Aligning Intentions: Acceptance and Rejection in Dialogue.

Julian J. Schlöder
ILLC / Amsterdam
j.j.schloder@uva.nl

Antoine Venant
IRIT / Toulouse
antoine.venant@irit.fr

Nicholas Asher
IRIT / Toulouse
asher@irit.fr

Abstract

We present a novel framework for a formal pragmatics and show applications with a special focus on the dynamics of agreement and disagreement. We are particularly interested in *intentions*. We circumvent the notoriously difficult task of axiomatizing agents' *internal* intentions by reducing *externalized* intentions to commitments to preferred futures. We give a formally precise account of both Stalknakerian rejections and more general rejections of arbitrary speech acts.

Many accounts of pragmatic reasoning and utterance interpretation refer to the speakers' intentions (Cohen et al. 1990 is a collection of related work). For example, the model of Asher and Lascarides 2013 verifies the following chain of inferences to govern agreement in a cooperative dialogue between A and B:[1]

$$C_A p \mathrel{|\!\sim} C_A I_A C_B p \mathrel{|\!\sim} I_B C_B p$$
$$\text{by (ISC)}\ C_A \varphi > C_A I_A C_B \varphi$$
$$\text{and (Co)}\ C_A I_A \varphi > I_B \varphi.$$

That is, by asserting that p, the speaker A establishes the commitment $C_A p$. By Intent to Share Commitment (ISC), A intends that B make the same commitment, $C_A I_A C_B p$. Then, Cooperativity (Co) infers that B adopts this intention, $I_B C_B p$. In her next move, B is expected to fulfill this intention by making an agreement move. The two axioms formalise two truisms about dialogues: that speakers want to be agreed with, and that, in cooperative settings, intentions are mutually adopted (Clark, 1996). A

problem with such accounts is that the notion *intention* is notoriously nebulous. Intentions are inherently private to the interlocutors. Hence it is hard to give a motivated semantics to operators like I_A above, *i.e.*, to say when $I_A \varphi$ is true and, if it is, what grounds its truth.

The first contribution of our work reduces propositions about dialogue intentions to propositions that have truth-conditions grounded in the external world. Informally, an intention for us restricts a space of possible futures. That is, we identify intentions with futures to which an agent commits. To this end, we integrate the language of *temporal modal logic* (where $\diamond$ means 'eventually') into an action model. We then can say that to-intend-that-α is the *committment* that α-will-happen, *i.e.*, in the language of Asher and Lascarides, $I_A \alpha$ is $C_A \diamond \alpha$. On first glance, this seems to overstate the matter in two ways: (i) one surely does not externalise (by the public commitment operator C_A) all internal intentions; and (ii) merely intending something is strictly weaker than claiming that it will happen. A brief summary of our counterarguments is as follows.

We concede (i), but argue that all intentions that are relevant to the dialogue are externalised: wherever required for utterance interpretation, they are presupposed or inferred and—as such—on the public record with the utterance. The overstatement (ii) is in fact not as severe as it seems, if such commitments to futures are only inferred by nonmonotonic inference. For individual actions α, it is clear that '*A intends α*' non-monotonically entails that '*A will do α*.'[2] If one voices an intention to a (linguistic) *joint* action, one can generally expect that cooperative interlocutors will partake.[3] Of course, these inferences are not mono-

---

[1] $\mathrel{|\!\sim}$ is defeasible inference, $>$ a default conditional, $C_A \varphi$ 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.

[2] Note that this separates intentions from mere desire. For example, I might desire to go to an expensive restaurant, but if I'm pressed for money, I will not intend it.

[3] For example, one speaks an utterance with the intent to be understood. Then, in speaking an utterance one also *com-*

tonic. Therefore we make use of default logic in our formalisation of intentions in dialogue.

We model temporal logic as KT4 modal logic and state the following basic coordination principles for commitments and temporal operators ($>$ denotes a default conditional).

(a) $C_A\varphi > C_A\neg\Diamond\neg C_A\varphi$. Commitments are not intended to be broken (*i.e.*, to be kept).

(b) $C_A\neg C_A\varphi > C_A\neg\Diamond C_A\varphi$. Withhold judgements are intended to be kept.

To do pragmatics in this language we formalise two truisms: interlocutors intend to reach eventual agreement, and if intentions align they are actualised.

(c) $C_A(\Diamond\Box(C_A\varphi \leftrightarrow C_B\varphi))$ and
$C_B(\Diamond\Box(C_A\varphi \leftrightarrow C_B\varphi))$
(Eventual agreement).

