ we also have that $\neg C_a\phi \in \{\psi : \psi \in \Sigma^a_w$ for each $w \in W\}$ with $a = a_i^{-1}$.*

Proof. Again, a very small model suffices. Let $\mathfrak{L}$ be defined from $\Phi = \{p\}$ and $\mathfrak{A} = \{a\}$. Then, let $\mathfrak{D}$ be the discourse model graphically represented below, where we omitted reflexive arrows. Further, let $\phi = p$.
Here, $+^D_a\neg p$ does not alter the initial model. Since $p \in \Sigma^a$ we do not have that $\neg C_a p$. □

$$\Sigma^a = \{\neg p, p\} \bigcirc$$

Importantly, this property holds only if we assume that commitments must be *consistent* – as Incurvati and Schlöder in fact do (see Incurvati and Schlöder (2017): 756). In principle, we could do so as well. We could impose restrictions on commitment sets in various ways. Alternatively, we could impose modal logical axioms. Again, we will not pursue any of these matters here. *Prima facie* such changes should not affect our first proposition.

4. Epistemic and doxastic extensions

In this section we will introduce one extension by adding knowledge. Extensions with belief or other modalities are done similarly and therefore will not be discussed. For readers unfamiliar with epistemic logic we refer to van Benthem (2011) once again. We will only provide the basic ingredients and leave specific implementations to the users, meaning that we will not go into any axiomatizations or other topics.

4.1. Adding knowledge

Definition 12 (Epistemic language)**.** Given a set of proposition letters Φ and a set of agents $\mathcal{A}$ (which we fix for specific applications), our modal language $\mathfrak{L}_K$ (K for *knowledge*) is given by the following inductive syntax:

$$\phi :: p \mid \neg\phi \mid \phi \wedge \phi \mid C_a\phi \mid K_a\phi, \text{ where } a \in \mathcal{A}$$

Definition 13 (Epistemic commitment frame)**.** Let $\mathfrak{L}$ be a language based on the set of agents $\mathcal{A}$. An *epistemic commitment frame* is a tuple $(W, \{\Sigma^a_w : a \in \mathcal{A}, w \in W\}, \{R_a : a \in \mathcal{A}\}, \{\simeq_a : a \in \mathcal{A}\})$, where W is a non-empty set of worlds, for each $a \in \mathcal{A}$ and $w \in W$, Σ^a_w is a set of formulas ϕ from $\mathfrak{L}$. We call Σ^a_w the *commitment set of a in w*. Last, for each $a \in \mathcal{A}$, R_a is a binary accessibility relation on W. We set: wR_av if and only if $\Sigma^a_w \subseteq \Sigma^a_v$. Similarly, $\simeq_a$ is a binary accessibility relation on W and is used for encoding knowledge. A common way to do so is to have $\simeq$ an equivalence relation.

Definition 14 (Epistemic commitment model)**.** An *epistemic commitment model* $\mathfrak{M}$ is a tuple $(\mathfrak{F}, V)$ consisting of an epistemic commitment frame $\mathfrak{F}$ and a valuation function V.

Definition 15 (Semantics)**.**

1. $\mathfrak{M}, w \Vdash p$ iff $w \in V(p)$
2. $\mathfrak{M}, w \Vdash \neg\phi$ iff not $\mathfrak{M}, w \Vdash \phi$
3. $\mathfrak{M}, w \Vdash \phi \wedge \psi$ iff $\mathfrak{M}, w \Vdash \phi$ and $\mathfrak{M}, w \Vdash \psi$
4. $\mathfrak{M}, w \Vdash C_a\phi$ iff for all $v \in W$, wR_av implies $\phi \in \Sigma^a_v$
5. $\mathfrak{M}, w \Vdash K_a\phi$ iff for all $v \in W$, $w \simeq_a v$ implies $\mathfrak{M}, v \Vdash \phi$

All other definitions from the last section can be adjusted in the same way and so we will not do so here. Belief can be added similarly.

This extension allows to study how the dynamics of commitments influences knowledge. We will, however, not attempt this here. What you cannot do with this system is to study how knowledge influences commitments. This requires an epistemic update operator. A very potential way to do this is to go a step further and to make use of *product updates* (cf. Baltag et al. (1998)). We will not undertake this here either.[15]

5. Conclusion

This paper addressed assertions and responses to assertions, i.e., confirmations with bare *yes* and contradictions with bare *no*, in commitment space semantics as presented in Krifka (2015). We showed that the formal analysis of these conversational acts is empirically unmotivated and thus asks for revision. We offered instead a new framework taking lead from the architecture of commitment space semantics. Such was motivated by the need to accommodate other phenomena as well, in particular knowledge, belief, and preferences (among other things). Hence, we moved to a modal framework. In this framework we provided empirically grounded anaylses of the above speech acts, thereby improving on commitment space semantics. We furthermore showed that with some more effort a treatment of the speech act of weak rejection (Incurvati and Schlöder (2017)) is possible as well. Last, we gave a very basic epistemic extension of the presented framework and indicated how to go even further.

References

Asher, N. and A. Lascarides (2008). Commitments, beliefs and intentions in dialogue. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (Londial)*, pp. 35–42.

Baltag, A., L. S. Moss, and S. Solecki (1998). The Logic of Public Announcements, Common Knowledge, and Private Suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '98, San Francisco, CA, USA, pp. 43–56. Morgan Kaufmann Publishers Inc.

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

Clark, H. H. and E. F. Schaeffer (1989). Contributing to Discourse. *Cognitive Science 13*, 259–294.

Farkas, D. F. and K. B. Bruce (2010). On Reacting to Assertions and Polar Questions. *Jounal for Semantics 27*, 81–118.

Gaudou, B., A. Herzig, and D. Longin (2006). Grounding and the expression of belief. In *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning*, KR'06, pp. 221–229. AAAI Press.

Incurvati, L. and J. J. Schlöder (2017). Weak Rejection. *Australasian Journal of Philosophy 95:4*, 741–760.

Krifka, M. (2015). Bias in Commitment Space Semantics: Declarative questions, negated questions, and question tags. In Sarah D'Antonio, Mary Moroney, and Carol Rose Little (Ed.), *Proceedings of SALT 25*, pp. 328–345.

Sidnell, J. (2010). *Conversation Analysis. An Introduction*. Wiley-Blackwell.

[15]There are, however, a few papers that propose formalism based on Baltag et al. (1998), for instance Gaudou et al. (2006) Asher and Lascarides (2008), and Venant and Asher (2015). These can easily accommodate knowledge and belief for that matter (in fact, Gaudou et al. (2006) and Asher and Lascarides (2008) include belief). These differ from the current proposal in their treatment of commitments: unlike here, commitments are treated as usual modalities – something we cannot say to be the case here.

Stalnaker, R. C. (1999). Assertion. In *Context and Content. Essays on Intentionality in Speech and Thought*, pp. 78–95. Oxford University Press.

Stalnaker, R. C. (2002). Common Ground. *Linguistics and Philosophy 25*, 701–721.

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press.

Venant, A. and N. Asher (2015). Dynamics of Public Commitments in Dialogue. In *Proceedings of the 11th International Conference on Computational Semantics*, pp. 272–282.