Aligning intentions: Acceptance and rejection in dialogue[1]

Julian J. SCHLÖDER — *ILLC, University of Amsterdam*
Antoine VENANT — *IRIT, Université Paul Sabatier*
Nicholas ASHER — *CNRS, IRIT*

Abstract. This paper presents an operational and grounded notion of *intention in dialogue* and links it to the *commitments* that speakers make in dialogue. We take these two concepts to then develop a conceptually sound way of doing formal pragmatics. Our model tackles a number of relevant phenomena: (i) we formally derive the illocutionary forces of the speech acts of *asserting* and *rejecting* a proposition; (ii) we give a suitable semantics to rejections of arbitrary speech acts, including rejections of rejections; (iii) we demonstrate how *cooperativity* is linked to how strongly the notion of *speaker commitment* is understood. That is, how tightly bound speakers are by their commitments directly influences how cooperative they are.

Keywords: dialogue, commitment, intention, acceptance, rejection

1. No mindreading

Nobody knows what's in someone else's head. This is a simple truism (Lepore and Stone, 2014). Yet, in conversation, we reason about our interlocutors' thoughts, desires and intentions—while the truth of these internal mental states remains obscure to us. This is a curious tension and it begs for resolution. At least two challenges are at hand here: in what sense can intentions be *public knowledge* as required for successful conversation, and where does that knowledge *come from.* The intuitive solution to both challenges lies in the follow-up truism that *we can only observe how other people behave*—and then draw conclusions about their mental states.

One promising approach, developed the furthest by Asher and Lascarides (2008, 2013), is to supply the interpreter of an utterance with some *default expectations* about what people engaged in dialogue intend. A simple example is that *askers of questions typically intend to get an anwer* and the less simple dual *utterances following questions are typically intended as answers*. Our main interest in this paper lies with the pragmatic principles underlying the sharing of knowledge. Asher and Lascarides provide us with the following default principles governing this process.

$$\textbf{(Intent to Share Commitment)} \quad C_A\varphi > C_A I_A C_B \varphi.$$

$$\textbf{(Cooperativity)} \quad C_A I_A \varphi > I_B \varphi.$$

Here, $>$ denotes a default conditional ($p > q$ expressing roughly that 'if p, then typically q'), $C_A\varphi$ means that the speaker A is publicly *committed* to the formula φ and $I_A\varphi$ means that the speaker A *intends* to establish a state that brings about φ. The first axiom expresses that, typically, commitments are intended to be shared. The second axiom expresses that, typically,

[1] Julian J. Schlöder has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. 607062. Nicholas Asher is supported by ERC grant no. 269427.

publicised intentions are shared. The sharing of commitment is supposed to be part of the purpose of an assertion (Stalnaker, 1978) and the uptake of intention to be a fundamental part of conversation (Clark, 1996). If we assume that a speaker A asserting the proposition p entails that A undertakes a commitment to p ($C_A p$), we can compute *agreement* as follows ($\approx\!\!\!\mid$ is a defeasible inference building on $>$).

$C_A p$ (A asserts that p).

$\models_{\approx}$ $C_A I_A C_B p$ by (Intent to Share Commitment).

$\models_{\approx}$ $I_B C_B p$ by (Cooperativity).

$\rightsquigarrow$ B will agree.

This is an appropriate derivation, but a few things are left open. First, regarding the meaning and truth-conditions of the operators I_A/I_B: while Asher and Lascarides (2003, 2008) give a modal semantics to these operators, it is unclear what this semantics represents. As said above, intentions are inherently private to the interlocutors; so it is difficult to say when $I_A\varphi$ is true and, if it is, what grounds its truth. We want to find a *grounded* and *operational* notion of intention. By *grounded* we mean a notion that ties in with observable behaviour; by *operational* a notion that is suitable for inference and admits motivated truth-conditional scaffolding.

Second, we wonder about the status of *Intent to Share Commitment* and *Cooperativity* as axioms. The question is what licenses their stipulation and in what sense, if any, they express truths. One can surely take them to be non-trivial *empirical* facts, but we think it more principled to derive them from more basic principles. Third, the above representation leaves open what happens in noncooperative situations: the *Cooperativity* axiom is crucial to the derivation above, but it is unclear what replaces it in noncooperative settings. This ties in with the second issue; we seek a more basic principle that generalises *Cooperativity* and reduces to it in appropriate circumstances.

Thus, our goal is to buttress a theory of intentions in formal pragmatics on conceptually solid footing. We proceed towards that goal as follows. In the next section we give a discussion of our conception of dialogue intentions as *commitments to preferred futures* and address some conceptual and practical challenges to this conception. We will show how this notion allows us to do formal pragmatics by formally defining the *basic principles* that govern intentionality in dialogue. Afterwards, we can derive some previously stipulative facts from them, including the illocutionary forces of assertion and rejection; also, we establish a link from cooperativity to commitment. Before concluding, we outline a truth-conditional semantics backing our derivations.

2. Intentions and commitment

Research on intentions divides up into two camps: one that studies intentions as attitudes towards actions or plans (Bratman, 1987) and one according to which intentions are propositional attitudes (Cohen and Levesque, 1990). Given that dynamic semantics (Groenendijk and Stokhof, 1991), in which the contents of a sentence form an action to update an information context, has blurred the line between actions and propositions, our model will have features of both approaches to intention. We will model intentions as being towards achieving a certain propositional *state*. Such states, naturally, are achieved through action, but we do not model an intention to the action.[2]

As mentioned, our concern lies with the meaning of intentional operators like I_A above. We first observe that having an intention has an inherently temporal component: an intention relates to some (potential) future act. Furthermore, intending an act or state α includes a certain dedication to undertake that act or bring forth that state. This is a stronger attitude than merely *preferring* α over not-α or desiring or wanting α. For a simple example, consider someone who desires to go out for fine dining, but cannot afford to do so. It is perfectly reasonable to attribute the cognitive attitude "she wants to go out, but chooses not to" to someone, but "she intends to go out, but chooses not to" strikes us as more than just a little odd.

