1

Making the Right Commitments in Dialogue

Nicholas Asher
IRIT
Université Paul Sabatier, Toulouse

Alex Lascarides
School of Informatics,
University of Edinburgh

September 28, 2008

Abstract

In this paper we address a problem a problem for analyses of conversation that are based on a robust notion of cooperation. The problem is that such accounts fail to say anything about conversations where a robust notion of cooperation is absent. Moreover, such conversations are not hard to find. They are all around us. We build on previous work in Segmented Discourse Representation Theory (SDRT)Asher and Lascarides (2003), we providea dynamic logic for constructing a formal representation of both cooperative and non cooperative dialogues. Our representations depict the individual commitments of each conversational participant, and we rely on SDRT's glue logic for constructing such representations for sample dialogues. Here we extend the SDRT's glue logic with a version of the logic of public announcement logic (PAL) to render it dynamic. This extension also leads to a modification of standard of PAL to include defaults—-commonsense reasoning determines which speech acts an agent is committed to. We also link public commitments to private attitudes—such as belief, desire and intention—within a seperate but related logic of cognitive modelling. This is also an extension of a dynamic logic of public announcement, and is designed for computing calculable implicatures and an agent's next dialogue move in a decidable manner. The cognitive logic synthesises different approaches to modelling agents, particularly those from the BDI literature and game theory. We show how game-theoretic principles can be used to derive axioms of cooperativity that are typically treated as primitive in BDI logics.

1 Introduction

A desirable feature of any theory of dialogue is that it should link discourse interpretation to general principles of agent rationality and cooperativity (Grice, 1975). For Grice, Griceans and Neo-Griceans, conversation is fundamentally cooperative, and cooperative in at least two ways: a basic level concerning the conventions that govern linguistic meaning, and a level concerning shared attitudes towards what is said. The basic level of cooperation is rather self-evident: we must cooperate and use words in the same way with roughly the same meaning in order to be able to communicate. Grice encodes second level of cooperation in his conversational maxims, which are in turn specifications of his Cooperative Principle.

[*] The first author would like to thank David Beaver for helpful discussions on the subject.

These have been very influential not only in philosophy but also in linguistics and AI, leading to a so-called 'mentalist approach', which treats dialogue as a function of the agents' mental attitudes, usually formalised with BDI (belief, desire, intention) logics that incorporate axioms such as Sincerity—one normally believes what one says (Allen and Litman, 1987, Grosz and Sidner, 1990, Lochbaum, 1995), and one normally tries to help one's interlocutors achieve their conversational goals, which in turn requires speakers to adopt shared intentions. Finally, mutually agreeing that a proposition p is true—in other words, grounding at Clark's (1996) level 4—occurs when these axioms validate an inference that p is mutually believed.

While this approach has been very influential, even a cursory examination of different types of conversation reveals that the second level of Gricean style cooperation is a mark of only a few types of conversations. Consider, for instance, a conversation between a prosecutor and a defendant whom he is cross examining. The prosecutor wants to get the defendant to say things that will show that he is guilty of the crime charged; and the defendant, if he is rational, will clearly not want to help the prosecutor achieve his conversational goals. If he can get away with it, he will lie, claim not to remember details that he is asked about and so on. As a second example, consider a debate between two political contestants. Each has the intention to show himself the better candidate, and thus they have opposed conversational goals. This last sort of conversation led linguists like Ducrot (1973) and Merrin (1999) to think of conversations as arguments, where each participant has the conversational goal to convince the others of the rightness of his opinion. But we think to say that all conversations are arguments is equally myopic. A more general point of view is needed.

Not only are shared intentions not a feature of many conversations, the grounding established by mutual agreement often fails to translate into mutual belief. Gaudou et al. (2006) provide the following example to motivate a distinction between agreement and belief.

(1) a. A to B (C out of earshot): C is stupid.
b. B to A (C out of earshot): I agree.
c. A to B (C in earshort): C is smart.

Intuitively, A and B agree that C is stupid. If B now utters *That's right*, then they would also agree that C is smart. So if agreement amounts to mutual belief, then A and B would hold contradictory mutual beliefs, making them irrational (given that mutual belief entails private belief). And this would be a wrong prediction: A is not irrational; he is disingenuous. Gaudou et al. (2006) therefore conclude that agreement is a function of *shared public commitments* as opposed to mutual beliefs, and one must define the group to which public commitments are made. But although public commitment and belief should be distinct, they must be linked: B should conclude that A is lying—in other words, that A cannot believe all the things that he has publicly committed to.

Dialogue (1) contrasts with dialogue (2); here, A 'drops' a commitment to (2a) in favour of (2b):

(2) a. A: It's raining.
b. B: No it's not.
c. A: Oh, you're right.

Accordingly, a theory of dialogue should distinguish A's illocutionary act in (1c) vs. his act in (2c), even though in both cases A asserts the negation of his prior assertion.

Mentalists may well reply that since their principles of cooperativity are defeasible rules, the sorts of conversational situations we have just evoked are not counterexamples to their maxims or the Cooperative principle. True enough, but the Mentalist program has nothing to say about non cooperative conversations. Part of the reason is simply that the mentalist program hasn't paid much attention to non cooperative conversations, but part of the reason is of theoretical interest as well. Almost all mentalist approaches to dialogue of which we are aware give short shrift to the detection and modelling of preferences of the dialogue participants. But modelling preferences is crucial to determine whether a conversation is cooperative or not. Roughly speaking, when the preferences of the dialogue participants are completely aligned (they share the same preferences), then we have a cooperative conversation and something that corresponds to a game of coordination. In other conversations like political debates, the preferences of the participants are opposed and we have something like a 0 sum game in game theory and a game of pure conflict. Building on our previous work (Asher (1993), Asher and Lascarides (2003), Lascarides and Asher (2008), Asher and Bonzon (2008), we propose in this paper to provide a framework that models preferences as they are learned from conversational moves and affected by conversational moves. We want to model the dynamic transitions that we think are fundamental to dialoguge—the flow from preferences to conversational actions (or conversational moves), which in turn affect preferences that in turn will lead to a new decision problem involving information about the preferences of both dialogue participants about what to say next. In fact these decision problems are game problems, and our framework will enable us to think of conversations, both strategic and cooperative, as forms of dynamic games. We believe such a framework that is needed to model strategic or non-cooperative conversations as well as cooperative ones adequately.

To accomplish our goal, we will need to do several things. First, following our previous work Asher (1993), Asher and Lascarides (2003) we will sharply distinguish the content of dialogue from the contents of the participants' cognitive states and then link the two. We need a term for what a speaker engages in when he or she makes a conversational move. Following Hamblin (1987), we will say that the speaker makes a *public commitment* to a content when he or she makes a conversational move—exactly what kind of commitment will depend on the nature of the speech act or speech acts made by the conversational move. Following Lascarides and Asher (2008) we will sketch out how we currently think of the logical form for dialogue. We will work within Segmented Discourse Representation Theory (SDRT Asher and Lascarides (2003)), and use *rhetorical relations* to distinguish A's commitments in (1) vs. (2): A is committed to $\mathit{Correction}(2a, 2b)$ in dialogue (2) thereby entailing that he is committed to the negation of (2a), while in (1) the fact that A's utterances are addressed to a different group suggests they are not rhetorically connected at all and he remains committed to both of his utterances.

This paper extends Lascarides and Asher (2008) by extending the notion of public commitment to discourse moves to the notion of public commitment to certain attitudes that can be defeasibly inferred from these moves. We thus connect a commitment based approach with a BDI one, preserving the insights of both and avoiding, hopefully, many of their weaknesses. We achieve this cognitive extension by reconstructing SDRT's cognitive logic (CL, Asher and Lascarides (2003))—a separate but related logic to that of dialogue content—to include the attitude of public commitment, and axioms that relate it in default ways to other, private, attidues.

The next step is to make the private attitudes and the public commitments made in dialogue moves interact. We need to be able to reason about possible conversational moves as a means to realizing preferences. This requires us to reason about hypothetical updates to commitments and then to cognitive states. We will to this end base CL on a dynamic logic of public announcement (Baltag et al., 1999), extended with default axioms of rationality and cooperavitity. The result will capture dynamic flow from preferences to dialogue moves and back to preferences that we mentioned above, and more generally the practical reasoning that goes on in conversation when speakers adjust their preferences and intentions in light of what's been said.

The *dynamic* influence of dialogue on cognitive states and *vice versa* is ignored in our earlier work (e.g., the version of CL from Asher and Lascarides (2003) is static). It is also ignored in theories of dialogue content such as Ginzburg's (Ginzburg, 1995a,b) that do not focus on axiomatising the link between dialogue content and cognitive attitudes. By adopting a dynamic CL, we will also refine the approach to dialogue from "dialogue games" (e.g., Amgoud (2003), MacKenzie (1979), McBurney and Parsons (2002), Walton and Krabbe (1995), Wooldridge (2000) *inter alia*). because in contrast to these approaches, the utilities for each possible dialogue move in CL need not be 'pre-defined' or quantified. Indeed, even a complete list of possible dialogue moves is unnecessary. Rather, CL will exploit the dynamics in the logic to infer qualitative statements about the *relative* utility of various states, and these relative utilities can change as the dialogue progresses and agents learn more about facts and about each other. Our approach can also be viewed as extending the Grounding Acts Model of dialogue (Traum, 1994, Traum and Allen, 1994): following Poesio and Traum (1997), it provides its update rules with a logical rationale for constraining the update effects on content vs. cognitive states in the way they do, because the axioms in CL will provide a general model of information transfer between dialogue content and cognitive states.

While we think that this dynamic interaction between preferences and conversation is extremely important and while we think the basic approach, general methods and solution concepts of game theory are essential in analyzing this dynamic interaction, we are fully aware that supposing that conversational agents actually use game theoretic methods to calculate optimal moves is cognitively and empirically implausible. Classical game theory makes assumptions about the cognitive capacities and application of those capacities in ordinary conversation that are difficult to maintain. And while evolutionary game theory has shown us how to do away with those assumptions when speaking of the adoption of linguistic conventions (Skyrms 1994, 2003) or the adoption of various general principles like pragmatic defaults (Asher, Sher and Williams 2001, van Rooij 2003), it is implausible to suppose that evolutionary game theory can guide actual conversational decisions based on preferences that have been recently revised in the light of what has been said. That would require the evolutionary stability of a plan library the size of which even Roger Schank couldn't dream of. Finally, standard game theory demands that each agent precise and complete information about the strategies and preferences of the other dialogue agents, something which in practice agents almost never have. Fortunately, we have found, we think, a way out of this dilemma. Building on work of (Boutilier et al., 1999, 2004, Domshlak, 2002), we can have a "poor man's" representation of preferences that when combined with the notion of Boolean

Games (Bonzon 2007) provides us with a a limited game theoretic representation in which the solution concepts are for the most part extremely simple in complexity (Bonzon, 2007). There are certain limitations; we do not represent probabilities and expected utilities, and so we cannot represent mixed equilibria or even decisions under risk in our framework. But we can express many things, enough, we hope, to model strategic conversation. In particular, we'll be able to derive some axioms of cooperativity that are typically treated as primitive in BDI approaches and in our own Asher and Lascarides (2003). In addition, we will be able to model both cooperative and non cooperative conversations.

Below, section 2 explores informally the illocutionary contribution of various speech acts in terms of public commitment. Section 3 addresses the problem of constructing the logical form of dialogue, which includes identifying the speech acts performed. This is done in a description logic called the *glue logic*: its formulae are interpreted as (partial) descriptions of logical forms; and its (default) axioms are used to infer how underspecified compositional semantics is resolved to pragmatically preferred values. Until now, SDRT's glue logic has been static, making it impossible to show within the logic itself how a discourse logical form *evolves* with new verbal exchanges. We rectify this here by making the glue logic dynamic, extending a logic of public announcement to include ways of reasoning about the default consequences of the messages that agents exchange. The tractability of public announcement logics makes it possible to keep the glue logic's computational complexity as it was in the static case.

We will then focus in Section 4 on the *logic of cognitive modelling* (CL), modelling the links between what's said and the agents' mental states.[1] This is not a full or deep logic of mental attitudes; dialogue content provides only partial and shallow information about this. SDRT's static CL from Asher and Lascarides (2003) cannot model how agents learn about each other as the dialogue progresses; nor how they adapt their dialogue moves in light of what they learn. We will rectify this here by going dynamic, again by using public announcement logic. We will study the interactions between various speech acts and attitudes in this setting, in particular modelling how information from the dialogue's logical form transfers into CL. Once we have the means to study the effects of speech acts on various attitudes like beliefs and intentions and *vice versa*, we will investigate strategic reasoning by participants about which speech acts to make in response to other speech acts.

2 The General Approach

Lascarides and Asher (2008), argue that relational speech acts or *rhetorical relations* (e.g., *Narration*, *Explanation* and *Acknowledgement*) are a crucial ingredient for an adequate model of agreement (or, equivalently, *grounding* at level 4 (Clark, 1996)). This distinguishes (1) from (2) (see earlier discussion). It also affords a straightforward model of implicit agreement: representing the illocutionary contribution of an agent's utterance via rhetorical relations accurately reflects his commitments to another agent's commitments, even when this is linguistically implicit. For example, Karen's utterance in (3c) commits her to (3b) thanks to the semantic consequences of the relational speech act $Explanation(3b, 3c)$ that she has

[1] One major motivation for this separation between CL and the logic of dialogue content is that calculable implicatures are not available as antecedents to surface anaphora such as VP ellipsis and pronouns; see Asher and Lascarides (2003) and Section 4.2 for further details.

performed. And because Mark uses (3b) to explain why he and Sharon are having a fight (3a), a commitment the speech act in (3b) arguably entails that Sharon is also committed to (3a). This example is from Sacks et al (1974, p717)

(3) a. Mark (to Karen and Sharon): Karen 'n' I're having a fight,
b. Mark (to Karen and Sharon): after she went out with Keith and not me.
c. Karen (to Mark and Sharon): Wul Mark, you never asked me out.

Karen's commitment to (3b) is not (monotonically) entailed by (3c)'s compositional semantics, nor from Karen's asserting it. Rather, committing to $\mathit{Explanation}(3b, 3c)$ entails a commitment to (3b), (3c), and a causal connection between them. Accordingly, recognising an implicit acknowledgement (or acceptance act in the terms of the Grounding Acts Model (Traum, 1994)), is logically co-dependent on recognising the relational speech acts performed and their semantics.