(d) $((C_A\Diamond\varphi) \wedge (C_B\Diamond\varphi)) > ((C_A\varphi) \wedge (C_B\varphi))$
(Alignment of intentions).

Underlying our analysis is a sophisticated analysis of speech acts that models them as actions in a dynamic logic: For A to make an assertion that p or to agree to p is to make an action that changes the commitment structure of A, *i.e.*, $C_A p$. We view speech acts as such actions and therefore can model the temporal operators as ranging over sequences of action-induced model transitions. In this model, we (i) generalise the Asher-Lascarides model above, (ii) give a formalisation of Stalnakerian rejection, and (iii) generalise on (i) and (ii) to formalise the acceptance and rejection of arbitrary speech acts.

For (i), we verify that $C_A p \mathrel{\v\!\sim} C_A\Diamond C_B p$, *i.e.*, that A intends that B agree. This corresponds to (ISC).[4] To *reject* an assertion, according to Stalnaker 1978, is to refuse to accept it; this distinguishes it from previous analyses of *correction* where a speaker asserts a revised version of an assertion. For (ii), we formalise the action *'B rejects p'* as effecting $C_B\neg C_B p$ (*i.e.*, to commit to not accepting) and derive $C_B\neg C_B p \mathrel{\v\!\sim} C_B\Diamond\neg C_A p$ (*i.e.*, B wants A to retract). However, a rejection also presupposes that the rejected proposition is *understood*; B understanding A is $C_B C_A p$ (Venant et al., 2014). We integrate this into our action structure and verify that it does not interfere with the above derivations.

In addition, *any* speech act can be rejected, *i.e.*, refused to be taken up (Austin, 1962). To our knowledge, there is no extant formal model for this. Under an Austinian conception of felicity, speech acts are made to *intentionally* bring forth a change; Clark 1996 characterises linguistic acts as *joint actions* in general. Thus, the rejection of a speech act is to refuse participating in its intended action. We account for this as follows: If A makes a speech act with intended effect φ, our model sees this as $C_A\Diamond\varphi$ (as shown above, our model derives that the effect of an assertion is to project acceptance, and the effect of rejecting-an-assertion is to prompt retraction). For B to take φ up is to adopt the intention, *i.e.*, $C_B\Diamond\varphi$. Upon such uptake, φ is realised by both speakers according to (d). Conversely, for B to reject a speech act with effect $\Diamond\varphi$ is to commit to negating the effect, *i.e.*, $C_B\neg\Diamond\varphi$. This generalises the narrower Stalnakerian conception of rejection described above. The effect of A asserting p is $\Diamond C_B p$, and B rejecting this is $C_B\neg\Diamond C_B p$. Our axioms verify that $C_B\neg\Diamond C_B p \mathrel{\v\!\sim} C_B\Diamond\neg C_A p$—the same as B rejecting by $C_B\neg C_B p$.

Applying this framework to further pragmatic phenomena, most notably to non-cooperative dialogue settings, we can show that different degrees of cooperativity correspond to structural properties of commitment structures. Also, we can equip our logic with a full model theory, giving truth conditions to statements about intentions. This is done using the *conversations as infinite games* framework of Asher and Paul 2013. In the model theory, a conversation is an unbounded sequence of speech acts; basic commitments (*i.e.*, commitments to propositions) are monotonic consequences of these speech acts, and temporal operators partition the space of sequence continuations.

References

Nicholas Asher and Alex Lascarides. 2013. Strategic conversation. *Semantics and Pragmatics*, 6:1–58.

Nicholas Asher and Soumya Paul. 2013. Infinite games with uncertain moves. In *1st SR Workshop*, p. 25–32.

John L. Austin. 1962. *How to do Things with Words*. Clarendon Press.

Herbert H. Clark. 1996. *Using language*. Cambridge University Press.

P. R. Cohen, J. Morgan, and M. Pollack, editors. 1990. *Intentions in Communication*. MIT Press.

Robert Stalnaker. 1978. Assertion. In P. Cole, editor, *Pragmatics (Syntax and Semantics 9)*. Academic Press.

Antoine Venant, Nicholas Asher, and Cedric Degremont. 2014. Credibility and its attacks. In *SemDial 18*, p. 154–162.

---

mits that it *can* be understood and thus that it, eventually, *will* be understood.

    [4]$C_B\Diamond C_B p$ (analogous Co), follows if commitments satisfy axiom 5: $\neg C_A\neg p \to C_A\neg C_A\neg p$.