We can conceptualise this dedication to action as a notion we already understand: *commitment*. We can say that an intention to α can be represented as a *commitment* (to oneself) to bring about α and a publicised intention to α as a *public commitment* (to one's interlocutors) to bring about α. It is useful to compare this to the felicity conditions for speech acts by Austin (1962). According to him, a successfully performed speech act must be made with the right intentions and, moreover, these intentions must be apparent to one's addressees so they can take them up properly. In the terminology we have started to develop here, a speech act publicises a commitment to a (potential) future state or act; uptake consists in the addressees sharing the commitment. Only after all participants have thereby fulfilled the intentional felicity conditions does the illocutionary force of the act apply. We will return to a model for illocutionary force in the next section.

Thus, we can express an intentional operator derivatively from operators we already understand quite well: the operators for *commitment* (C_A, C_B) and *temporal modal operators* which we will put as $\Diamond$ ('eventually') and $\Box$ ('from now on', equivalent $\neg\Diamond\neg$).[3] We can thus represent 'A intends that α' as $C_A\Diamond\alpha$ ('A commits that eventually α obtains'). There is, however, a sleight of hand in this representation: we only understand commitment as *public* commitment, but have no grasp on (not publicised) commitments to oneself. This, however, is not a problem for our project. As far as intentions are required for the understanding of a speech act, they ought to be (implicitly or explicitly) publicised. In fact, the intentions going along with a speech act must be *apparent* to the addressee of that act.[4] Making explicit how implicit intentions are apparent from observable behaviour is the major drive behind the pragmatics we develop in the next section.

[2] We ignore here certain pathological cases where someone does the right thing for the wrong reasons; e.g. intending that α, executing an ill-conceived plan to bring forth α, and just by chance achieve α in the process nonetheless.

[3] More complex temporal logics such as LTL (Pnueli, 1977) would arguably be a more appropriate choice here. For simplicity we remain with the normal modal logic language. We formalise temporal modal logic below as KT4; we do not need additional machinery, and KT4 seems to express universal truths about temporal reasoning.

[4] In noncooperative settings the apparent intentions might be purposefully misleading; nonetheless, one needs to display *some* intention (even if it is dishonest) when making a speech act.

For now, we must address a more pressing concern. Our semi-formal logical form $C_A\Diamond\alpha$ for the intention that α is indistinguishable from the claim that α *will happen* (if one understands claiming that p as the undertaking of a commitment to p). Surely, claiming that something will happen is much stronger than intending it. The difference is, however, smaller than it appears. As argued above, an intention α incurs a dedication to bringing forth α. It is a minor step from there to conclude that the publicised intention to α amounts to the claim that one will cause α unless something unexpected happens. As a case in point, notice that it is absurd to intend something impossible.[5] In this sense, we can understand intentions as *defeasible* commitments: they reduce to the claim that the content of the intention will obtain *unless* circumstances unexpectedly prevent this.

With the preliminaries settled, we can now formulate a principle that we take to be *partly constitutive* of commitment in dialogue. Undertaking a commitment includes a certain dedication to upholding the commitment. Quite literally, a commitment that is undone without much ado is not a commitment. Thus we consider it a first principle of commitment that one does not intend to undo the commitment. We can therefore stipulate the following as fundamental axioms for our pragmatics.

Definition 1 (Basic Principles)**.** The following axioms are (partly) constitutive of commitment:

(a) $C_S\varphi > C_S\Box C_S\varphi$ (for each speaker S).

(b) $C_S\neg C_S\neg\varphi > C_S\Box\neg C_S\neg\varphi$ (for each speaker S).

(c) $C_S\Diamond C_S\Diamond\varphi > C_S\Diamond\varphi$ (for each speaker S).

The principle (a) states that if one makes a commitment, one intends to keep it (equivalently, does not intend to break it; $C_S\neg\Diamond\neg C_S\varphi$). Principle (b) states that if someone commits that they *might* commit to something (understanding $\neg C_S\neg$ as the dual to C_S), they also intend to keep that commitment. Note the difference to $\neg C_S\neg\varphi$ which merely expresses that φ is compatible with S's commitments; in this case, S can easily commit to either φ or $\neg\varphi$ later without dropping any earlier commitment. In contrast, $C_S\neg C_S\neg\varphi$ expresses that S has *committed* to potentially also committing to φ—hence committing to $\neg\varphi$ requires dropping this earlier commitment. Finally, principle (c) expresses that an intention to intend something reduces to a single intention.

Note that we model intentions as intending *states*, not *acts*. If someone intends an act, completion of that act should resolve the intention. In our case, with speakers intending states, once the state is achieved the intention is preserved as the intention to *maintain* that state.

These axioms have privileged epistemic status over the principles *Intent to Share Commitment* and *Cooperativity* we discussed in the Introduction. The principles in Definition 1 are *constitutive* of commitment: that one intends to keep a commitment lies at the core of what it means to commit. We also maintain the Basic Principles in noncooperative settings; if commitment is the

[5] There might be situations where a speaker claims to intend something objectively impossible; in such cases, the speaker is mistaken. This seems to be similar phenomenon to someone claiming to *know* something false.

basic operation in dialogue and these principles are foundational to commitment they need to be maintained (at least as a pretense) even in noncooperative dialogue.[6] A prototypical noncooperative action is a lie; lies only work since the defeasible inference from public commitment to private belief is active even in noncooperative settings (i.e. *ceteris paribus* both parties are still assumed to speak truths). Similarly, the Basic Principles allow the noncooperative move of making false commitments—commitments one has no intention of keeping (e.g. an unfaithful promise).[7] The possibility to be insincere extends to publicised intentions as well: one can commit to intending something without actually intending it—hence (c) is a default as well.