More generally, we proposed that the commitments of each agent at a given turn in the dialogue is a Segmented Discourse Representation or SDRS (Asher and Lascarides, 2003): this is a hierarchically structured set of labelled rhetorical relations, as shown in each cell of Table 1—the proposed logical form for the dialogue (3):[2]

For simplicity, we have omitted the semantic representations of the clauses π_1 to π_3, and we have also adopted a convention that the root label of the agent i's SDRS for turn j is π_{ji}. In future, we may also refer to the content associated with a label π as K_π.

The logical form of dialogue is the logical form of each of its turns (where a turn boundary occurs whenever there is a change in speaker). The logical form of each turn is a set of

[2] We briefly recapitulate the definitions of SDRSs discussed Asher (1993) and Asher and Lascarides (2003), *inter alia*. We suppose that SDRS-formulae are constructed from a modal language which can express action statements via modal operators as well as questions using λ abstracts and a dedicated operator ?, a set of labels π, π_1, π_2, etc., and a set of relation symbols for discourse relations: R, R_1, R_2, etc. The set $\mathcal{L}$ of well-formed SDRS-formulae is defined as consisting of formulas of the first order language, together with:

1. If R is an n-ary discourse relation symbol and $\pi_1, \ldots, \pi_n$ are labels, then $R(\pi_1, \cdots, \pi_n) \in \mathcal{L}$.
2. For $\phi, \phi' \in \mathcal{L}, (\phi \wedge \phi'), \neg\phi \in \mathcal{L}$.

Definition 1 An SDRS

Let $\mathcal{L}$ be the set of SDRS-formulae. Then an SDRS is a triple $\langle \Pi, \mathcal{F}, \mathit{last} \rangle$, where:

- Π is a set of labels; i.e. $\Pi \subseteq$ vocab-2.
- *last* is a label in Π (intuitively, this is the label of the content of the last clause in the discourse); and
- $\mathcal{F}$ is a function which assigns each member of Π a member of $\mathcal{L}$.
- The relation $\succ$ that is the transitive closure of the *immediately outscopes* relation on labels Π as defined by $\mathcal{F}$ (i.e., π immediately outscopes π' iff $\mathcal{F}(\pi)$ contains as a literal either $R(\pi'', \pi')$ or $R(\pi', \pi'')$ for some relation R and label π'') satisfies the following two constraints. First, it forms a well-founded partial order, and secondly it has a unique root (that is, there is a unique label π_0 such that $\forall \pi \in \Pi$, $\pi_0 \succeq \pi$).

When there is no confusion, we may write $\langle \Pi, \mathcal{F} \rangle$ instead of $\langle \Pi, \mathcal{F}, \mathit{last} \rangle$. Note that labels don't need to form a tree under $\succ$. This reflects the fact that a single clause can make more than one illocutionary contribution to the context; see also Wolf and Gibson (2005).

Turn	Mark's SDRS	Karen's SDRS
1	π_1	$\emptyset$
2	$\pi_{2A} : \mathit{Explanation}(\pi_1, \pi_2)$	$\emptyset$
3	$\pi_{2A} : \mathit{Explanation}(\pi_1, \pi_2)$	$\pi_{3B} : \mathit{Explanation}(\pi_1, \pi_2) \wedge \mathit{Explanation}(\pi_2, \pi_3)$

Table 1: A representation of dialogue (3).

SDRSs—one for each dialogue agent, representing *all* his current public commitments, from the beginning of the dialogue up to the end of the turn in question.[3] And each agent constructs the SDRSs for all other agents as well as his own. For the sake of simplicity, we ignore the potential for misunderstandings in this paper. And so A and B will both build Table 1 (up to alphabetic variance on variables) as the logical form of (3).[4]

A dialogue participant may be committed in each turn not only to the (relational) speech acts that he performed in that turn, but also to the content expressed by prior speech acts, even if they were performed by another agent—for example, $\mathit{Explanation}(\pi_1, \pi_2)$ forms a part of Karen's commitments for the third turn, even though Mark uttered this speech act since it was Mark who uttered π_2. In this case, it is necessary to include $\mathit{Explanation}(\pi_1, \pi_2)$ in Karen's commitments for the third turn, because intuitively she is committed to their having a fight because she went out with Keith, and not just to the compositional semantics of π_2. Indeed, this contextually-specific content of π_2 is agreed upon at this point, and it follows from the semantic definition of $\mathit{Explanation}(\pi_1, \pi_2)$, but not from that of $\mathit{Explanation}(\pi_2, \pi_3)$, the speech act that Karen performed by uttering π_3. Lascarides and Asher (2008) proposed a number of general principles for computing these logical forms for dialogue: the principles predicted the semantic scope of implicit and explicit endorsements and challenges, and in particular provided the basis for adding $\mathit{Explanation}(\pi_1, \pi_2)$ to Karen's commitments in the third turn of (3).

Lascarides and Asher (2008) also provide a dynamic semantics for the so-called Dialogue SDRSs or DSDRSs, such as the one shown in Table 1. Roughly, where D is the set of dialogue participants, the context of evaluation C_d (d standing for dialogue) is a set of dynamic contexts

[3] Representing prior but ongoing commitments in the current turn is motivated by a desire to avoid revision in the model theory. See Lascarides and Asher (2008) for details.

[4] Here is the technical definition of a DSDRS that we gave in Lascarides and Asher (2008). Let D be a set of dialogue participants. Then a Dialogue SDRS (or DSDRS) is a tuple $\langle n, T, \Pi, \mathcal{F}, \mathit{last}\rangle$, where:

- $n \in \mathcal{N}$ is a natural number (intuitively, $j \leq n$ is the j^{th} turn in the dialogue);
- Π is a set of labels;
- $\mathcal{F}$ is a function from Π to the SDRS-formulae $\mathcal{L}$;
- T is a mapping from $[1, n]$ to a function from D to SDRSs, such that each SDRS is drawn from Π and $\mathcal{F}$. That is, if $T(j)(a) = \langle \Pi^a_j, \mathcal{F}^{d_i}_j, \mathit{last}^a_j \rangle$ where $j \in [1, n]$ and $a \in D$, then $\Pi^a_j \subseteq \Pi$ and $\mathcal{F}^a_j =_{def} \mathcal{F} \upharpoonright \Pi^a_j$ (that is, $\mathcal{F}^a_j$ is $\mathcal{F}$ restricted to Π^a_j), and $\mathit{last}^a_j \in \Pi^a_j$.
- $\mathit{last} =_{def} \mathit{last}^d_n$, where d is the (unique) speaker of the last turn n, and last^d_n is the 'last' label in $T(n)(d)$ (note that each turn has a unique speaker, because turn boundaries occur whenever the speaker changes, even if this is mid-clause).

For convenience, we may refer to the SDRS $T(j)(a)$ as $T^a(j)$.

In words, T maps each turn and dialogue agent to an SDRS. This definition reflects the logical forms proposed in Lascarides and Asher (2008) and illustrated in Tables 1 to 3.

Turn	A's SDRS	B's SDRS
1	$\pi_1 : K_{\pi_1}$	$\emptyset$
2	$\pi_1 : K_{\pi_1}$	$\pi_{2B} : \mathit{Correction}(\pi_1, \pi_2)$
3	$\pi_{3A} : \mathit{Correction}(\pi_1, \pi_3) \wedge \mathit{Acknowledgement}(\pi_2, \pi_3)$	$\pi_{2B} : \mathit{Correction}(\pi_1, \pi_2)$

Table 2: The logical form of dialogue (2).

for interpreting SDRSs—one for each agent $a \in D$. Thus, where C^i_a and C^o_a are respectively an input and output context for evaluating an SDRS:

$$C_d = \{\langle C^i_a, C^o_a \rangle : a \in D\}$$

The semantics of a dialogue turn $T = \{S_a : a \in D\}$ is the product of the CCPs its SDRSs S_a:

$$C_d [\![T]\!]_d C'_d \text{ iff } C'_d = \{\langle C^i_a, C^o_a \rangle \circ [\![S_a]\!]_m : \langle C^i_a, C^o_a \rangle \in C_d, a \in D\}$$

And given that a turn represents all an agent's commitments from the beginning of the dialogue, the CCP of a dialogue overall is that of its last turn. Dialogue entailment is then defined in terms of the entailment relation $\models_m$ afforded by the semantics $[\![.]\!]_m$ of SDRSs (m stands for monologue):

$$T \models_d \phi \text{ iff } \forall a \in D, S_a \models_m \phi$$

Thus $\models_d$ defines shared public commitments, and we assume that ϕ is agreed upon in turn T among D iff $T \models_d \phi$. Similar definitions hold for agreement among a subgroup $D' \subset D$: i.e., for all $a \in D'$, $S_a \models_m \phi$. This definition of agreement means that the DSDRS of (3) makes the following agreed upon or grounded at level 4 at the end of the conversation: Mark and Karen are having a fight because she went out with Keith and not Mark.

The logical form of dialogue (2) is Table 2. The inference in the glue logic that B is committed to $\mathit{Correction}(\pi_1, \pi_2)$ stems from the incompatibility between K_{π_1} and K_{π_2} (we assume this incompatibility is transferred into the glue language, as shown in Section 3) and the general glue-logic principle that the necessary semantic effects of a speech act are normally sufficient for inferring it has been performed. The content of (2c) (labelled π_3) supports a glue-logic inference that π_3 acknowledges π_2. This resolves its underspecified content to entail K_{π_2}, and so $\mathit{Correction}(\pi_1, \pi_3)$ is also inferred (as shown), via the same reasoning that led to the inference $\mathit{Correction}(\pi_1, \pi_2)$. The dynamic interpretation of this DSDRS is consistent, even though the SDRSs from turn 2 are not consistent with each other. Furthermore, it predicts that the proposition that it's not raining is agreed upon by the end of the conversation.

In contrast, the fact that (1c) is designed to be overheard by C while (1ab) is not forces a glue-logic inference that they are not rhetorically linked at all; see the logical form in Table 3. This DSDRS is inconsistent, because A's SDRS for turn 3 is inconsistent. It predicts that A and B agree on (1a). Should B endorse (1c), then the updated DSDRS would make them agree on all propositions (since the shared entailments would be inconsistent).

Since we focussed only on propositions in Lascarides and Asher (2008), the input and output contexts C^i_a and C^o_a for interpreting SDRSs were world-variable assignment pairs (i.e., $C^i_a, C^o_a \in W \times F$, where W is the set of possible worlds and F is the set of partial variable assignment functions), following the distributive dynamic semantics of SDRS-formulae from Asher and

Turn	A's SDRS	B's SDRS
1	$\pi_1 : K_{\pi_1}$	$\emptyset$
2	$\pi_1 : K_{\pi_1}$	$\pi_{2B} : \mathit{Acknowledgement}(\pi_1, \pi_2)$
3	$\pi_{3A} : K_{\pi_1} \wedge K_{\pi_3}$	$\pi_{2B} : \mathit{Acknowledgement}(\pi_1, \pi_2)$

Table 3: The logical form of (1).

Lascarides (2003). But what does it mean to be publicly committed to the content of a question, and what responses can be regarded as resolving that question to the satisfaction of the questioner? The dynamic semantics of questions from Groenendijk and Stokhof (1982) and Asher and Lascarides (2003) assumes that the output context of a question is a set of dynamic propositions (the true direct answers). This is of a different semantic type from the input and output contexts for interpreting propsitions, making the dynamic semantic interpretation of DSDRSs as defined above highly problematic. We therefore need to modify the semantics so that questions and propositions yield contexts of the same semantic type, as done in Asher (2007) who follows Groenendijk's (2003) semantics of questions. The modified semantics of DSDRSs yields a straightforawrd interpretation of an agent publicly committing to the content of a question and hence also the concept of a question being grounded and resolved. While the semantic type of the contexts C^i_a and C^o_a will change, the relationship between the interpretations $[\![.]\!]_d$ and $[\![.]\!]_m$ of dialogue and 'monologue' remain as in Lascarides and Asher (2008).

3 Computing the Right Commitments

In Asher and Lascarides (2003) we argue that the logic for constructing the logical form of dialogue should be decidable, so as to provide a competence model of language users who by and large agree on what was said (if not its cognitive effects, Lewis (1969)). This glue logic must involve nonmonotonic reasoning and hence consistency tests, because one never has complete information about the dialogue context, including the speaker's intentions. So to remain decidable, the glue logic must be separate from, but related to, the logic in which one *interprets* that logical form. SDRT achieves this by exploiting developments in underspecified semantics (e.g., Copestake et al. (2005), Egg et al. (2001)).

An underspecified semantics is a *partial description* of the *form* of the intended logical form. Many grammars derive an underspecified logical form (ULF) of sentences because syntax underdetermines semantic scope, lexical senses and antecedents to anaphoric expressions. The glue logic in SDRT builds ULFs, and accordingly it has only restricted access to what the candidate logical forms *mean* (for formal details, see Asher and Lascarides (2003)).

The glue logic derives a logical form (or, more accurately, a *partial description* of a logical form) through commonsense reasoning; this enables it to make predictions about which interpretation is pragmatically preferred. This is formalised using default axioms that predict rhetorical connections. Its rules have the following form, where the symbols α and β are metavariables ranging over the labels of discourse segments in the DSDRS representation, $>$ is a weak conditional used to formalize defaults, and ? is a variable in the glue-logic language, that indicates that the value of some constructor in the fully-specific logical form is currently unknown:

- `Glue Logic Schema:` $\lambda :?(\alpha,\beta) \wedge \mathit{Info}(\alpha,\beta,\lambda)) > \lambda : R(\alpha,\beta,\lambda)$

In words: if the segment labelled β is to be connected to the segment labelled α with a rhetorical relation whose value we don't know yet, and the result is to to be a part of the dialogue segment λ and moreover $\mathit{Info}(\alpha,\beta,\lambda)$ holds of the content labelled by λ, α and β, then normally the rhetorical relation is R. The conjunct $\mathit{Info}(\alpha,\beta,\lambda)$ is cashed out in terms of the (underspecified) logical forms that α, β and λ label, and the rules are justified either on the basis of underlying linguistic knowledge, world knowledge, or knowledge of the cognitive states of the conversational participants. Thus glue logic axioms encapsulate *prima facie* default inferences about which types of speech act were performed, on the basis of a shallow representation of the content and context of the utterances. These default axioms give rise to a nonmonotonic proof theory $\mid\!\sim_g$. Intuitively, this gives us the default preferences for resolving underspecified aspects of logical form (including identifying antecedents to anaphora) that are generated by compositional semantics and other contextual information.

In the glue-logic axioms below, specific renditions of the conjunct $\mathit{Info}(\alpha,\beta,\lambda)$ will *look like* SDRS-formulae, but as we stressed above they have a different *interpretation* because only some, but not all, of the dynamic semantic content is transferred into the glue logic. Dynamic semantic consequences arising from the subsitution of equalities, $\wedge$-elimination, $\vee$-elimination and $\exists$-introduction are transferred over. But $\exists$-elimination is not transferred over, and so the glue logic loses the logical equivalence between, say, the SDRS-formulae $\neg\exists x\neg\phi$ and $\forall x\phi$.

Asher and Lascarides (2003) provide several glue logic axioms. Some of these follow from their formalization of axioms of rationality and cooperativity. `Q-Elab` and `IQAP`, where $\mathit{int}(\beta)$ means that β has interrogative mood, are examples:

- `Q-Elab:` $(\lambda :?(\alpha,\beta) \wedge \mathit{int}(\beta)) > \lambda : \mathit{Q\text{-}Elab}(\alpha,\beta)$
- `IQAP:` $(\lambda :?(\alpha,\beta) \wedge \mathit{int}(\alpha)) > \lambda : \mathit{IQAP}(\alpha,\beta)$

In words, `Q-Elab` states that if a question is attached to α with a discourse relation, then normally that relation is *Q-Elab*: in other words, all its possible answers will help to elaborate an executable plan for achieving the intention or SARG that prompted the speaker to say α. If, on the other hand, an utterance is attached to a question, then normally the relation is *IQAP*. These axioms *short-circuit* conversational implicatures (Morgan, 1975): while validating these axioms is achieved through reasoning about cognitive states (see Asher and Lascarides (2003)), the premises of these axioms don't talk about beliefs or intentions at all, and instead rest entirely on sentence mood.

Discourse coherence is not a yes/no matter; coherent discourses can vary in quality. SDRT assumes that the degree of coherence influences the SDRS that's built. SDRT represents degree of coherence as a partial order on all fully-specific interpretations. This partial order adheres to some very conservative assumptions about what factors contribute to coherence. Definition 2, taken from Asher and Lascarides (2003), specifies these factors, and this is now used to rank DSDRSs, as well as SDRSs, into a partial order.[5]

[5] A more formal definition of `MDC` is given in Asher and Lascarides (2003), and the partial orders over SDRSs

Definition 2 Maximising Discourse Coherence

Discourse is interpreted so as to maximise discourse coherence, where the ranking among interpretations are encapsulated in the following principles:

1. All else being equal, the more rhetorical connections there are between two items in a discourse, the more coherent the interpretation.
2. All else being equal, the more anaphoric expressions whose antecedents are resolved, the higher the quality of coherence of the interpretation. Moreover, anaphoric expressions that are resolved to values that lead to $\mid\!\sim_g$-consequences for a particular rhetorical relation are preferred over values that are logically unrelated to any rhetorical connection.
3. Some rhetorical relations are inherently scalar. For example, the quality of a *Narration* is dependent on the specificity of the common topic that summarises what went on in the story. All else being equal, an interpretation which maximises the quality of its rhetorical relations is more coherent than one that doesn't.
4. All else being equal, the number of labels in the semantic representation is minimal, so long as minimising the number of labels does not create semantic anomalies among the rhetorical relations in the representation (for example, $\pi_0 : \mathit{Contrast}(\pi_1, \pi_2) \wedge \mathit{Condition}(\pi_2, \pi_3)$ is anomalous because the first speech act 'asserts' K_{π_2} while the second does not, and this anomaly is removed by changing the relative scope of the acts: $\pi_0 : \mathit{Contrast}(\pi_1, \pi)$, $\pi : \mathit{Consequence}(\pi_2, \pi_3)$).

Definition 3 defines discourse update for DSDRSs. It is adapted from the definition of discourse update from Asher and Lascarides (2003) for SDRSs. As in original SDRT, the representation of the discourse context updated with new information includes all the $\mid\!\sim_g$ consequences of the old and the new information. So update always *adds* constraints to what the discourse means. If there is underspecified information about which of the available labels the new content attaches to, then update is conservative, and generalises over all the possibilities (see the second part of Definition 3).

Definition 3 Discourse Update for DSDRSs

Simple Update. We first define how one updates a context with new information β, given a particular available attachment site α.

Let $T(d, m, \lambda)$ be a formula in the ULF-language $\mathcal{L}_{\mathit{ulf}}$ which means that the label λ is a part of the SDRS $T^d(m)$ in the DSDRS being described. So the ULF-formula

it defines is easily extended into a formal definition of MDC that ranks DSDRSs into a partial order: roughly put, we take a conservative view that one DSDRS D_1 is more coherent than another D_2 if (a) they are comparable (i.e., they consist of the same number of turns and the same dialogue participants); and (b) each SDRS for each dialogue participant and turn in D_1 is equal or more coherent than the SDRS for that dialogue participant and that turn in D_2.

$\lambda :?(\alpha,\beta) \wedge T(d,m,\lambda)$ specifies that the new information β is to be attached to the DSDRS as a part of the SDRS $T^d(m)$.

Furthermore, let σ be a set of (fully-specified) DSDRSs, and let $Th(\sigma)$ be the set of all ULFs that partially describe the fully-specific DSDRSs in σ. And let ψ be either (a) a ULF $\mathcal{K}_\beta$, or (b) a formula $\lambda :?(\alpha,\beta) \wedge T(d,m,\lambda)$ about attachment, where $Th(\sigma) \models_{ulf} \mathcal{K}_\beta$ (in other words, $\mathcal{K}_\beta$ partially describes the set of DSDRSs in σ. Then $\sigma + \psi$ is a set of DSDRSs defined as follows:

1. $\sigma + \psi = \{\tau :\ \text{if } Th(\sigma), \psi {\mid\!\sim_g} \phi \text{ then } \tau \vdash_{ulf} \phi\}$, provided the result is not $\emptyset$;
2. $\sigma + \psi = \sigma$ otherwise.

Discourse Update. Suppose that A is the set of available attachment points in the old information σ for the new information β. Then the power set $\mathcal{P}(A)$ represents all possible choices for what labels α_i in σ the new label β is actually attached to. $update_{\text{SDRT}}$ is neutral about which member of $\mathcal{P}(A)$ is the 'right' choice, because $update_{\text{SDRT}}(\sigma, \mathcal{K}_\beta)$ is the *union* of DSDRSs that result from a sequence of +-operations for each member of $\mathcal{P}(A)$ together with a stipulation that the last element of the updated DSDRS is β.

Definition 3 makes $Th(\sigma) \subseteq Th(update_{\text{SDRT}}(\sigma, \mathcal{K}_\beta))$. In other words, constructing logical form is monotonic (Shieber, 1986), in that the representation of the discourse context is always an elementary substructure of the representation of the dialogue updated with the current utterance, even if the current utterance denies earlier content. However, the logical form remains a product of complex default reasoning, since identifying the speech acts that were performed involves commonsense reasoning with the linguistic and non-linguistic context.

While the old and new information are combined by discourse update so as to increase the constraints imposed on the form of the logical form, it typically doesn't yield a specific enough description to identify a *unique*, fully specific logical form. The Principle MDC then ranks those alternative, fully specific logical forms.

Definition 3 *uses* the entailment relation ${\mid\!\sim_g}$ but it is *external* to it. It is impossible to reason about updates *within* a static glue logic. As we argued in Section 1, we need to make this modal description logic dynamic. We do this via a simple application of public announcement logic (PAL) (Baltag et al., 1999). In simple PAL one can perform an action which is to announce a particular formula. PAL has a possible worlds semantics, and the effect of such an announcement is to change the model, restricting the states in the output model to those in which the announced formula is true. SDRT's glue logic can thus be reconceived in terms of the effects of announcing a formula: the states of the model are, as they have always been, fully specified DSDRSs. And so announcements will eliminate DSDRSs from the input model, leaving only those DSDRSs in the output that satisfy the announcement. As Definition 3 suggests, we need to specify the effects of three sorts of announcements:

1. $\mathcal{K}_\beta$—such an announcement is essentially the ULF of an utterance, or utterance segment.

2. $\lambda :?(\alpha,\beta) \wedge T(d,j,\lambda)$—such an announcement provides information about the choice of an attachment in the SDRS $T^d(j)$ (i.e., the semantic representation of d's commitments for dialogue turn j).

3. *last* $= \beta$—such an announcement requires that all 'states' in the model have β as the last entered element in the DSDRS.

If all consequences of one's announcements were *monotonic*, then we could stay within simple PAL. However, as Definition 3 makes plain, *nonmonotonic* consequences of announcements determine the DSDRSs, because identifying the speech act performed by an announcement and hence its illocutionary effects (including specific values for underspecified semantic conditions in the compositional semantics of the utterance) are generally a product of commonsense reasoning. These aspects of interpretation form a crucial part of the logical form of dialogue, enabling accurate predictions about implicit and explicit agreement. To support these nonmonotonic consequences we need another operation—not simple announcement but *announcement ceteris paribus.*

We will transform the glue logic's original model theory into a model theory for interpreting anouncement *ceteris paribus.* As we said earlier, the original models consist of a set of states or DSDRSs; in addition, there is a function $*$ from a state and a set of states to a set of states (this encapsulates normality and is used to interpret the modal conditional $>$), and a valuation function V for assigning values to the non logical constants of the ULF language. Its semantics, barring the details of the valuation function, is as follows:

- $\mathcal{M}, s \models \phi$ iff $s \in V(\phi)$ for atomic ϕ
- $\mathcal{M}, s \models \phi \wedge \psi$ iff $\mathcal{M}, s \models \phi$ and $\mathcal{M}, s \models \psi$
- $\mathcal{M}, s \models \neg\phi$ iff $\mathcal{M}, s \not\models \phi$
- $\mathcal{M}, s \models \phi > \psi$ iff $*^{\mathcal{M}}(s, [\![\phi]\!]^{\mathcal{M}}) \subseteq [\![\psi]\!]^{\mathcal{M}}$, where $[\![\phi]\!] = \{s' \ : \ M, s' \models \phi\}$

Converting this static model theory into a dynamic one involves (a) extending the object language to express public announcements; and (b) defining how models are transformed by such announcements in interpretation.

As is standard in PAL, we add a modality $[!\phi]$ to the object language, to express the announcement that ϕ. The formula $[!\phi]\psi$ means that ψ follows from announcing ϕ. The values of ϕ that are of interest to us are the ULFs $\mathcal{K}_\beta$, an assumption about attachment as expressed in the formula $\lambda :?(\alpha, \beta) \wedge T(d, j, \lambda)$, and *last* $= \beta$. All of these announcements are $>$-free (although $\mathcal{K}_\beta$ may contain a predicate symbol that stipulates that the SDRS being described contains the distinct modal connective $>$ from the vocabulary of DSDRSs, given in footnote 2. So for the rest of this section, we will assume that announcements are $>$-free. We also need to extend the standard PAL language to express the *ceteris paribus* consequences of announcements; we therefore introduce a new modality $[!\phi]^{cp}$, and $[!\phi]^{cp}\psi$ means that ψ normally follows from announcing ϕ.

Having extended the language's syntax, let's now refine the glue logic's model theory into a dynamic one where announcements transform the models.

Definition 4 A PAL model theory for the glue logic

Let $\mathcal{M} = \langle S, *, V\rangle$ be a model. So S is a set of DSDRSs; $*$ is a function from $S \times (S \times S)$ to $S \times S$, and V is a valuation function. We define $\mathcal{M}^\phi$ and $\mathcal{M}^{cp(\phi)}$ as follows:

- $\mathcal{M}^{\phi} = \langle S^{\phi},\ *^{\mathcal{M}}|S^{\phi}, V\rangle$, where
  - $S^{\phi} = S^{\mathcal{M}} \cap [\![\phi]\!]$
- $\mathcal{M}^{cp(\phi)} = \langle S^{cp(\phi)}, *^{\mathcal{M}}|S^{cp(\phi)}, V\rangle$, where
  - $S^{cp(\phi)} = \{s' \in S\ :\ Th(\mathcal{M}), \phi \mid\!\sim \psi \rightarrow \mathcal{M}, s' \models \psi\}$

The (dynamic) interpretation of $[!\phi]\psi$ and $[!\phi]^{cp}\psi$ are as follows:

- $\mathcal{M}, s \models [!\phi]\psi$ iff $\mathcal{M}^{\phi}, s \models \psi$
- $\mathcal{M}, s \models [!\phi]^{cp}\psi$ iff $\mathcal{M}^{cp(\phi)}, s \models \psi$

In words, $\mathcal{M}^{\phi}$ is a model that's formed by eliminating all states from $\mathcal{M}$ that fail to satisfy the monotonic consequences of announcing ϕ; and $\mathcal{M}^{cp(\phi)}$ is formed by eliminating from $\mathcal{M}$ all states that fail to satisfy the nonmonotonic consequences of announcing ϕ. Note that *ceteris paribus* announcements, like 'simple' announcements, presuppose that the announcement is true; i.e. $S^{cp(\phi)} \subseteq [\![\phi]\!]$. Moreover, since $>$ is supra-classical, $S^{cp(\phi)} \subseteq S^{\phi}$. In words, this means that the logical form of an utterance is always a partial description of the logical form for the entire dialogue; it does not mean that the dynamic interpretation of the utterance is true (or even that the speaker is committed to it). Given the glue logic axioms, the nonmonotonic consequences of an announcement will typically express information about rhetorical connections (e.g., $\lambda : R(\alpha, \beta)$ for some particular relation R), or specify values of other underspecified elements introduced by the grammar, such as antecedents to anaphoric expressions. The glue logic axioms also ensure that consequences of a ceteris paribus announcement also express which commitments from prior turns are ongoing commitments. For instance, for dialogue (3), the glue-logic axioms will ensure that $\mathcal{M}, s \models [!\mathcal{K}_{\pi_3}]^{cp}\pi_{3A} : \mathit{Explanation}(\pi_1, \pi_2)$, where $\mathcal{M}$ is the model constructed by updating with utterances π_1 to π_3 (in that order), and $s \in \mathcal{M}$; see Lascarides and Asher (2008) for details of the relevant axiom that guarantees this.

We can now define discourse update within the dynamic logic in a very simple way. We imagine that the set of DSDRSs σ is simply the set of states of a model $\mathcal{M}_{\sigma}$:

- `Dynamic Simple Update:`
  $\sigma + \phi \vdash \psi$ iff $Th(\sigma) \models [!\phi]^{cp}\psi$

To define full DSDRS update, we simply take Boolean combinations of *ceteris paribus* updates, thereby matching the update process defined in Definition 3 (the second paragraph).