To complete the picture it is left to give meaning to the temporal operators. As for the present model we only require future-oriented operators, we characterise them as the normal modal logic KT4.

Definition 2 (Temporal Logic)**.** For the sake of simplicity, we model temporal operators as KT4 modal logic:

(K) $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$.

(T) $\varphi \rightarrow \Diamond\varphi$. (If something is true now, it is also true eventually.)

(4) $\Diamond\Diamond\varphi \rightarrow \Diamond\varphi$. (If it is eventually the case that eventually φ, then eventually φ.)

3. Formal pragmatics of acceptance and rejection

In the previous section we have developed the foundation to our formal pragmatics. We now develop a framework on top of this foundation that allows us to model the pragmatics of acceptance and rejection in dialogue.

3.1. Asserting and rejecting propositions

The first task is to define what asserting and rejecting a proposition means with respect to public commitment. As already hinted at in Section 1, it is intuitive to think of an *assertion that p* as the undertaking of a commitment that *p* holds. This intuition sketches a general plan to capture the semantic effects of *A* asserting *p*: one needs to update a *conversational record* with the information that *A* committed to *p*.

While this intuitive idea constitutes a standard line of research (Brandom, 1983; Asher and Lascarides, 2008), putting it to work requires us to make some fundamental modeling choices: primarily, we need to define what the conversational record *is* and how it *changes* during conversation. That is, we need a formal representation of conversational records from which com-

[6] This corresponds roughly to what Asher and Lascarides (2013) call *basic cooperative*: the minimal cooperativity required to engage in conversation at all.

[7] One can argue that such insincere commitments are not commitments at all. We agree in principle, but the insincere speaker still *feigns* a commitment. And since we are concerned with observable behaviour only, we need to deal with such deceit and the consequences it has in a dialogue.

mitments can be easily projected. Then we need to capture the *immediate* effects of an assertion on this record; making the model too smart, so to speak, would go against our project of finding minimal, constitutive principles.

We can rely on existing work for this. Commitments are representable in static modal logic models (Asher and Lascarides, 2008), and Venant and Asher (2015) show how assertions enact *dynamic* updates of these models. Their model in particular accounts for updates associated with assertions that dispute previous commitments (undertaken by oneself or others). Furthermore, Venant and Asher (2016), building on Asher and Lascarides (2003), give a general method to link such models to a complex language for semantic representation and interpretation in a discourse context.

It is however not obvious how one ought to understand the *rejection of p* in such a framework. There is a clear intuition that rejection operates dually (in some way) to assertion; that is, rejection is the speech act that cancels or reverses an assertion. An old idea is to reduce rejection to assertion by claiming that a rejection of p is equivalent to the assertion of $\neg p$ (Frege, 1952; Rumfitt, 2000). This idea does not seem to capture the breadth of linguistic data. For instance, Grice (1991) provides us with an example where a rejection of a proposition p is not equivalent to asserting that that p is false.[8]

(1) If you say "X or Y will be elected," I may reply "That's not so; X or Y *or* Z will be elected." Here, too, I am rejecting "X or Y will be elected" not as false but as *unassertable*. (Grice, 1991: p. 82, his emphasis)

Similarly, Stalnaker (1978) remarks in a footnote that to reject is not to assert a negative, but to refuse to accept (p. 87, footnote 9). Given Grice's example and similar cases we agree with Stalnaker on this. Incurvati and Schlöder (2017) give multiple equivalent ways to formalise this conception; their commitment account models a speaker A rejecting p as undertaking a commitment *not* to commit to p, $C_A \neg C_A p$. Our Basic Principles (Definition 1) predict that this is correct: $C_A \neg C_A p$ entails—by the Basic Principles—$C_A \Box \neg C_A p$. This can be read as A publicising the intention not to commit to p and thereby not to agree to p. This coincides with Stalnaker's remark and with our intuition that a rejection is the dual speech act to assertion.

Definition 3 (Assertion and Rejection)**.** Let A be a speaker and p be a proposition.

In asserting that p, A undertakes the commitment that $C_A p$.

In rejecting p, A undertakes the commitment that $C_A \neg C_A p$.

3.2. Deriving illocutionary force

We now have described what commitments a speaker undertakes in performing the speech acts of assertion and rejection. The next question is what the *purpose* of these speech acts is. We

[8] Others have made similar arguments, e.g. Dickie (2010) and Walker (1996).

follow Austin (1962) in saying that one makes a speech act with a particular intention. We can formalise this as follows.

Definition 4 (Speech-Act Related Goals)**.** If a speaker A makes a speech act and we can derive from the resulting commitment structure that $C_A\Diamond\gamma$ then γ is a *Speech-Act Related Goal (SARG)* of that act.

Trivially, according to the Basic Principles in Definition 1, it is a SARG of any speech act to maintain the commitments undertaken by making that act. However, committing to a proposition is rarely if ever a goal in and of itself, so there must be something more to their purpose. To answer this question, we first must ask what the overall purpose of conversation itself is; since speakers make speech acts to engage in conversation, their purpose can only be understood with respect to the purpose of the conversation.

The traditional conception has it that the goal of a conversation is to *exchange information* (Stalnaker, 1978; Skyrms, 2010). In other work (Asher et al., 2016; Asher and Paul, 2016), some of us have argued that conversations have many purposes that are not plausibly analysable as an exchange of information. For example, two politicians in a debate are often not exchanging information with each other but rather setting forth arguments and providing information to convince an audience, which might be not necessarily they themselves, but their peers, a committee, or an electorate. Thus, Asher et al. (2016), Asher and Paul (2016) abstract away from all these *particular* circumstances and conceive of conversations, at utmost generality, as being about achieving certain winning conditions with respect to an abstract *Jury*.