- `Dynamic Discourse Update:`
  Let $\mathcal{M}$ be a model that satisfies a set σ of DSDRSs, and let the ULF $\mathcal{K}_{\beta}$ be new information. Let $\Sigma_1, \ldots \Sigma_n$ be all the jointly compossible attachment sites of β, chosen from the set of all possible attachment sites for each DSDRS in in σ. Let k_i be an enumeration of the compossible attachment sites in Σ_i, $1 \leq i \leq n$. Then

$$\begin{aligned}
&\mathit{Update}(\mathcal{M}, \phi_{\beta}) \vdash \psi \text{ iff } \forall s \in S^{\mathcal{M}},\\
&\mathcal{M}, s \models [!(\mathcal{K}_{\beta} \wedge \mathit{last} = \beta)](\\
&\qquad \big[!(\lambda^1_1 :?(\alpha^1_1, \beta) \wedge T(d, j, \lambda^1_1) \wedge \ldots \wedge \lambda^1_{k_n} :?(\alpha^1_{k_1}, \beta) \wedge T(d, j, \lambda^1_{k_1}))\big]^{cp} \psi \wedge\\
&\qquad \big[!(\lambda^2_1 :?(\alpha^2_1, \beta) \wedge T(d, j, \lambda^2_1) \wedge \ldots \wedge \lambda^2_{k_2} :?(\alpha^2_{k_2}, \beta) \wedge T(d, j, \lambda^2_{k_2}))\big]^{cp} \psi \wedge\\
&\qquad \ldots \wedge\\
&\qquad \big[!(\lambda^n_1 :?(\alpha^n_1, \beta) \wedge T(d, j, \lambda^n_1) \wedge \ldots \wedge \lambda^n_{k_n} :?(\alpha^n_{k_n}, \beta) \wedge T(d, j, \lambda^n_{k_n}))\big]^{cp} \psi)
\end{aligned}$$

The nonmonotonic inference relation $\mid\!\sim$ afforded by the modal connective $>$ is that of Commonsense Entailment (Asher and Morreau, 1991). It is decidable, and validates the following intuitively compelling patterns of inference:

Defeasible Modus Ponens: $\phi, \phi > \psi \mid\!\sim \psi$

Penguin Principle: If $\phi \vdash \psi$ then $\phi, \phi > \chi, \psi > \neg\chi \mid\!\sim \chi$

Nixon Diamond: If $\phi \not\vdash \psi$ and $\psi \not\vdash \phi$ then
$\phi, \psi, \phi > \chi, \psi > \neg\chi \not\mid\!\sim \chi$
$\phi, \psi, \phi > \chi, \psi > \neg\chi \not\mid\!\sim \neg\chi$

The computational complexity of PAL is typically demonstrated by proving reduction axioms (e.g., Balbiani et al. (2007)). We provide the standard reduction axioms for PAL, plus an additional one for *ceteris paribus* announcements:

- **Reduction Axioms and Rules**
  1. $[\phi]p \leftrightarrow (\phi \rightarrow p)$
  2. $[\phi](\psi \wedge \chi) \leftrightarrow ([\phi]\psi \wedge [\phi]\chi)$
  3. $[\phi]\neg\psi \leftrightarrow \neg[\phi]\psi$
  4. $[\phi](\psi > \chi) \leftrightarrow ([\phi]\psi > [\phi]\chi)$
  5. $\frac{\Gamma \vdash [\phi]^{cp}\psi}{\Gamma, \phi \mid\!\sim \psi}$

The last rule is easily seen to be valid, once we note that $\Gamma \vdash [\phi]^{cp}\psi$ means that ψ is a nonmonotonic consequence of $\Gamma \cup \{\phi\}$.

A corresponding reduction rule for *defeasible* inferences to public announcements like (4) is not valid, however, because of the peculiarities of nonmonotonic consequence relations.

(4)
$$\frac{\Gamma \mid\!\sim [\phi]^{cp}\psi}{\Gamma, \phi \mid\!\sim \psi}$$

Many nonmonotonic reasoning systems or logics have problems with the so called "drowning problem" noted by Benferhat et al. (1993). Here, the problem has to do with examples involving nested conditionals of the sort proposed in Asher (2004). For instance in (5) and (6), $\Gamma \mid\!\sim [\phi]^{cp}\psi$, but $\Gamma, \phi \not\mid\!\sim \psi$; the default given by ϕ 'drowns out' the needed conclusions of Γ when they are mixed together.[6]

(5) a. Γ: $\{A, A > D, A > ((B \wedge E) > C)\}$
b. ϕ: $B \wedge E \wedge ((A \wedge E) > \neg D)$
c. ψ: C

[6] It is easily shown that $\Gamma \mid\!\sim (B \wedge E) > C$ for both (5) and (6), and using this we can show that $\Gamma \mid\!\sim [\phi]^{cp}\psi$, since this is equivalent to taking all the nonmonotonic consequences of Γ and then taking the nonmonotonic consequences of those together with ϕ to yield ψ. However, we cannot show that $\Gamma, \phi \mid\!\sim \psi$ since $\Gamma, \phi \not\mid\!\sim (B \wedge E) > C$, which is essential to deriving C.

(6) a. Γ: $\{A, A > D, A > (D > ((B \wedge E) > C))\}$

b. ϕ: $B \wedge E \wedge \neg D$

c. ψ: C

Nevertheless since $\mid\!\sim$ is decidable, the reduction axioms that are valid ensure that our extension of PAL is decidable as well.

We have now made SDRT's glue logic dynamic. This allows a dialogue agent to *reason* about what the update of the DSDRS will be after his contribution, and this in turn allows us to inject an element of *planning* into the analysis of dialogue moves. But to make this possibility a reality requires reasoning about how public announcements affect, and are affected by, attitudes such as belief and intention. After all, planning dialogue moves involves identifying actions—in this case, public announcements—that will in all likelihood fulfill ones intentions, which in turn are chosen on the basis of one's preferences, and the likelihood of satisfying those preferences. Accordingly, we now examine SDRT's other shallow logic, the logic of cognitive modeling, to model these links among public announcements, commitments and the other attitudes.

4 Cognitive Modelling

SDRT's general architecture is highly modular, making its cognitive logic CL separate from, but related to, the way a dialogue is interpreted (see Chapter 9 of Asher and Lascarides (2003)), much as the glue logic is so related. This has the advantage of allowing us to exploit agent rationality and cooperativity to validate inferences about conversational implicature where this is appropriate, while avoiding reasoning with cognitive states entirely when computing 'conventional' implicatures (Grice, 1975), arising from the presence of discourse connectives like *but* or intonation. The cognitive logic in SDRT also has the advantage of a decidable consequence relation, so that constructing the semantic representation of dialogue remains decidable.

The model theory of CL is quite different from that of GL. Whereas GL is a description logic for reasoning about the form of DSDRSs, in CL one reasons about how a given agent models his own cognitive state and those of others, as is standard from the BDI literature. The states in a CL model therefore represent the epistemic possibilities about the agents' commitments, beliefs, intentions and preferences, given what has been said in the dialogue so far and given occasional inputs from the deep cognitive model, which Asher and Lascarides (2003) modeled as an oracle. So formulas like $\lambda : R(\alpha, \beta)$ are interpreted differently in CL from GL: whereas in GL this formula (partially) describes a bit of discourse structure, in CL it describes the content of the speech act β in a particular discourse context. So CL must capture some of the entailments of $\lambda : R(\alpha, \beta)$ from the dynamic logic of DSDRSs; if it did not, then CL could not concern itself with the cognitive consequences of public commitments that are made in discourse. However, following Asher and Lascarides (2003), we do not assume that *all* DSDRS-entailments are transferred into CL so as to ensure that CL remains decidable.

Accordingly, the transfer of formulae ϕ from the language of DSDRSs into a shallower form in

CL preserves some of its consequences but not all of them:[7]

Definition 5 Transfer from DSDRSs to CL

A transfer function τ from SDRS-formulae ϕ to CL-formulae ϕ^τ is defined as follows:

1. $\phi^\tau = \phi$, for atomic SDRS-formulae ϕ
2. $(\phi \wedge \psi)^\tau = (\phi^\tau \wedge \psi^\tau)$
3. $(\neg\phi)^\tau = \neg\phi^\tau$
4. $(\phi > \psi)^\tau = (\phi^\tau > \psi^\tau)$
5. $(\exists x\phi)^\tau = p_{\exists x\phi}$, where p is an atomic variable.
6. $(\pi : \phi)^\tau = \pi : \phi^\tau$
7. $(?\lambda x_1 \ldots \lambda x_n \phi)^\tau = int(x_1, \ldots, x_n, \phi^\tau)$, where *int* is (poly-morphic) predicate symbol (standing for *interrogative*).
8. $(!\delta\phi)^\tau = imp(\phi^\tau)$, where *imp* is a predicate symbol (standing for *imperative*).

Definition 5 ensures that reasoning about the cognitive effects of dialogue content ϕ enjoys access to some of ϕ's entailments from the dynamic logic for interpreting dialogue, such as what's entailed by conjunction, negation, and who said what. But existentially quantified SDRS-formulae lose their structure in the transfer to CL, and so CL does not recognise the logical relationship between, say, the SDRS-formulae $\neg\exists x\neg\phi$ and $\forall x\phi$. CL also eschews the action operator δ and the λ-abstracts that form part of the logical form of imperatives and questions respectively, so as to maintain CL's decidability. However, not all information is lost in CL from imperatives and interrogatives: the transfer function τ preserves the information that a given utterance was an imperative, or that it was an interrogative, via the predicate symbols *imp* and *int*. It simply lacks information about all their entailments, although depending on how the interpretation of the predicate symbols *imp* and *int* are constrained by meaning postulates, one could preserve some of the SDRS-entailments.

This transfer of information from DSDRSs to CL is very similar to the transfer of information from DSDRSs to GL; while GL needed to partial access to the logic of dialogue content so as to reason effectively about the pragmatically preferred interpretation of a dialogue, CL needs it to ensure that agents engage in effective practical reasoning about each other. In addition, so that one can reason in CL about the cognitive effects of different dialogue updates, we assume that all $>$-free information from GL is transferred into CL. Technically, there is a translation function τ' from the language of GL to that of CL such that if ϕ is a formula of GL that is $>$-free, then $(\phi)^{\tau'} = \phi$. This doesn't harm the complexity of CL, because GL is quantifier free.

Since CL models reasoning about cognitive states, we add modalities to its object language for expressing various mental attitudes. One important attitude, given our definition of agreement, is that of public commitment: $\mathcal{P}_{a,D}\phi$ means that a publicly commits to ϕ to the group of agents D. Semantically, we make this modality K45 (one commits to all the consequences of one commitments, and one has total introspection on commitments, or lack

[7] In Asher and Lascarides (2003), we dropped all structure; the information that ϕ was simply transferred into CL as the propositional variable p_ϕ. Here, we preserve some logical properties of DSDRSs (according to their dynamic sematics), but not all of them.

of them), and following Gaudou et al. (2006) we also add axioms Ax1 (a commitment to D is a commitment to all its subgroups) and Ax2 (there is a group commitment by x and y to D iff x and y both make that commitment to D):

K: $\mathcal{P}_{a,D}(\phi \rightarrow \psi) \rightarrow (\mathcal{P}_{a,D}\phi \rightarrow \mathcal{P}_{a,D}\psi)$

4: $\mathcal{P}_{a,D}\phi \rightarrow \mathcal{P}_{a,D}\mathcal{P}_{a,D}\phi$

5: $\neg\mathcal{P}_{a,D}\phi \rightarrow \mathcal{P}_{a,D}\neg\mathcal{P}_{a,D}\phi$

Ax1: For any $D' \subseteq D$, $\mathcal{P}_{a,D}\phi \rightarrow \mathcal{P}_{a,D'}\phi$

Ax2: $\mathcal{P}_{\{x,y\},D}\phi \leftrightarrow (\mathcal{P}_{x,D}\phi \wedge \mathcal{P}_{y,D}\phi)$

So the models $\mathfrak{M}$ of CL have suitably constrained accessibility relations $R^{\mathcal{P}_{a,D}} \subseteq W \times W$ for all a and D. Motivation for axiom K on commitments is that, following Hamblin (1987), commitments should not lack logic entirely. Axiom K ensures that what a commitment *means* affects exactly what the commitment is to, and it thus allows an agent to persuade another to drop a commitment on the grounds of its consequences. The introspection axioms 4 and 5 are intuitively valid on the grounds that without them, agents can be ignorant about their commitments. But the anomalous response (2c′) to B's dispute (2b) suggests that such ignorance is counterintuitive:[8]

(2) a. A: It's raining.
b. B: No it's not.
c′. A: ??I didn't know I conveyed to you that it was raining.

One could explain the anomaly of this dialogue by replacing axiom 4 with $\mathcal{P}_{a,D}\phi \rightarrow \mathcal{B}_a\mathcal{P}_{a,D}\phi$ (where $\mathcal{B}_a$ is belief). But this make sthe model theory unnecessarily complex: instead of simply incorporating a transitive accessibility relation for commitments, this triggers complex interactions among the accessibility relations for belief and commitments, making dynamic updates harder to compute. Finally, without axiom Ax1, CL would undergenerate agreement compared with what's agreed upon among dialogue agents at the level of logical form. For instance, in a dialogue setting where agents enter and leave the room while a dialogue progresses, and where that dialogue features no disputes, two agents x and y may share a public commitment to ϕ, but to different groups D_1 and D_2 respectively, where $y \in D_1$ and $x \in D_2$. Without Ax1, CL would not predict that x and y agree that ϕ. But the logic of DSDRSs does predict such an agreement.

Commitments lack axiom D, because one can make anything a matter of public record. As a result, $\mathcal{P}_{a,D}(p \wedge \neg p)$ is satisfiable, reflecting A's public commitments in (1). This contrasts with the belief modality $\mathcal{B}_a$, which is KD45, making belief is rational and the accessibility relation $R^{\mathcal{B}_a} \subseteq W \times W$ in $\mathfrak{M}$ transitive, euclidean and serial.

Like the glue logic, we will make CL dynamic by extending a dynamic PAL (Baltag et al., 1999); the extensions are once again designed to support reasoning about the *default* consequences

[8] Utterance (2c′) is to be distinguished from a defence that A is misunderstood; e.g., *I didn't mean that it's raining.*

of public announcements, this time including (default) links to cogntive states. To ensure that CL reflects the commitments in DSDRSs, we assume that agents announce to the dialogue participants certain commitments to SDRS-formulae (or, more accurately, their shallow form as determined by τ). Thus a speaker a uttering K_π to D will result in CL-based reasoning with the modality $[!\mathcal{P}_{a,D}\tau(K_\pi)]^{cp}$. As in GL, *ceteris paribus* announcements in CL are defined in terms of a modal connective $>$; thus CL models $\mathfrak{M}$ include a function $*$ from worlds and propositions (i.e., $W \times W$) to propositions, which defines normality and is used to interpret $\phi > \psi$.

Our basic announcements in CL bring about a particular sort of transition on models, one which updates the alternativeness relation for commitments to include the consequences ψ of a ceteris paribus announcement, so long as adding ψ to the commitments is consistent. Actually, given the way we have set things up, each turn commits a speaker to commitments from earlier turns, unless he disavows one of those commitments, and as we'll see shortly, this glue-logic constraint on constructing DSDRSs is reflected in CL too. We define the model transition for commitments and for beliefs below:

- $\mathfrak{M} \mapsto \mathfrak{M}_{\phi,a,D} : R^{\mathcal{P}_{a,D}}_{\phi,a,D} = (?\top; R^{\mathcal{P}_{a,D}}; ?\phi)$
- $\mathfrak{M} \mapsto \mathfrak{M}_{\flat_a\phi}$: $R^{\mathcal{B}_a}_{\flat_a\phi} = (?\top; R^{\mathcal{B}_a}; ?\phi)$

The following clause defines the interpretation of announcements of commitments. In words, should an agent say ϕ to D, then his commitments are updated, to include the nonmonotonic consequences of this announcement (so long as this is consistent with the input model):

- `Announcements of Commitment:`
  $\mathfrak{M}, w \models [!\mathcal{P}_{a,D}\phi]^{cp}\psi$ iff $\mathfrak{M}^{cp(\phi)}_{\phi,a,D}, w \models \psi$

We also assume that any GL inferences χ about the illocutionary effects of announcing ϕ are consequences of the announcement in CL. More formally:

- `Transferring Commitments:`
  $$\frac{\Gamma \vdash_{\mathrm{GL}} [!\phi^{a,D}]\chi}{\Gamma \vdash_{\mathrm{CL}} [!\mathcal{P}_{a,D}\phi]\mathcal{P}_{d,D}\chi}$$

In words, all agents are publicly committed to (the shallow form of) their SDRSs for the turn where they announce ϕ, as inferred via discourse update in the glue logic GL. The formulae ϕ (and χ) from `Transferring Commitments` will typically be partial descriptions of rhetorical connections among labels in a DSDRS, as well as partial descriptions of other constructors (e.g., quantifiers, the imperative or interrogative modalities, predicate symbols in the SDRS vocabulary, terms, and so on).

As we said earlier, the glue logic ensures that the ceteris paribus consequences of the current announcement include speech acts from prior turns that are ongoing commitments. This is why the dynamic interpretation of the DSDRS overall is that of its last turn, with the input context of evaluation being the dialogue-initial state. Therefore, since the formula χ in `Transferring Commitments` is a shallow representation of all the relational speech acts that are current commitments, including those that are ongoing commitments from prior turns, the input model $\mathfrak{M}$ that is transformed by the current announcement should be the one that represents the dialogue initial state, rather than the output model from the prior turn. This way, the public commitments in CL reflect the semantic properties of the DSDRSs. But it does not reflect *all* the semantic properties (e.g., the existentially quantified formulae), to ensure that CL remains decidable. For instance, consider the logical form of dialogue (2) in Table 2. The above transfer rules for computing A's commitments in CL from turn 2 ensures that the output model satisfies $\mathcal{P}_{A,\{A,B\}}\phi_{\pi_1}$,where ϕ_{π_1} is the shallow CL-representation of *It's raining.* But the quite different output model from turn 3 is one that satisfies $\mathcal{P}_{A,\{A,B\}}(\neg\phi_{\pi_1} \wedge \phi_{\pi_2})$, where ϕ_{π_2} is the shallow representation of *Oh, you're right (that it's not raining).* A's public commitments have changed dynamically, but we can only detect this by comparing the output CL-model from turn 2 vs. that from turn 3.

The following three definitions, of CL's models, language and interpretation, summarises the technical discussion so far:

Definition 6 Models of CL

A model $\mathfrak{M}$ of CL is a tuple $\langle W^{\mathfrak{M}}, *^{\mathfrak{M}}, R^{\mathcal{P}_{a,D}}, R^{\mathcal{B}_a}, R^{\mathcal{I}_a}, V^{\mathfrak{M}}\rangle$ where:

- $W^{\mathfrak{M}}$ is a set of possible worlds.
- $*$ is a function from the worlds and propositions (i.e., subsets of $(W^{\mathfrak{M}} \times W^{\mathfrak{M}})$ to propositions.
- For each agent a and group D, $R^{\mathcal{P}_{a,D}} \subset W^{\mathfrak{M}} \times W^{\mathfrak{M}}$ is an accessibility relation on worlds that validates the axioms K45, Ax1 and Ax2.
- For each agent a, $R^{\mathcal{B}_a} \subset W^{\mathfrak{M}} \times W^{\mathfrak{M}}$ is an accessibility relation on worlds that validates KD45 (so it is transitive, serial and euclidean).
- For each agent a, $R^{\mathcal{I}_a} \subset W^{\mathfrak{M}} \times W^{\mathfrak{M}}$ is an accessibility relation on worlds that matches the axioms D4 (so it is serial and euclidean).
- For each agent a, R^{Pref_a} is a transitive linear order (we'll see in Section 4.4 that this relation is defined in terms of $*$).
- V is a valuation function which assigns to each atomic formulae of CL a subset of $W^{\mathfrak{M}}$.

We have already discussed how announcements in CL effect a model transition, updating commitments. We now define the full syntax and semantics of CL.

Definition 7 The CL Language

The language of CL is defined as follows:

1. Where ϕ is an SDRS-formula, $(\phi)^{\tau}$ is a formula of CL (see Definition 5).
2. All GL-formulae are CL-formulae.
3. If ϕ and ψ are CL formulae, then $\neg\phi$, $\phi \wedge \psi$, $\phi > \psi$, $\mathcal{P}_{a,D}\phi$, $\mathcal{I}_a\phi$, $\mathcal{B}_a\phi$ and $Pref_a(\phi, \psi)$ (meaning agent a prefers ψ to ϕ) are all formulae.

4. If ϕ and ψ are CL formulae, then $\flat_a\phi$, $\sharp_a\phi$ and $\heartsuit_a(\phi,\psi)$ are all formulae ($\flat_a$, $\sharp_a$ and $\heartsuit_a$ are action operators that effect an update on beliefs, intentions and preferences respectively).
5. If ϕ and ψ are CL formulae, then $[!\mathcal{P}_{a,D}\phi]\psi$ and $[!\mathcal{P}_{a,D}\phi]^{cp}\psi$ are CL formulae.

Definition 8 The Semantics of CL

Let $\mathfrak{M}$ be a CL model. Then we start by interpreting 'static' fragments of CL:

- For atomic CL-formulae ϕ $\llbracket\phi\rrbracket^{\mathfrak{M}} = \{w \in W^{\mathfrak{M}} : V^{\mathfrak{M}}(\phi)(w) = 1\}$
- $\llbracket\phi \wedge \psi\rrbracket^{\mathfrak{M}} = \llbracket\phi\rrbracket^{\mathfrak{M}} \cap \llbracket\psi\rrbracket^{\mathfrak{M}}$
- $\llbracket\neg\phi\rrbracket^{\mathfrak{M}} = W^{\mathfrak{M}} \setminus \llbracket\phi\rrbracket^{\mathfrak{M}}$
- $\llbracket\phi > \psi\rrbracket^{\mathfrak{M}} = \{w : *^{\mathfrak{M}}(w, \llbracket\phi\rrbracket^{\mathfrak{M}}) \subseteq \llbracket\psi\rrbracket^{\mathfrak{M}}\}$
- $\llbracket\mathcal{P}_{a,D}\phi\rrbracket^{\mathfrak{M}} = \{w : \forall w' \text{ st } R^{\mathcal{P}_{a,D}}(w,w'), w' \in \llbracket\phi\rrbracket^{\mathfrak{M}}\}$
- $\llbracket\mathcal{I}_a\phi\rrbracket^{\mathfrak{M}} = \{w : \forall w' \text{ st } R^{\mathcal{I}_a}(w,w'), w' \in \llbracket\phi\rrbracket^{\mathfrak{M}}\}$
- $\llbracket\mathcal{B}_a\phi\rrbracket^{\mathfrak{M}} = \{w : \forall w' \text{ st } R^{\mathcal{B}_a}(w,w'), w' \in \llbracket\phi\rrbracket^{\mathfrak{M}}\}$
- $\llbracket Pref_a(\phi,\psi)\rrbracket^{\mathfrak{M}} = \{w : R^{Pref_a}(w, \llbracket\phi\rrbracket^{\mathfrak{M}}, \llbracket\psi\rrbracket^{\mathfrak{M}}\}$ [9]

We now define model transitions, which are used to define the action operators (i.e., announcements, *ceteris paribus* announcements, $\flat_a$, $\heartsuit_a$ and $\sharp_a$):

- $\mathfrak{M} \mapsto \mathfrak{M}_{\phi,a,D} : R^{\mathcal{P}_{a,D}}_{\phi,a,D} = (?\top; R^{\mathcal{P}_{a,D}}; ?\phi)$
- $\mathfrak{M} \mapsto \mathfrak{M}_{\flat_a\phi}$: $R^{\mathcal{B}_a}_{\flat_a\phi} = (?\top; R^{\mathcal{B}_a}; ?\phi)$
- $\mathfrak{M} \mapsto \mathfrak{M}_{\sharp_a\phi}$: $R^{\mathcal{I}_a}_{\flat_a\phi} = (?\top; R^{\mathcal{I}_a}; ?\phi)$
- $\mathfrak{M} \mapsto \mathfrak{M}_{\sharp_a\phi}$: $R^{\mathcal{I}_a}_{\flat_a\phi} = (?\top; R^{\mathcal{I}_a}; ?\phi)$
- $\mathfrak{M} \mapsto \mathfrak{M}_{\heartsuit_a(\phi,\psi)}$: $R^{\mathcal{I}_a}_{\heartsuit_a(\phi,\psi)} = (?\top; R^{Pref_a}; ?(\phi,\psi))$

The dynamic component of CL effects the following model transitions:

- $\mathfrak{M}, w \models [!\mathcal{P}_{a,D}\phi]\psi$ iff $\mathfrak{M}^{\phi}_{\phi,a,D}, w \models \psi$
- $\mathfrak{M}, w \models [!\mathcal{P}_{a,D}\phi]\psi$ iff $\mathfrak{M}^{cp(\phi)}_{\phi,a,D}, w \models \psi$
- $\mathfrak{M}, w \models \flat_a\phi$ iff $\mathfrak{M}_{\flat_a\phi}, w \models \mathcal{B}_a\phi$
- $\mathfrak{M}, w \models \sharp_a\phi$ iff $\mathfrak{M}_{\sharp_a\phi}, w \models \mathcal{I}_a\phi$
- $\mathfrak{M}, w \models \heartsuit_a(\phi,\psi)$ iff $\mathfrak{M}_{\heartsuit_a(\phi,\psi)}, w \models Pref_a\phi$

We have already discussed how this semantics in CL, together with the axiom `Transferring Commitments`, ensures that a's public commitments in CL reflect his commitments as represented in the dynamic interpretationof dialogue content. In addition, we can now add axioms to CL to effect changes to a's other mental attitudes as the dialogue proceeds. This involves stating axioms that link a's public commitments commitments to other mental attitudes, especially preferences.

[9] In Section 4.4, this will be defined as a $>$-formula, and hence its semantics is definable by $*$ in the model.

4.1 Other Attitudes: the primacy of preference

As we argued in Section 1, an adequate link between discourse interpretation and cognitive states must separate public commitments from private belief, hence agreement from mutual belief. It must also separate commitments to preferences and to intentions from the private, actual preferences and intentions. In Asher and Lascarides (2003), we followed most of the extant work on dialogue and attitudes by focussing on the links between what we're now calling commitments and the attitudes of belief and intention. We ignored preferences. This was a consequence of buying in at least partly to the Gricean mentalist perspective, according to which conversation is, above all, a matter of information transfer and of the cooperative realization of joint intentions.

We now realize that neglecting preferences is a mistake. First of all conversation is pursued for many ends, not just information transfer or the coordination of joint intentions. Furthermore, from our present perspective the transfer principles for commitments to other attitudes are themselves based on whether the conversational participants have preferences that cohere with each other in a particular way. What we mean by this will be clearer after we look a bit closer at preferences.

Preferences are distinct from intentions. First, preferences can persist even after they are realised; intentions do not. Secondly, preferences can be contrary to fact: one might prefer to be skiing rather than in a meeting, while nevertheless being at the meeting. More generally, intentions are an outcome of a deliberation about preferences and what is believed to be the case. The role of preferences in action is quite general: people normally intend to do things that maximise their preferences. In particular, an agent's utterances generally follow and reflect his preferences.

Preferences influence dialogue moves, and the dialogue moves one observes another agent do are also a valuable source of information for inferring his preferences. This information flow between preferences and dialogue moves allows an agent to make decisions about what dialogue move to do next, based on his calculations about which moves maximise his preferences and those of others. Thus an important component of dialogue move analysis concerns reasoning about what move will maximise one's preferences.

Such reasoning almost inevitably involves reasoning about other persons's preferences and this involves reasoning of the sort described by classic game theory. A game consists of a set of players and a set of strategies. Each strategy is assigned a real valued payoff or utility for each player. Typically the payoff for each player is a function of the player's own strategy and the strategies of the other players. Classic game theory assumes common knowledge of all strategies and payoffs, but this assumption can be relaxed. For instance, in signalling games, it is often assumed that player 1 (sender) knows which state she is in whereas player 2 does not—and thus assigns probabilities to the states that 1 may signal about. A *Nash Equilibrium* (NE) is a combination of strategies that is optimal in the sense that no player of the game has a reason to deviate unilaterally from that strategy. A *Strict Nash Equilibrium* is one in which each strategy in the combination of the NE is uniquely optimal.

Games naturally fit into a description of conversational moves. An example of a dialogue game is shown in Table 4. R(ow) and C(olumn) are considering putdown moves (P_R and P_C) vs. non-putdown moves in a conversation motivated by showing off. In effect this game is

2/1	P_C	$\neg P_C$
P_R	$0, 0$	$3, -3$
$\neg P_R$	$-3, 3$	$4, 4$

Table 4: Simple Putdown Game

a dynamic one where either R or C goes first with the other responding. It, and those that follow, merit an extensive form representation, but we will ignore this for reasons of space. The first column is where C plays a putdown move while in the second he does not; similarly for agent R in rows 1 and 2. The cells indicate the utilities or values for agents R and C respectively for each combination of moves. Note how the utitility values for both R and C are influenced by what *both* agents do. By specifying the utilities for each player on all strategies, the game describes the complete preferences of each player on all strategies. The game from Table 4 has two NE: $(\neg P_R, \neg P_C)$ and (P_R, P_C).