For the purposes of this paper, we will assume that these winning conditions are expressible in a standard propositional language and thus can be put as the goal to *convince the Jury*. On our commitment account, agreement between two interlocutors is to share a commitment. Hence we can fix another basic principle: both speakers engaged in a dialogue *intend to establish their commitments as shared with the Jury*. While in Definition 1 we recorded basic principles of *commitment*, we now fix this principle as partly constitutive of the activity of *conversing*.

Definition 5 (Basic Principles)**.** Let A and B be speakers engaged in conversation and let J be the Jury; let Φ be the finite set of issues raised in the dialogue. The following axiom is (partly) constitutive of conversing:

(d) $C_A\Diamond\Box\bigwedge_{\varphi\in\Phi}(C_A\varphi \leftrightarrow C_J\varphi)$ and $C_B\Diamond\Box\bigwedge_{\varphi\in\Phi}(C_A\varphi \leftrightarrow C_J\varphi)$.
Speakers intend to reach agreement with the Jury on the issues Φ.

The Basic Principle (d) states that both speakers are committed to aligning with the Jury on all issues φ raised in the conversation. This does not mean to agree to φ, but rather to reach a state where the Jury *shares* a commitment with regards to φ; that is, either both the speaker and the Jury commit to φ or both do not commit to φ. This principle is an idealisation. First, dialogues might be aborted at any point and so leave issues open. Second, speakers might have an argument just for the fun of it (as in debate club), or to convince the Jury of their conversational *ability* and not primarily of the *issues* they raise (as in debate competitions).

The first point does not impugn our general point that it is still their *goal* to resolve issues; the second point concerns situations that are, arguably, still conducted under a pretense of adherence to (d).

Now we are able to derive some particular SARGs of assertion and rejection. The following derivations show that the goal of an assertion is to prompt agreement, and the goal of a rejection is to prompt retraction of a previous commitment (if there was one). First suppose that speaker A asserts the proposition p.[9]

$C_A p$ (A asserts that p).

$\mid\!\approx$ $C_A \Box C_A p$ by Definition 1, (a).

$\mid\!\approx$ $C_A \Box\Box C_A p$ by Definition 2, (4).

$\mid\!\approx$ $C_A \Diamond\Box C_J p$ by Definition 5, (d).

$\mid\!\approx$ $C_A \Diamond C_J p$ by the dual of Definition 2, (T).

$\rightsquigarrow$ A wants the Jury to share the commitment to p.

Now suppose that speaker B rejects p.

$C_B \neg C_B p$ (B rejects p).

$\models$ $C_B \neg C_B \neg\neg p$ by Double Negation Introduction.

$\mid\!\approx$ $C_B \Box\neg C_B \neg\neg p$ by Definition 1, (b).

$\mid\!\approx$ $C_B \Box\neg C_B p$ by Double Negation Elimination

$\mid\!\approx$ $C_B \Box\Box\neg C_B p$ by Definition 2, (4).

$\mid\!\approx$ $C_B \Diamond\Box\neg C_J p$ by Definition 5, (d).

$\mid\!\approx$ $C_B \Diamond\neg C_J p$ by the dual of Definition 2, (T).

$\rightsquigarrow$ B wants the Jury to be not committed to p.

These two SARGs seem to be of particular significance with respect to the Basic Principle (d), and therefore express the conventional purpose that assertion and rejection serve in a dialogue. Thus, we call them the *illocutionary forces* of the speech acts assertion and rejection.

A special case of the latter derivation is in the situation where A has already convinced the Jury of p. Then B's intention is for the Jury to *retract* the commitment to p. This is in particular the case in everyday conversation where the Jury is the speakers themselves; that is, B wants to convince A and A wants to convince B. Then we derive $C_A \Diamond C_B p$ as the illocutionary force of

[9] Again $\mid\!\approx$ is defeasible inference with defeasible modus ponens on $>$. That is, $p > q, p \mid\!\approx q$ unless that inference is blocked by information contradicting q.

assertion and $C_B\Diamond\neg C_A p$ as the illocutionary force of rejection. Hence, if A asserts that p, B's rejection of p voices the intention that A retract her commitment to p. Moreover, this means we derive $C_A p \mathrel{|\!\!\approx} C_A\Diamond C_B p$ from basic principles, instead of stipulating the axiom *Intent to Share Commitment* we discussed in Section 1.

3.3. Rejecting rejections and other speech acts

The previous discussion dealt exclusively with rejecting a (previously asserted) proposition. But, in fact, any speech act can be rejected, cancelling its effects (Searle, 1969). A particular case is the rejection of a previous rejection.

(2) a. A: It is the case that p.
b. B: No, it's not.
c. A: Yes, it is.

It is not obvious—in English—that A's utterance in (2c) is a rejection of B's rejection in (2b), rather than a re-assertion of A's own (2a). However, (2) can be translated to German as follows.

(3) a. A: Es ist der Fall, dass p.
b. B: Nein, ist es nicht.
c. A: Doch, ist es.

In this case, A's utterance in (3c) contains the negative polarity item *doch* which requires a negated antecedent. Thus, the utterance that (3c) must rhetorically relate to is (3b) and it is implausible to model it as a re-assertion of p. English lacks a particle like *doch*, but in some cases a rejection syntactically mirrors the shape of a previous rejection. The following are attested examples from the AMI corpus (Carletta, 2007).

(4) a. A: That's dependent on the television.
b. B: No, I don't think so.
c. A: I do know so.

(5) a. A: Mushroom is a vegetable.
b. B: I don't think it is.
c. A: It's vegetable.

In (4), the form of (4c) mirrors (4b), suggesting that (4c) functions as a rejection of the rejection (4b). In contrast, the form of (5c) suggests that it indeed works as a re-assertion of (5a). We note that it is not our goal to argue that rejection-of-rejection and re-assertion have different *effects*; surely, both (4c) and (5c) have the effect of A indicating that she stands by her original assertion. Rather, we want to verify that our approach squares with the data showing that one can reject a rejection. That is, we check that we can make sense of these cases without engaging in undue sleights of hand (such as simply defining rejections-of-rejections to be re-assertions).

Naïvely reproducing the semantics for rejection from Definition 3 will not do. In (6) we have added the resulting commitments, according to Definition 3, to the dialogue from (2). That is, (6a) results in A undertaking a commitment to p and (6b) in B undertaking a commitment not to share A's commitment. However, representing the rejection-of-a-rejection in (6c) as A committing not to share in B's commitment is odd.

(6) a. A: It is the case that p. $C_A p$
b. B: No, it's not. $C_B \neg C_B p$
c. A: Yes, it is. $??C_A \neg C_A \neg C_B p$

Apparently, the way we characterised rejection in Definition 3 is specific to the rejection of a proposition. Rejecting an *act* (other than assertion) appears to be different. Let's try to make sense of this. Speech acts can be *cancelled* which means that their force does not obtain (Searle, 1969), i.e. that the associated intention is not fulfilled. We model rejection now as the speech act that realises the function of cancellation. That is, what a rejection cancels is the *illocutionary force* of its target act; or, more generally, a rejection can target a SARG. Put differently, in a Clarkian (Clark, 1996) understanding of conversation, every move in a dialogue proposes a joint action; execution of the action realises the move's illocutionary force. Thus we can say that to reject a speech act is to *refuse participation* in that act's joint action.

Definition 6 (Rejecting Arbitrary Speech Acts)**.** Suppose a is a speech act with SARG α. To reject that SARG is to commit not to participate in α; that is, not to intend α. Thus, if speaker B rejects α, B is undertaking a commitment to $C_B \neg \Diamond \alpha$. We assume that all else being equal the SARG of a that represents a's illocutionary force is the relevant SARG here.

We can first verify that this generalises on the narrower definition from Definition 3. That is, we can reproduce that rejecting an assertion of p (as in 2b) results in the commitment $C_B \neg C_B p$, i.e. in rejecting the asserted proposition. Since here we talk about the speakers engaging in joint actions, we can model the Jury as the speakers themselves. Then the illocutionary force of an assertion that p is $C_A \Diamond C_B p$. By to Definition 6 the rejection of that act is $C_B \neg \Diamond C_B p$. Hence:

$C_B \neg \Diamond C_B p$ (B rejects $\alpha = \Diamond C_B p$).

$\models$ $C_B \Box \neg C_B p$ by a modal logic validity.

$\models$ $C_B \neg C_B p$ by the dual of Definition 2, (T).

Note in particular that these inferences are non-defeasible; that is, it is an unavoidable consequence of rejecting an assertion (in the sense of Definition 6) to reject the asserted proposition (in the sense of Definition 3).

Now we can check that Definition 6 also felicitously models rejections-of-rejections. In (6b), the illocutionary force we attribute is $C_B \Diamond \neg C_A p$ (as derived earlier). Rejecting this force yields:

$C_A \neg \Diamond \neg A C_A p$ (A rejects $\alpha = \Diamond \neg C_A p$).

$\models$ $C_A \Box C_A p$ by a modal logic validity.

$\models$ $C_A C_A p$ by the dual of Definition 2, (T).

$\rightsquigarrow$ A is committed to her commitment that p.

The intuitive reading of this derivation is that A is confirming or renewing her commitment to p as made in (6a). We think that this is correct for rejecting a rejection.

3.4. Cooperative commitments

Everything said so far goes for cooperative as well as noncooperative settings. Our final goal is to reproduce the *Cooperativity* axiom from Section 1. This axiom models that cooperative speakers adopt publicised intentions. Put differently, once an addressee has *understood* the SARG of an act, she is expected to *take up* that goal. We now demonstrate that this kind of cooperativity varies with the assumptions made on the strength of commitments. Consider the following three modal logic axioms.

$$\text{(D) } \neg C_A\bot \qquad \text{(4) } C_A\varphi \rightarrow C_A C_A\varphi \qquad \text{(5)}\neg C_A \rightarrow C_A\neg C_A\varphi.$$

These axioms formalise some sort of sincerity on commitments: (D)+(5) expresses that one cannot claim to have made a commitment (C_AC_Ap) that one in fact has not made ($\neg C_Ap$); (D)+(4) that a speaker cannot make a commitment to a proposition (C_Ap) and simultaneously reject it ($C_A\neg C_Ap$). To see that these partially express cooperativity, note that (D)+(5) rule out the noncooperative act in (7a) and (D)+(4) the one in (7a′).

(7) a. A: I claimed earlier that p. C_AC_Ap
(when, in fact, A never asserted p, i.e. $\neg C_Ap$*)*
a.′ A: I never claimed that p. $C_A\neg C_Ap$
(when, in fact, A asserted p previously, i.e. C_Ap*)*

Claims similar to (7) can be found, e.g., in political debates. This demonstrates that in noncooperative settings one cannot trust one's interlocutor to be sincere about their own commitments. Thus, we refer to KD45 operators C_A as *cooperative commitments*.[10] We now show that if we axiomatise our commitment structures to be KD45 models, we derive the cooperative principle that *understanding entails acceptance*; this inference is of course only made defeasibly, since rejections can occur for innocuous reasons in cooperative settings as well. This precisely formalises the claim of Schlöder and Fernández (2014) that a cooperative addressee is expected to take up a proposal unless there is a reason not to. Apparently, speakers are aware of this default expectation, as evinced by the following attested example from the British National Corpus (Burnard, 2000).

(8) a. A: Do you agree with that?
b. B: I have no reason to disagree. Yes.
(BNC, file FMN, lines 492–493)

First we need to clarify what is meant by 'understanding' on the commitment-based account. Venant et al. (2014) already provide us with a model: understanding is also an attitude that ought to be displayed publicly if it is to be conversationally relevant (this appears to be the

[10] Incurvati and Schlöder (2017) also derive KD45 as the modal logic of cooperative commitment from independently motivated principles.

point of backchannel utterances; also recall our discussion of publicised intentions). Making something public, on the commitment-based account, is always the undertaking of a commitment. Now, according to Venant et al., B understanding that A committed to φ is precisely B committing to the fact that A made that commitment, $C_B C_A \varphi$. We adopt this notion and can now demonstrate that under cooperative commitments, understanding an intent defeasibly entails sharing the intent.