It is worthwhile reflecting on the status of the Gricean maxims, in particular the maxim of quality, which Grice (and Lewis) take to be basic and which we formalized in terms of Sincerity and Competence in Asher and Lascarides (2003). These function as transfer principles from commitments to private attitudes. They are a formalization of the Mentalist approach, and an heir to Grice. Given our dynamic CL, Sincerity and Competence would look like this:

- Sincerity: $\mathcal{P}_{a,D}\phi > \flat_a\phi$
- Competence: $\mathcal{B}_b\mathcal{B}_a\phi > \flat_b\phi$

In the same vein we should reflect on the status of similar transfer and cooperative principles for intentions of Asher and Lascarides (2003), which we rewrite in our new formalism:

- Sincerity for Intentions: $\mathcal{P}_{a,D}\mathcal{I}_a\phi > \sharp_a\phi$
- Cooperativity: $(b \in D \wedge \mathcal{P}_{a,D}\mathcal{I}_a\phi) > \sharp_b\phi$

These defeasible principles and the maxims that underlie them seem largely irrelevant in this conversational context. Regardless of whether or not R or C believe what they say, the preferences will lead them to say one thing in one situation and perhaps something completely contrary in another. Yet the situations do not affect what they believe, and so they will most likely violate the maxim of quality in this game. Similarly Cooperativity clearly is not at work. Suppose R announces an intention that she will play a putdown move. C may not intend that R play that putdown move. For example, if this were a cooperative game where agreements prior to play are possible, then C might even try to argue R out of her intention to get to an agreement on playing $\neg P$.

To make the point even clearer, consider again the situation of a cross examination of a defendant in a trial by the prosecutor. The prosecutor might have two options, asking a direct question like *Did you commit the crime?* or asking a question that will attempt to lull the defendant into thinking of the prosecutor as being on his side. Let us assume that the defendant has two options—to tell the truth or lie. If he has no wish to confess, he will

L/D	T	L
D	$3, -3$	$1, -1$
PI	$-1, 1$	$2, -2$

Table 5: The Courtroom Game

S/T	A	$\neg A$
Q	$3, 3$	$-3, -3$
$\neg Q$	$1, 1$	$0, 0$

Table 6: The Student Teacher Game

most clearly lie when confronted with the direct question. We can imagine a description of the respective players (L/D) and their preferences for the direct question D and the indirect question I, and either Lying (L) or Truthtelling (T) in Table 5. The game from Table 5 has two NE: $(\neg I, T)$ and (D, L). In such a situation the defendant will lie if faced with the prosecutor's direct question. Let us suppose that the prosecutor asks the direct question with the intention of getting the answer he wants. Clearly the defendant will not take on that intention. He will lie in an effort to thwart the prosecutor's intention. Sincerity and Cooperativity are irrelevant here in such truly adverserial dialogues, although work in game theory has been done to show that some information offered is still credible.

On the other hand, there are plenty of conversations where the defaults we have specified above are operative. Let us take a familiar example to those who work on dialogue systems. Suppose we are in a situation where a pupil is attempting to solve a problem with the help of a teacher. They both want the student to solve the problem. Suppose the student asks a question of the teacher with the intention of getting a true and helpful answer. The teacher will follow Cooperativity and take that intention on board, producing an answer that is true and as helpful as she can make it. The student upon hearing the teacher's answer will naturally take her to believe her answer and by Competence make it his own belief. Such sort of dialogues are what we might call *defeasibly payoff symmetric* in that the players have roughly the same utilities for the same actions. Let's just look at one conversational turn in such a dialogue, where the student asks a question or doesn't ask a question and the teacher provides the answer or doesn't. The game described in Table 6 is completely payoff symmetric. This game has only one NE (Q, A). It would appear that in such games the maxims are in force as are all of the defaults. They are not violated often, if at all. Because defeasible payoff symmetry doesn't have a nice ring to it, we'll call conversational situations that meet the defeasibly payoff symmetry requirement *Grice Cooperative* or *GC*. Actually, GC is not a black and white affair though we will treat it so. Preferences can be more or less aligned; as Crawford and Sobel (1982) show the more preferences are aligned, the more the information is credible. Because we have a defeasible symmetry, we can let many conversations be GC even if speakers have different utilities on quite a few strategy profiles.

We now rewrite our transfer and cooperativity principles with GC in mind.

- `Sincerity Revised:` $(\mathcal{P}_{a,D}\phi \wedge GC) > \flat_a \phi$

- `Competence Revised:` $(\mathcal{B}_b\mathcal{B}_a\phi \wedge GC) > \flat_b\phi$
- `Sincerity for Intentions Revised:` $(\mathcal{P}_{a,D}\mathcal{I}_a\phi \wedge GC) > \sharp_a\phi$
- `Revised Cooperativity:` $(b \in D \wedge \mathcal{P}_{a,D}\mathcal{I}_a\phi \wedge GC) > \sharp_b\phi$

We can prove that these defaults hold, and we can prove slight generalizations of them if we look at the boundary conditions for the stabilization of the use of the simple defaults in the sort of game set up considered by Asher, Sher and Williams (2001).[10] But even these principles are defaults because of examples like (1). In the context of (1), we might well assume that we are in a GC situation. Let p be the (shallow) CL representation of *C is stupid*. Then given A's SDRS for the third turn, `Transferring Commitments` yields $\mathcal{P}_{A,\{A,B\}}(p \wedge \neg p)$ in CL. This satisfies the antecedent to `Sincerity`, but $\flat_A(p \wedge \neg p)$ is not inferred because $\mathcal{B}_A(p \wedge \neg p)$ is inconsistent ($\mathcal{B}_A$ satisfies axiom D). $\mathcal{P}_{A,\{A,B\}}p$ and $\mathcal{P}_{A,\{A,B\}}\neg p$ are also true (by axiom K); they both satisfy the antecedent of `Sincerity`, but their consequences are mutually inconsistent, and so neither is inferred. Thus, A has made a contradictory commitment, but remains rational. And assuming that all agents mutually believe the CL axioms, B detects from A's inconsistent current commitments that he's lying, and without further information B does not know what A believes: p, or $\neg p$, or neither.

Let's now turn to the issue of belief transfer and mutual belief. As is standard, mutual believe ($MB_{x,y}\phi$) is defined in terms of belief using a fixed point equation:

- `Mutual Belief:` $MB_{x,y}\phi \leftrightarrow (\mathcal{B}_x(\phi \wedge MB_{x,y}\phi) \wedge \mathcal{B}_y(\phi \wedge MB_{x,y}\phi))$

On this definition, $MB_{x,y}\phi$ entails an ω sequence of nested belief statements: $\mathcal{B}_y\phi, \mathcal{B}_y\mathcal{B}_x\phi, \ldots$ and $\mathcal{B}_y\phi, \mathcal{B}_x\mathcal{B}_y\phi, \ldots$. We will denote a formula that starts with $\mathcal{B}_x$, and alternates with $\mathcal{B}_y$ to a nesting of depth n as $\mathcal{B}^n_{(x,y)}\phi$. It is straightforward to verify that the following scheme is sound.

- `Induction Scheme for Mutual Belief:`

$$\begin{array}{rl}
\text{Assume} & \Gamma\mid\!\sim\mathcal{B}_y(\phi \wedge \mathcal{B}_x\phi) \wedge \mathcal{B}_x(\phi \wedge \mathcal{B}_y\phi) \\
\text{and for any } n\text{:} & \dfrac{\Gamma\mid\!\sim\mathcal{B}_y(\phi \wedge \mathcal{B}^n_{(x,y)}\phi) \wedge \mathcal{B}_x(\phi \wedge \mathcal{B}^n_{(y,x)}\phi)}{\Gamma\mid\!\sim\mathcal{B}_y(\phi \wedge \mathcal{B}^{n+1}_{(x,y)}\phi) \wedge \mathcal{B}_i(\phi \wedge \mathcal{B}^{n+1}_{(y,x)}\phi)} \\
\text{Then:} & \Gamma\mid\!\sim MB_{i,j}\phi
\end{array}$$

We're now in a position to prove that grounding a propoosition leads defeasibly to mutual belief are linked, but unlike the BDI account we do not postulate an equivalence or even a one way entailment between grounding and mutual belief. Where $D = \{x, y\}$, the **proof** that $\mathcal{P}_{\{x,y\},D}\phi\mid\!\sim MB_{x,y}\phi$ is as follows (we assume that the background conversational situation is GC).

[10] Actually for Competence Revised, we would need to ensure more than just GC, we would need to make sure that the penalties for being wrong are high enough to affect the utility of making a particular discourse move. See also Hurd (1995) and Lipman (2003). One actually

1. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_x\phi$ `Sincerity`
2. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_y\phi$ `Sincerity`
3. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_y\mathcal{B}_x\phi$ 1; cognitive axioms are mutually believed
4. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_y(\phi \wedge \mathcal{B}_x\phi)$ 2, 3; $\mathcal{B}$ is KD45
5. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_x\mathcal{B}_y\phi$ 2; cognitive axioms are mutually believed
6. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim \mathcal{B}_x(\phi \wedge \mathcal{B}_y\phi)$ 1, 5; $\mathcal{B}$ is KD45
7. $\mathcal{P}_{\{x,y\},D}\phi \wedge GC \mid\!\sim MB_{x,y}\phi$ 4,6; Axioms are mutually believed and induction scheme

□

Thus grounded propositions are normally mutually believed; e.g., it is in (3) and (2), but not (1). In fact, agreement is not a necessary condition for mutual belief, as well as not being a sufficient one.

Together with `Sincerity` and `Competence` and the assumption that the conversational situation is GC, we can model how an assertion leads to belief transfer, and ultimately to mutual belief (by default). The **proof** of this starts with the premise $\mathcal{P}_{a,D}\phi$ and that the conversation is GC, where $D = \{a, b\}$, for this is the cognitive hallmark of a asserting ϕ, and we now show that $\mathcal{P}_{a,D}\phi \mid\!\sim MB_D\phi$:

1. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_a\phi$ `Sincerity`
2. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_b\phi$ 1; `Competence`
3. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_a\mathcal{B}_b\phi$ 2; CL is mutually believed
4. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_a(\phi \wedge \mathcal{B}_b\phi)$ 1,3; $\mathcal{B}_a$ is KD45
5. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_b\mathcal{B}_a\phi$ 1, CL is mutually believed
6. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim \mathcal{B}_b(\phi \wedge \mathcal{B}_a\phi)$ 2,5; $\mathcal{B}_b$ is KD45
7. $\mathcal{P}_{a,D}\phi \wedge GC \mid\!\sim MB_D\phi$ 4,6; Induction Scheme

□

The above proof shows how an individual's public commitment to ϕ yields a default inference that ϕ is mutually believed among all the dialogue participants assuming GC. But ϕ isn't *grounded* unless *all* agents publicly commit to it: that is, they must all make an announcement, from which ϕ nonmonotonically follows. Hence mutual belief is not a sufficient condition for grounding; nor is it a necessary one (as shown in our analysis of (1)).

Given that the various defaults we have set up have game theoretic justifications, the reader is probably wondering why we are even using these defaults once we have game theory? The reason is that these principles can be used as shorthand rules, Millian abstractions over the basic calculus of utility. Below we will argue (as we already claimed in the introduction) that we can't really use classical game theory because the information and computational capacities demanded by it are not reasonable demands to place on ordinary conversational agents. The shorthand rules for inferring attitudes from commitments are useful, however, even when the precise information pertaining to preferences that we have used here for illustrative purposes is lacking.

4.2 Other Principles

We have so far focussed on public commitment and belief, but dialogue influences, and is influenced by, intentions and desires as well. One intuitively compelling link is `Intent to Ground`: if a commits to ψ, then normally he commits to the intention that his interlocutors commit to it too, if they have not done so already:

- `Intent to Ground:` $(b \in D \wedge \mathcal{P}_{a,D}\phi \wedge \neg\mathcal{P}_{b,D}\phi) > \mathcal{P}_{a,D}\mathcal{I}_a\mathcal{P}_{b,D}\phi$

This default axiom defines just one (default) illocutionary purpose for *all* types of speech acts—i.e., an intention that the contribution be grounded. As is well known from Searle (1969), different types of speech acts are linked to other intentions as well; e.g., the (default) intention behind a question is to know its answer. We have formalised such links in SDRT elsewhere (e.g., Asher and Lascarides (2001, 2003)).

We'll see shortly that `Intend to Ground` and `Cooperativity` predict why agents tend to make moves that ground each other's contributions. In addition, the axioms of `Cooperativity`, `Sincerity for Intentions`, introspection on intentions ($\mathcal{I}_a\phi \rightarrow \mathcal{I}_a\mathcal{I}_a\phi$), and axioms that link various speech act types to their illocutionary purpose all ensure that the intentions behind a's (current) announcement become by default the intentions of all agents in D. Thus what one agent says in a dialogue can affect another agent's subsequent behaviour.

Asher and Lascarides (2003) use Aristotle's Practical Syllogism to link beliefs, intentions and the *choice* that marks one's preferred way of achieving goals:

- `Practical Syllogism (PS):`
  $(\mathcal{I}_a(\psi) \wedge \mathcal{B}_a((\phi > \psi) \wedge \mathit{choice}_a(\phi, \psi)) >$
  $\mathcal{I}_a(\phi)$

This is a general axiom that permits an agent to plan his actions so as to achieve his goals. In words, if a intends that ψ, and he believes that ϕ normally leads to ψ and moreover ϕ is a's choice for achieving ψ, then normally a intends that ϕ. `PS` was used to infer an agent's beliefs and intentions from his behaviour, and *vice versa*. For instance, Asher and Lascarides (2003) demonstrate its role in validating the glue-logic axioms `Q-Elab` and `IQAP` in CL. So even though these defaults for inferring speech acts depend only on sentence mood, their justification exlpoits reasoning about mental states.

In this context, one can imagine, for instance, that the preferred way of achieving a's intention to commit to D to content ϕ is for a to make a *ceteris paribus* announcement ψ with consequences $\mathcal{P}_{a,D}\phi$: in other words, there is some announcement ψ such that $\mathcal{B}_a[!\mathcal{P}_{a,D}\psi]^{cp}\phi$ and $\mathit{choice}_a(!\mathcal{P}_{a,D}\psi, \mathcal{P}_{a,D}\phi)$. The `Practical Syllogism` then predicts that the agent a will achieve his intention to commit to ϕ by uttering ψ. With this assumption about ways to achieve commitments, we're now in a position to see why agents tend to make moves that ground each others contributions. For suppose that a makes an announcement to D that entails $\mathcal{P}_{a,D}\phi$. Then by `Intent to Ground`, $\mathcal{P}_{a,D}\mathcal{I}_a\mathcal{P}_{b,D}\phi$ follows. So by `Sincerity of Intentions` and `Cooperativity`, $\mathcal{I}_a\mathcal{P}_{b,D}\phi$ follows. Using `PS`, and the assumption that one chooses to make public commitments through announcements, it will follow that there is some announcement ψ' such that $\mathcal{I}_b[!\mathcal{P}_{b,D}\psi']\phi$. In other words, b will make an annoucement which results in the shared public commitment that ϕ, making ϕ grounded.