Theorem 7 (Cooperative Commitment)**.** Assume that C_A and C_B satisfy the KD45 axioms and that A considers B her Jury and vice versa. Then $C_B C_A \Diamond\varphi \mathrel{|}\!\approx C_B \Diamond\varphi$ (and the same for A and B switched).

Proof. $C_B C_A \Diamond\varphi$ (assumption).

$\mathrel{|}\!\approx$ $C_B \Box C_B C_A \Diamond\varphi$ by Definition 1, (a).

$\mathrel{|}\!\approx$ $C_B \Box\Box C_B C_A \Diamond\varphi$ by Definition 2, (4).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A C_A \Diamond\varphi$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A \Diamond\varphi$ by (D)+(5).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_B \Diamond\varphi$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond C_B \Diamond\varphi$ by Definition 2, (T).

$\mathrel{|}\!\approx$ $C_B \Diamond\varphi$ by Definition 1, (c). □

Note that Theorem 7 is an accurate analogue of the *Cooperativity* axiom of Asher and Lascarides.

For the two speech acts we have defined here, we can do one better. Theorem 7 shows that understanding a *goal* entails uptake of that goal. The following derivations show that for assertion and rejection it suffices to understand the *act* itself.

$C_B C_A p$ (B understands A's assertion of p).

$\mathrel{|}\!\approx$ $C_B \Box C_B C_A p$ by Definition 1, (a).

$\mathrel{|}\!\approx$ $C_B \Box\Box C_B C_A p$ by Definition 2, (4).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A C_A p$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A p$ by (D5).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_B p$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond C_B p$ by Definition 2, (T).

$\rightsquigarrow$ B intends to share the commitment to p.

$C_B C_A \neg C_A p$ (B understands A's rejection of p).

$\mathrel{|}\!\approx$ $C_B \Box C_B C_A \neg C_A p$ by Definition 1, (b).

$\mathrel{|}\!\approx$ $C_B \Box\Box C_B C_A \neg C_A p$ by Definition 2, (4).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A C_A \neg C_A p$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box C_A \neg C_A p$ by (D5).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box \neg C_A p$ by (D4).

$\mathrel{|}\!\approx$ $C_B \Diamond\Box \neg C_B p$ by Definition 5, (d).

$\mathrel{|}\!\approx$ $C_B \Diamond \neg C_B p$ by Definition 2, (T).

$\rightsquigarrow$ B intends not to commit to p.

The assumption that A and B consider each other their Jury stands to reason in everyday cooperative dialogue. Depending on how one models the Jury, the above derivations and Theorem 7 can also be recovered in a more general setting. For instance, if the Jury is *also* considered to be bound by the Basic Principles (a)–(d), Theorem 7 can be maintained, since then intentions percolate from A over J to B. However, in general, the Jury is a mere overhearer of the conversation and not a participant. So it is potentially bound by a different set of principles.

4. Model theory

A model theory can be obtained by understanding the speech acts we consider as moves in an unbounded conversational game (Asher and Paul, 2013). We start with both participants initial commitments, as represented in a Kripke model satisfying the appropriate axioms. Computing a speech act leads to a model transition; we abstract away from explicitly computing these transitions (though it can be done; see Venant and Asher (2015)). Rather, we consider a conversation to be a sequence of public commitment models, where we tacitly assume that the transition from one model to the next was due to one participant undertaking a commitment by way of making speech act.

Thus, the complete tree of potential conversations is a tree in which each dot is a Kripke model for commitment. This allows us to understand the temporal operators $\Diamond$ and $\Box$ as quantifying over potential future commitment states. To define defeasible inference we modify Commonsense Entailment (Asher and Morreau, 1990) to include a temporal dimension.

Definition 8 (Point)**.** Let W be a set of worlds. A *point* on W is a model for three speaker public commitment modal logic (A, B and the Jury J), i.e. a structure $t = \langle W, V^t, R^t_A, R^t_B, R^t_J \rangle$ where V is a valuation and R^t_A, R^t_B and R^t_J are accessibility relations.

Definition 9 (Intention Model)**.** A *intention model* is a tuple $M = \langle W, P, S, T, * \rangle$ where:

- W is a set of worlds.
- P is a set of points on W.
- $S \subseteq P^{<\omega}$ is a tree. That is, $\emptyset \in S$ and S is a set of finite sequences on P such that if $x \in S$ and y is an initial segment of x, then also $y \in S$. If $x \in P^{<\omega}$, $x = \langle t_1, t_2, ..., t_n \rangle$, we write $x \oplus t = \langle t_1, t_2, ..., t_n, t \rangle$ for the extension of x with t.
- T is a function that maps an $x \in S$ and a $w \in W$ to a maximal branch in S extending x i.e. $T(x,w) \subseteq S$ is a set of finite sequences such that:
  - $x \in T(x,w)$ and for all $y \in T(x,w)$, x is an initial segment of y.
  - $\bigcup T(x,w)$ is infinite or there is no $t \in P$ such that $(\bigcup T(x,w) \oplus t) \in S$.
  - $T(x \oplus t, w) = T(x,w) \setminus \{x\}$.
- $*$ is a function $* : W \times \mathscr{P}(S \times W) \to \mathscr{P}(S \times W)$ with:

- for all $w \in W$ and $X \subseteq S \times W$, $*(w,X) \subseteq X$.
- for all $w \in W$ and $X,Y \subseteq S \times W$, $*(w,X \cup Y) \subseteq *(w,X) \cup *(w,Y)$.

For $x \in S$ write $M^x = \langle W, V^x, R^x_A, R^x_B, R^x_J \rangle$ for the last point in x.

Intuitively, S is represents possible conversations along transitions between the commitment structures P. A finite sequence represents a finite conversation and the last point in the sequence is the current commitment state in that conversation; intuitively T assigns *a future timeline* to every world in that state. Note that time at a world is linear, and the temporal modals are evaluated with respect to the same world at different points in time.

We use T to give truth-conditions to the temporal operators and $*$ to give truth-conditions to defeasible conditionals. The function $*$ assigns to a world and a set of sequence-world pairs (i.e. the extension of a proposition, computed globally on the entire tree) the set of sequence-world pairs where the proposition holds in a typical ('normal') manner.

Definition 10 (Truth)**.** Truth is defined in a model $M = \langle W, P, S, T, * \rangle$ relative to an $x \in S$.

- $M,x,w \models p$ iff $w \in V(p)$,
- $M,x,w \models \neg\varphi$ iff $M,x,w \not\models \varphi$,
- $M,x,w \models \varphi \wedge \psi$ iff $M,x,w \models \varphi$ and $M,x,w \models \psi$,
- $M,x,w \models C_S\varphi$ iff for all $v \in R^x_S(w)$, $M,x,v \models \varphi$,
- $M,x,w \models \Diamond\varphi$ iff there is a $y \in T(x,w)$ such that $M,y,w \models \varphi$.
- $M,x,w \models \varphi > \psi$ iff $\forall (y,v) \in *(w, \{(y,v) \in S \times W \mid M,y,v \models \varphi\})$, $M,y,v \models \psi$.

Write $[\![\varphi]\!] = \{(y,v) \in S \times W \mid M,y,v \models \varphi\}$ for the global extension of φ.

Note that this definition of $\Diamond$ satisfies KT4—and quite a lot more. The assumptions we have made in the preceding discussion were intentionally chosen to be minimal, *basic* assumptions. They are not *complete* for this model theory, as this would require some stronger assumptions. We leave an investigation into these for further work. Our goal here is to demonstrate that this semantic framework is expressive enough to define admissibility conditions corresponding to our basic principles.

Definition 11 (Semantic Axioms for the Basic Principles)**.** The Basic Principles correspond to the following structural axioms on intention models.

(a) $C_A\varphi > C_A\Box C_A\varphi$.

$$\forall X \subseteq \mathscr{P}(S \times W),\ w \in W:$$
$$*(w, \{(y,v) \in S \times W \mid \forall v' \in W : vR^y_A v' \to (y,v') \in X\})$$
$$\subseteq \{(y,v) \in S \times W \mid \forall v' \in W \Big(vR^y_A v' \to \forall y' \in T(y,v')\big(\forall v'' \in W (v'R^{y'}_A v'' \to (y',v'') \in X)\big)\Big)\}.$$

(b) $C_A \neg C_A \neg \varphi > C_A \Box \neg C_A \neg \varphi$.

$$\begin{aligned}&\forall X \subseteq \mathscr{P}(S \times W),\ w \in W :\\ &*(w, \{(y,v) \in S \times W \mid \forall v'' \in W (vR^y_A v'' \to \exists v' \in W : v''R^y_A v' \wedge (y,v') \in X)\})\\ &\subseteq \{(y,v) \in S \times W \mid \forall v' \in W \Big(vR^y_A v' \to \forall y' \in T(y,v') \big(\exists v'' \in W : v'R^{y'}_A v'' \wedge (y,' v'') \in X \big) \Big) \}.\end{aligned}$$

(c) $C_A \Diamond C_A \Diamond \varphi > C_A \Diamond \varphi$.

$$\begin{aligned}&\forall X \subseteq \mathscr{P}(S \times W),\ w \in W :\\ &*(w, \{(y,v) \in S \times W \mid \forall v' \big(vR^y_A v' \to \exists y' \in T(y,v') \forall v'' \in W :\\ &\qquad\qquad (v'R^{y'}_A v'' \to \exists y'' \in T(y',v'') : (y'',v'') \in X) \big) \})\\ &\subseteq \{(y,v) \in S \times W \mid \forall v' \in W \Big(vR^y_A v' \to \exists y' \in T(y,v') : (y',v') \in X \Big) \}.\end{aligned}$$

(d) $C_A \Diamond \Box (C_A \varphi \leftrightarrow C_J \varphi)$.
$\forall x \in S,\ w \in W :\ \forall v (wR^x_A v \to \exists y \in T(x,v) \forall z \in T(y,v) : R^z_A = R^z_J)$.

Analogously for the axioms for speaker B.

The axioms for (a)–(c) look complicated, but are in fact just structural decodings of the truth-conditions of the Basic Principles. The soundness of (d) is trivial, and we present the soundness proof for axiom (a).

Theorem 12 (Soundness)**.** On intention models M where the axioms for (a) from Definition 11 holds: $\forall x \in S\ \forall w \in W : M,x,w \models C_A \varphi > C_A \Box C_A \varphi$.