`PS` is incomplete, because the relation choice_a is treated as primitive, with CL lacking the reasoning that agents engage in for finding optimal ways of achieving goals. In the example above, where the intention is to commit to some content ϕ, there may be many announcements

ψ that will achieve this goal. By treating *choice*$_a$ as a primitive predicate symbol, CL fails to model reasoning about which of these announcements would be best (for his purposes). So we need to add to CL axioms for reasoning about *preferences*, and define *choice* in terms of this. This means that we'll have to tackle preference directly within CL.

4.3 Representing preferences efficiently

A game problem is (minimally) a goal for each player, a set of strategies for attaining that goal and a preference ordering over the strategies.[11] So we need to be able to represent preferences of agents. In standard game theory, a player's preferences or utilties is a function from a strategy profile (a proposed action by each player) to a real number—standard game theory provides calculations of expected utility that combine probabilities over actions with the preferences for each player. This sort of calculation is far too complex to be part of CL, which is a shallow logic for rough and ready decisions about discourse moves. In addition, the preferences of one's conversational partners are often known with imprecision, and sometimes not at all.

To maintain a computationally effective CL, we need a model of simple strategic reasoning that nevertheless approximates the types of interactions between expected moves and utility that game theory addresses. but that moves away from the quantitative representation of standard game theory to a more ordinal conception. Can this be done? The short answer is "yes."

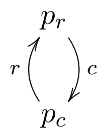
There exist several simpler representations available for strategic reasoning. *CP-nets* (Boutilier et al., 1999, 2004, Domshlak, 2002) provide one such qualitative model for Boolean games (Bonzon, 2007)—games where like Table 4 each player controls propositional variables which he or she can make true or false (think of these as descriptions of actions that the agent performs, or not).[12] CP-nets can be used to express preference relations in a relatively compact manner, and reasoning with them is shown to be efficient—they are designed to fully exploit the independence among the various conditions that affect an agent's preferences.

A CP-net for an individual agent has two components: a directed *conditional preference graph* (CPG), which defines for each feature F its set of parent features $P(F)$ that affect the agent's preferences among the various values of F; and a *conditional preference table* (CPT), which specifies the agent's preferences over F's values for every combination of parent values from $P(F)$. The CP-net for a Boolean game consists of a CP-net for each player. The CPG for each player identifies the parent features for those features over which the agent has control (e.g., for R in the put-down game, we need the features that influence his preferences for the put-down move p_r over which he has control, and similarly for C). The CPT for each player then reflects the relative utilities of his moves in the various conditions imposed by the parent features.

For example, the CP-net for the 'put down' game from Table 4 is shown in Figure 1. p_c stands for C doing a put down move; similarly for p_r. The dependencies among features for

[11] When there is only one player we have a decision problem.

[12] We could alternatively use a propositional operator for preferences, following van Benthem et al. (2005). However, standard modal operators do not allow us to define strict preferences (i.e. an asymmetric relation), and that seems to be rather unintuitive.



The CPG

Preferences for C:		$\neg p_r \succ_c p_r$
	p_r:	$p_c \succ_c \neg p_c$
	$\neg p_r$:	$\neg p_c \succ_c p_c$
Preferences for R:		$\neg p_c \succ_r p_c$
	p_c:	$p_r \succ_r \neg p_r$
	$\neg p_c$:	$\neg p_r \succ_r p_r$

The CPT

Figure 1: The CP-net for Table 4's 'Put Down' Game.

2/1	P_C	$\neg P_C$
P_R	$0, 0$	$3, -3$
$\neg P_R$	$-3, 3$	$0, 0$

Table 7: Second Put Down Game

each agent are shown with labelled arcs in the CPG. The CPT then distinguishes among the conditional preferences for agents R and C; e.g., $\neg p_r : \neg p_c \succ_c p_c$ stipulates that C prefers not to put down R rather than put him down, if R does not put down C. The 'joint' CP-net in Figure 1 is compact, and the method for detecting best moves is no worse than the complexity of the rest of CL. The semantics of CP-nets ensures that its conditional *ceteris paribus* preferences generate a total order $\succeq$ over all possible combinations of values of all features. Roughly put, the logic of CP-nets adheres to the following two (ranked) principles when generating this total order: first, one prefers values that violate as few conditional preferences as possible; and second, violating a (conditional) preference on a parent feature is worse than violating the preference on a daughter feature. So the total preference orderings for R and C for the CP-net in Figure 1 is as follows:

(7) $$(\neg p_r \wedge \neg p_c) \succ_c (\neg p_r \wedge p_c) \succ_c (p_r \wedge p_c) \succ_c (p_r \wedge \neg p_c)$$
$$(\neg p_r \wedge \neg p_c) \succ_r (p_r \wedge \neg p_c) \succ_r (p_r \wedge p_c) \succ_r (\neg p_r \wedge p_c)$$

In line with the game in Table 4, the orderings on $\succ_c$ and $\succ_r$ yield two NE: $(\neg p_r \wedge \neg p_c)$ and $(p_r \wedge p_c)$.[13]

Changing the original preferences slightly so as to make the game completely symmetric (as shown in Table 7) yields different Nash equilibria, and this is also reflected in its CP-net. Here, although the graph dependencies are the same as in Figures 1, the conditional preference statements corresponding to the game in Table 7 are different:

[13] That neither R nor C should deviate from $(\neg p_r \wedge \neg p_c)$ should be obvious from the fact that it's the most preferred state for both R and C. R should not deviate from his strategy p_r, given that C has chosen p_c, because $(p_r \wedge p_c) \succ_r (\neg p_r \wedge p_c)$. Similarly, C should not deviate from p_c, given that R has chosen p_r, because $(p_r \wedge p_c) \succ_c (p_r \wedge \neg p_c)$. Hence $(p_r \wedge p_c)$ is also a Nash Equilibrium.

(8)
$$\begin{array}{lrl}
\text{Preferences for C:} & & \neg p_r \succ_c p_r \\
 & p_r : & p_c \succ_c \neg p_c \\
 & \neg p_r : & p_c \succ_c \neg p_c \\
\text{Preferences for R:} & & \neg p_c \succ_r p_c \\
 & p_c : & p_r \succ_r \neg p_r \\
 & \neg p_c : & p_r \succ_r \neg p_r
\end{array}$$

This CP-net yields different total orders over the states for both agents:

(9)
$$\begin{array}{l}
(p_r \wedge \neg p_c) \succ_r (\neg p_r \wedge \neg p_c) \succ_r (p_r \wedge p_c) \succ_r (\neg p_r \wedge p_c) \\
(\neg p_r \wedge p_c) \succ_c (\neg p_r \wedge \neg p_c) \succ_c (p_r \wedge p_c) \succ_c (p_r \wedge \neg p_c)
\end{array}$$

In line with the game shown in Table 7, the ordering in (9) yields only one Nash equilibrium $(p_r \wedge p_c)$. However, if the scores in Table 7 for $(\neg P_R, \neg P_C)$ are changed from $(0, 0)$ to $(3, 3)$, then the state $(\neg P_R, \neg P_C)$ becomes a (weak) Nash Equilibrium in the game.[14] However, the CPT for this revised game is still (8), and so the weak Nash equilibrium fails to show up in the CP-net representation. This simple example exposes that CP-nets have some limitations, with their predicted Nash equilibria sometimes deviating from those of the games they purport to represent. However, Bonzon (2007) demonstrates that representations of Boolean games as CP-nets captures all pure Nash Equilibria accurately so long as some quite general conditions are met (e.g., the CP-net induces an acyclic graph of dependencies among the game variables).

Game theory tells us what the optimal strategy is for a set of players in case there is a unique Nash equilibrium. But in many cases there are many equilbria. In such cases as already emphasized by Lewis, something other than preferences and intentions must drive an agent to act, and we don't specify what that might be.

One general case where an agent is unable to identify a unique optimal solution arises when he has incomplete knowledge about the other agents' preferences, and hence incomplete knowledge about the joint CP-net for him and his interlocutors. For instance, suppose that agents R and C are engaged in a put-down game. R has the preferences that are represented in Figure 1. But R does not know what C's preferences are. In particular, he does not know if C is a 'non-jerk', with his preferences as shown in Figure 1, or a jerk. By "jerk", we mean that C prefers to perform the putdown move p_c, regardless of the conditions. And so his CP-net is as shown in Figure 2—which shows that there are no dependencies on p_c—as opposed to that in Figure 1.

p_c Preferences for C: $p_c \succ_c \neg p_c$

Figure 2: The CP-net for the put-down game for the jerk C

[14] A *weak Nash equilibrium* (weak NE) is a combination of strategies S such that no other combination of strategies is strictly preferred to S by any player; a *strong Nash equilibrium* (strong NE) is a combination of strategies that is at least as preferred as any other possible combination of strategies by every player. With CP-nets and ordinal preferences in general, there is no guarantee that preferences of agents are total over all the possible actions. This seems cognitively reasonable (see Bonzon (2007) for a discussion). And it motivates the need to distinguish between these two types of Nash equilibria, with the distinction arising in games that encompass incomparable strategies.

If C is indeed a jerk, then the combined CP-nets for R and C would make R's unique optimal move p_r. However, if C is not a jerk, but rather has the preferences shown in Figure 1, then as we have already seen R's unique optimal move is $\neg p_r$. Our definition of $choice_R$ entails that if R does not know whether the C's CP-net is that in Figure 1 or Figure 2, he will have to determine his choice of how to act not only on the basis of his (partial) knowledge of preferences, but by other factors as well.

One of the interesting and well-established results in empirical research on game theory is that the social relationships among the players affect the payoffs and hence the players' behaviours. For instance, Sally (2001), demonstrates that people default to non-risky behavior, but may engage in more risky play after establishing empahty with other players. In a dialogue setting, this means that one's dialogue strategy (or equivalently, the dialogue move one chooses in the dialogue game) is affected by information about other players that's gathered from their previous conversational moves. In turn, this reasoning leads to an analysis of why R might play certain conversational moves in a dialogue—to establish what sort of person C is, or to build trust between C and R for engaging in a cooperative endeavor. In the case we are considering the cooperative endeavor is rather trivial: it is only to mutually pat each other on the back or to refrain from putting the other down. But in fact, such put down games themselves might be useful for establishing what sort of person one is dealing with. R might engage in just such a sort of game, to see how C acts when the stakes are low, before making conversational moves towards other ends where the penalties for lack of cooperative or 'non-jerk'-like behavior are much higher.

4.4 Reasoning about Preferences in CL

Agents need to act even when they lack complete information about other agents' preferences. To do this effectively, an agent must not only be able to exploit to the full the partial information he has about the CP-nets of others, but also use the dialogue to infer more information about their CP-nets. Let's start, then, with the task of representing just a part of a CP-net, in a way that allows an agent to reason with that partial information about preferences within CL. As shown in Lang et al. (2003), one can translate CP-nets into a conditional logic. We can do the same with the weak conditional $>$ that is already a part of CL. In fact, there are several alternative ways of doing it. First, we introduce a predicate OK that labels a world as being a good outcome (Asher and Bonevac, 2005), where OK is always strictly preferred to $\neg OK$.[15]

- `Defining` $\succ_a$ `in terms of` $>$:
  $\phi : \psi \succ_a \neg\psi \Leftrightarrow \phi \rightarrow (\neg((\phi \wedge \psi) > \neg OK_a) \wedge ((\phi \wedge \neg\psi) > \neg OK_a))$

This gives an ordering where some of the normal $\phi\wedge\psi$ worlds are better than all the normal $\phi\wedge\neg\psi$ worlds, which is the ordering that Lang et al. (2003) adopt. Accordingly, the unconditional preference $\psi \succ_a \neg\psi$ is equivalent to $\top \rightarrow (\neg(\psi > \neg OK_a) \wedge (\neg\psi > \neg OK_a))$. From now on we will use $\succ$ within CL as a defined relation symbol; $\phi \succ_a \psi$ is our definition for $Pref_a(\psi, \phi)$,

[15] We can think of six alternative ways of defining the conditional preference $\succ_a$ for agent a, but here we provide the one that seems most reasonable to us at present.

the primitive in the CL language for expressing preferences. We will use both interchangeably below.

We now have a means for translating statements from a CP-net into CL. The langauge in CL is much more expressive than that of CP-nets; we saw that CP-nets cannot represent the game theoretic solution to the game given in Table 7 if the utility for the action (P_R, P_C) is $(3, 3)$. But with the translation into $>$-conditionals, we can define $\sim$ or indifference easily within CL using $\succ_a$ defined as above as: $\neg(\phi \succ_a \psi) \wedge \neg(\psi \succ_a \phi)$. This allows us write down the partial order preferences for both R and C such that it is clear that there are two NEs:

- $((\neg p_c \wedge \neg p_r) \succ_r (p_c \wedge p_r)) \wedge ((\neg p_c \wedge p_r) \sim_r (\neg p_c \wedge \neg p_r)) \wedge$ $((\neg p_c \wedge p_r) \succ_r (p_c \wedge p_r)) \wedge ((\neg p_c \wedge p_r) \succ_r (p_c \wedge \neg p_r))$
- $((\neg p_c \wedge \neg p_r) \succ_c (\neg p_c \wedge p_r)) \wedge ((\neg p_c \wedge \neg p_r) \succ_c (p_c \wedge p_r)) \wedge$ $((p_c \wedge \neg p_r) \sim_c (p_c \wedge \neg p_r)) \wedge ((p_c \wedge \neg p_r) \succ_c (\neg p_c \wedge p_r))$

In order to model how an agent learns another agent's preferences from his utterances, we need certain public announcements to lead to changes in an agent's preference order. We can make sense of updating preferences in the usual way: it means adding one or more conditionals of the form given in clause 3 above, and the action operator $\heartsuit_a$ is designed to effect these updates on the model. We ensure, furthermore, that the accessbility relation R^{Pref_a} in the model (see Definition 6) has the requisite properties that are rendered by its definition in terms of $>$.