Proof. Fix x and w. It is to show that $*(w, [\![C_A \varphi]\!]) \subseteq [\![C_A \Box C_A \varphi]\!]$. Instantiate the axiom for (a) for $X = [\![\varphi]\!]$ to obtain the sets:

$$\begin{aligned}&*(w, \{(y,v) \in S \times W \mid \forall v' \in W : vR^y_A v' \to (y,v') \in [\![\varphi]\!]\})\\ =&*(w, \{(y,v) \in S \times W \mid \forall v' \in W : vR^y_A v' \to M,y,v' \models \varphi\})\\ =&*(w, \{(y,v) \in S \times W \mid M,y,v \models C_A \varphi\})\\ =&*(w, \{[\![C_A \varphi]\!]\})\\ \subseteq&\{(y,v) \in S \times W \mid \forall v' \in W \Big(vR^y_A v' \to \forall y' \in T(y,v') \big(\forall v'' \in W (v'R^{y'}_A v'' \to (y',v'') \in [\![\varphi]\!]) \big) \Big) \}\\ =&\{(y,v) \in S \times W \mid \forall v' \in W \Big(vR^y_A v' \to \forall y' \in T(y,v') \big(\forall v'' \in W (v'R^{y'}_A v'' \to M,y',v'' \models \varphi) \big) \Big) \}\\ =&\{(y,v) \in S \times W \mid \forall v' \in W \big(vR^y_A v' \to \forall y' \in T(y,v') (M,y',v' \models C_A \varphi) \big) \}\\ =&\{(y,v) \in S \times W \mid \forall v' \in W (vR^y_A v' \to \forall M,y,v' \models \Box C_A \varphi)\}\\ =&\{(y,v) \in S \times W \mid M,y,v \models C_A \Box C_A \varphi\}\\ =&[\![C_A \Box C_A \varphi]\!]\end{aligned}$$

This is precisely the truth-condition for $C_A \varphi > C_A \Box C_A \varphi$. □

5. Conclusion

The work we have presented here is intended to be foundational. We offer an understanding of *commitment* in conversation that is grounded in the basic principles we take to be partly constitutive of the concept. Moreover, this understanding allows us to straightforwardly understand the elusive notion of *intention* without requiring additional machinery. The result is a model that goes well beyond using commitment as a mere scorekeeping device on the conversational record. By taking *public commitments* as basic observable data, we validate sophisticated inferences from *what is observed* to *what is intended*.

By conceptualising speech acts as the undertaking of particular commitments, the model then explains what illocutionary forces are and what it means to take up a speech act. We can also define the dual of uptake—cancellation—through our general account of rejection: not just as the rejection of a proposition, but as the rejection (cancellation) of an arbitrary speech act. We have demonstrated how this works for the particular case of rejecting a rejection.

Our *Basic Principles* are designed to apply in full generality to cooperative and noncooperative situations. We are able to distinguish cooperativity by modulating additional constraints on what it means to *commit*: by strengthening how tightly speakers are bound by their commitments, we can exclude particular noncooperative moves, and ultimately arrive at the result that cooperative speakers *take up what they understand* (if they can).

In future work we intend to extend this discussion to further speech acts, in particular questions. As mentioned in the Introduction, there are *prima facie* natural intuitions of what speakers *intend* when asking or replying to questions. This project, however, faces some challenges. We first need to include suitable propositional semantics for questions into our (as of yet very simple) commitment structures. Then, it is well possible that there are additional principles that constitute commitment with respect to *questions*; i.e. our Basic Principles potentially underspecify what it means to be committed to a question.

References

Asher, N. and A. Lascarides (2003). *Logics of Conversation.* Cambridge University Press.

Asher, N. and A. Lascarides (2008). Commitments, beliefs and intentions in dialogue. In J. Ginzburg, P. Healey, and Y. Sato (Eds.), *Proceedings of the 12th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, pp. 29–36.

Asher, N. and A. Lascarides (2013). Strategic conversation. *Semantics and Pragmatics 6*, 1–58.

Asher, N. and M. Morreau (1990). Commonsense entailment: A modal theory of nonmonotonic reasoning. In *European Workshop on Logics in Artificial Intelligence*, pp. 1–30. Springer.

Asher, N. and S. Paul (2013). Infinite games with uncertain moves. In *1st International Workshop on Strategic Reasoning*, pp. 25–32.

Asher, N. and S. Paul (2016, July). Evaluating conversational success: Weighted message exchange games. In J. Hunter, M. Simons, and M. Stone (Eds.), *20th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, New Jersey, USA.

Asher, N., S. Paul, and A. Venant (2016). Message exchange games in strategic conversation.

Journal of Philosophical Logic, 1–50.

Austin, J. L. (1962). *How to do Things with Words.* Clarendon Press.

Brandom, R. (1983). Asserting. *Noûs 17*(4), 637–650.

Bratman, M. (1987). *Intentions, Plans and Practical Reason.* Harvard University Press.

Burnard, L. (2000). *Reference Guide for the British National Corpus (World Edition).* Oxford University Computing Services.

Carletta, J. (2007). Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus. *Language Resources and Evaluation 41*(2), 181–190.

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

Cohen, P. and H. Levesque (1990). Intention is choice with commitment. *Artificial Intelligence 42*, 213–261.

Dickie, I. (2010). Negation, anti-realism, and the denial defence. *Philosophical Studies 150*(2), 161–185.

Frege, G. (1952). Negation. In P. Geach and M. Black (Eds.), *In Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.

Grice, H. P. (1991). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Groenendijk, J. and M. Stokhof (1991). Dynamic predicate logic. *Linguistics and Philosophy 14*, 39–100.

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

Lepore, E. and M. Stone (2014). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.

Pnueli, A. (1977). The temporal logic of programs. In *Foundations of Computer Science,*, pp. 46–57.

Rumfitt, I. (2000). “Yes” and “No”. *Mind 109*(436), 781–823.

Schlöder, J. J. and R. Fernández (2014). Clarification requests on the level of uptake. In *Proceedings of the 18th SemDial Workshop on the Semantics and Pragmatics of Dialogue*.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*, Volume 626. Cambridge university press.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information.* Oxford University Press.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics (Syntax and Semantics 9).* Academic 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.

Venant, A. and N. Asher (2016). Ok or not ok? commitments in acknowledgments and corrections. In *Semantics and Linguistic Theory*, Volume 25, pp. 595–614.

Venant, A., N. Asher, and C. Degremont (2014). Credibility and its attacks. In *Proceedings of the 18th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, pp. p. 154–162.

Walker, M. A. (1996). Inferring acceptance and rejection in dialogue by default rules of inference. *Language and Speech 39*(2-3), 265–304.