Agents can learn about other agents' preferences through conversation, because an agent's utterances can reveal his preferences, just as it can reveal other kinds of private attitudes like beliefs and intentions. For instance, R could learn whether C is a jerk or non-jerk by engaging in a put-down game, and observing how C acts. But reasoning from an agent's utterances to his preferences requires Sincerity about preferences, just as inferring other kinds of attitudes like beliefs and intentions from utterances require this. The Sincerity conditions for preferences look much like those for beliefs and intentions, save that we also allow an agent to infer information about another agent's preferences from his public commitments to *intentions*, as well as to preferences:

- `Sincerity on Preferences:` $(\mathcal{P}_{a,D} Pref_a(\phi, \psi) \wedge GC) > \heartsuit_a(\phi, \psi)$

This axiom is a default because an agent can lie about his preferences just as he can lie about his beliefs. But it does enable an agent to reason defeasibly from an agent's commitment to certain preferences to what his actual preferences are. And doing so will enable an agent to augment his existing partial information about the joint CP-nets that he and his interlocutors share as the dialogue proceeds.

Preferences have links to other attitudes as well. We can use the preferences as defined in a CP-net to refine the predicate *choice* in the Practical Syllogism of Asher and Lascarides (2003). In effect *choice* is defined relative to the current CP-net G for agent i (in fact, G may be a set of CP-nets for several agents that interact, as shown in Figure 1).

- `Definition of Choice:`
  $choice_i(\phi, \psi)$ iff there is a unique optimal solution s for i in his current situation, such

that $s > \psi$ and ϕ is a conjunction of values of features controlled by i that is a part of s.

For example, if agent R intends to coordinate with C by either patting him on the back or putting him down (so this is ψ), and R's knowledge about his preferences and those of C are those in Figure 1, then R's *choice* ϕ for achieving his goal is $\neg p_r$ (i.e., pat C on the back). This stems from the NE that is uniquely optimal (i.e., a state that is preferred above all other NEs). The above definition reveals a particular choice only when there is a *unique* optimal solution.

We can now derive Asher and Lascarides (2003)'s principle of the practical syllogism and our default of Revised Cooperativity. We need one more axiom within CL to do this; it captures the principle that agents are pay off maximizers; intentions to act are determined by those actions that maximize the agent's utility. And in turn given an intention to bring about ψ, if our agent is a utility maximizer, then ψ must be preferred to all other courses of action open to the agent.

- `Maximizing Utility`:

  a $choice_i(\phi, \psi) > \mathcal{I}_i(\phi \wedge \psi)$
  b $\mathcal{I}_i\psi > \psi \succ_i \phi$

Maximizing Utility entails that an agent will have an intention to do ϕ normally if and normally only if ϕ maximizes the agent's utility in the current conversational situation. This provides a very strong link between agents' preferences and intentions. In addition from `Maximizing Utility` allows us easily to derive a rule for inferring preferences.

- `From Commitments to Intentions to Preferences`
  $(\mathcal{P}_{a,D}\mathcal{I}_a\phi \wedge GC) > \heartsuit_a(\phi, \psi)$

`Maximizing Utility` also enables us to derive Asher and Lascarides (2003)'s formulation of the Practical Syllogism in almost trivial fashion:

1. $(\mathcal{I}_a(\psi) \wedge \mathcal{B}_a(\phi > \psi) \wedge choice_a(\phi, \psi)) \mid\!\sim choice_a(\phi, \psi)$ Classical Logic and $\mid\!\sim$ is supraclassical
2. $(\mathcal{I}_a(\psi) \wedge \mathcal{B}_a(\phi > \psi) \wedge choice_a(\phi, \psi)) \mid\!\sim) \mathcal{I}_i(\phi \wedge \psi)$ 1; `Maximizing Utility`
3. $\mid\!\sim (\mathcal{I}_a(\psi) \wedge \mathcal{B}_a(\phi > \psi) \wedge choice_a(\phi, \psi)) > \mathcal{I}_a\phi$ 2; $>$ validates Weak Deduction.[16]

□

Asher and Lascarides (2003) used the Practical Syllogism to infer an agent's beliefs and intentions from his behaviour, and *vice versa*. Now, by incorporating approximations of game-theoretic principles into CL, we can do all this without the Practical Syllogism as a separate principle. Thus by synthesising the game theory and BDI approaches within our logic, we have actually deepened the model of rationality and cooperativity; game-theory principles to a large extend predict why rational and cooperative agents behave the way that they do.

We can also use the powerful link between intentions and preferences to derive the original Cooperativity axiom of Asher and Lascarides (2003). First, note that we can use our representation of preferences to define a GC situation or a game that is defeasibly payoff symmetric.

Definition 9 A game is GC just in case for any of its players a and b:

$$(\phi \succ_a \psi) > (\phi \succ_b \psi)$$

Now suppose we make a rationality assumption, that if you intend ϕ then normally you publicly commit to the intention:

- `Commit to Intentions:` $\mathcal{I}_a\phi > \mathcal{P}_a\mathcal{I}_a\phi$

Then the defining principle for defeasibly symmetric payoff games or GC, together with `Preferences to Intentions`, `Commitments to Preferences` and `Commit to Intentions` makes what Asher and Lascarides (2003) call `Cooperativity` derivable within the logic of CL:

- `Cooperativity:` $\mathcal{I}_a\phi > \mathcal{I}_b\phi$

The **proof** of `Cooperativity` assumes that we start with a cooperative game GC, and an assumption that $\mathcal{I}_a\phi$:

1. $\mathcal{I}_a\phi \mid\!\sim \mathcal{P}_a\mathcal{I}_a\phi$ `Commit to Intentions`
2. $\mathcal{I}_a\phi \mid\!\sim (\phi \succ_a \psi)$ 1; `Maximize Utility`
3. $\mathcal{I}_a\phi \mid\!\sim (\phi \succ_b \neg\phi)$ 2; G is CG;
4. $\mathcal{I}_a\phi \mid\!\sim \mathcal{I}_b\phi$ 3; `Maximize Utility`
5. $\mid\!\sim \mathcal{I}_a\phi > \mathcal{I}_b\phi$ 4; $>$ validates Weak Deduction.

□

This proof leads us to establish as a corollary our default of Revised Cooperativity. Of course in many games and in many dialogue situations, we cannot assume that GC holds. For example the courtoom scenario is one. So we can't derive the original form of Cooperativity and our revised Cooperativity will not apply, since the antecedent of the defeasible rule is not satisfied. What we can infer is that B will intend ϕ, if A does, as long as ϕ is part of $B's$ choice for realizing his preferences. Moreover, we can still use `Intent to Ground` even in non GC situations to infer a commitment to grounding.

Once again, we see that incorporating some simple game-theoretic principles into CL allows us to *predict* axioms of rationality and cooperativity within the logic, rather than simply stating them as axioms of the logic. Here, we have shown how information about another agent's preferences influence his own intentions and those of others, and hence also the next moves in the conversation.

4.5 Inferring commitments to preferences from conversation

Given the centrality of preferences in our new CL logic, it is all important to discover how to extract commitments to preferences from conversational moves. Here we must be cautious. While an inference about an agent's actual preference may stem from his commitment to it, identifying a commitment to a preference can be a highly complex task. Perhaps the simplest case is one where a dialogue participant plainly states a commitment to a preference:

Turn	A's SDRS	B's SDRS
1	$\pi_1 :?\lambda x$dinner-at(x)	$\emptyset$
2	$\pi_1 :?\lambda x$dinner-at(x)	$\pi_{2B} : \mathit{IQAP}(\pi_1, \pi_2) \wedge \mathit{Elaboration}(\pi_2, \pi_3)$
3	$\pi_{3A} : \mathit{Correction}(\pi_{2B}, \pi_4)$	$\pi_{2B} : \mathit{IQAP}(\pi_{1A}, \pi_2) \wedge \mathit{Elaboration}(\pi_2, \pi_3)$
4	$\pi_{3A} : \mathit{Correction}(\pi_{2B}, \pi_4)$	$\pi_{4B} : \mathit{Ackn}(\pi_{3A}, \pi_5) \wedge \mathit{Result}(\pi_5, \pi_6)$

Table 8: The DSDRS for Dialogue (11).

(10) I prefer to eat at Chop Chop.

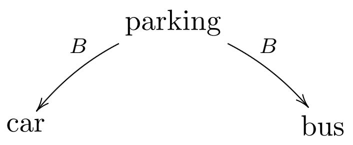
In dialogues within task-oriented domains, commitments to preferences are often directly stated, much as they are in (10). And indeed, detecting commitments to preferences is already a feature of many functioning dialogue management systems (e.g., Traum and Allen (1994)). But often, inferring a commitment to a preference is more complicated because the commitment is less direct, as illustrated in (11). Intuitively, A (and any bystanders) can infer B's commitments to certain preferences given his utterances (π_2, π_3 and π_5). These can in turn be represented as a CP-net over a simple game with three propositional veriables, two of which B has control over, as shown in Figure 3.

(11) π_1. A: Where shall we go to dinner?
π_2. B: Let's go to Chop Chop.
π_3. B: I'll drive.
π_4. A: There's no parking.
π_5. B: OK.
π_6. B: Let's go by bus.

This dialogue's semantic representation is shown in the DSDRS in Table 8. These dialogue moves must have an impact on commitments to preferences: intuitively, B's utterances in turn 2 reveal that B has a decision problem in mind where driving to Chop Chop is a possible option and indeed it is one that he prefers in that problem. In effect he presupposes that there is parking. However, once we update with A's assertion that there's no parking, and B's acknowledgement of that, they are jointly committed to there being no parking. By using a CP-net in which there is a variable for parking, perhaps controlled by A or perhaps just controlled by a third player "nature", B now makes a commitment about preferences relative to $\neg$parking. We can represent B's commitment to preferences, as revealed by this dialogue, using three features, one for parking, one for going by car and one for going by bus with the dependencies among them as shown in the CPG in Figure 3. The conditional preferences are shown in the CPT in Figure 3 too.[17]

A can update his representation of B's conditional preferences by building a CP-net from the responses that B makes to both A's question and to A's *Correction* of the presupposition of

[17] An alternative representation consists of two features, parking and mode-of-transport, where mode-of-transport has three possible values: car, bus or neither. This would 'hardwire' into the preference statements that going by both car and bus is not on option (as far as stating preferences are concerned, at any rate).



The CPG

parking:	car $\succ_B$ ¬car
parking:	¬bus $\succ_B$ bus
¬parking:	bus $\succ_B$ ¬bus
¬parking:	¬car $\succ_B$ car

The CPT

Figure 3: The CPG for (11)

B's answer. This will enable *A* to make the appropriate subsequent moves in the dialogue, depending of course on *A*'s own preferences.

B's answer π_2 to the question π_1 revealed his preference for where to eat dinner (not shown in the CP-net of Figure 3), which concerns only how to get to Chop Chop). Arguably, even here the link from the form of *B*'s utterance to his preferences is relatively direct, given that π_2 is a request. But the link between the linguistic form of a response to a question and the preference it reveals need not be quite that direct, making the preference stem entirely from the fact that it's a response to a question. For instance, consider (12), where (12b) does not provide an answer to (12a), although it does help to search for an answer (and hence is linked to (12b) with *Q-Elab*:

(12) a. A: What shall we have for dinner?
b. B: How about a nice roast?

Intuitively, *B*'s question (12b) reveals that *B* prefers to have roast for dinner. One can explain this intuition on the grounds that questions like (12)a) (and π_1 in (11)) furnish a decision problem for the respondent; and in responding to that question (whatever linguistic form that response might take), the respondent provides his interlocutor with the means for reconstructing how *B* has solved the problem. In dialogue (12) reconstructing that solution is relatively complex (and we don't go into details here). *B*'s response is linked to *A*'s question with *Q-Elab*, but in addition there is a further implicature within CL for this response, which is that *B* commits to a certain conditional preference: given that *A* and *B* have to eat dinner, *B* prefers a roast to non-roast (non-roast options are left underspecified). The antecedent of the conditional preference is in fact a presupposition of *wh*-question. Thus, defeasibly a response to a *Wh*-question that generates a decision or game problem yields a preference conditional upon the presupposition of the question. Not all *wh*-questions yield a decision problem, at least in normal circumstances. Some simply ask for factual information, such as the questions in (13):

(13) a. Who has submitted a paper to the conference?
b. Why did the mixture heat up?
c. How did you get here?

We leave for future research the delicate question of how to axiomatise the inferences to preferences from dialogue moves. Asher and Bonzon (2008) propose a set of rules based on a study of how agents express commitments to preferences with dialogue moves in scheduling dialogues from the Verbmobil corpus and in tourist advising dialogues from a French corpus collected by researchers in Grenoble. Some of the rules depend on lexical choice, while others depend on answers to questions or other pairs of dialogue moves. These rules provide us with something like a dynamic semantics between utterance moves and commitments to preferences. That is, a dialogue move, or rather a sequence of dialogue moves, ϕ define transitions from one set of commitments to preferences to another one.

One might wonder whether the (conditional) preference that's inferred from the semantic representation of (12) should then become a part of that semantic represenation, thereby importing inferences from CL back into the DSDRS in a very direct way. But doing this would affect the coherence of subsequent dialogue moves in counterintuitive ways. For instance, if (12b) really *says* that B prefers a roast, then we predict that we should be able to anaphorically refer to that preference by using, for example, an ellipsis.[18] But (12b) cannot be followed with (12c), meaning that A prefers to have a roast too:

(12) c. A: I do too.
c′. A: I would like that as well.

Such linguistic tests show that the preference inferred from (12b) is not part of the *meaning* of (12b), but it is an implicature. The CL within SDRT validates agents' inferences to such implicatures while they interpret discourse, and it is clear that people do infer implicatures such as this, since one can coherently respond to (12b) with (12c′). The roast, which is part of the meaning of (12b), is an available antecedent for the anaphor *that* (see Asher and Lascarides (2003) for formal definitions of available antecedents to anaphora). The cue phrase *as well* is a presupposition trigger and is bound to the preference that is inferred for B within CL from his utterance (12b); but presuppositions are different from so called "surface" anaphors like VP ellipsis, because they can be bound to implicatures made about the cognitive states of agents—in (14), for instance, the presupposition of *too* is bound to the inference made about A's private beliefs:

(14) a. A: There's going to be a storm.
b. B: I believe that too.

So, given this data on anaphora and presuppositions, we will for now assume that inferences about commitments to preferences are made within CL, and not transferred into the glue logic or into DSDRSs.

We've now discussed how dialogue participants constantly make new commitments—not only commitments to propositions, requests and to questions and the relations that link them together (including illocutionary effects), but also to various attitudes, such as beliefs, intentions and preferences. On the one hand, we've seen how a participant's particular preferences are affected by commitments that are revealed in a dialogue—via default axioms such `Commitments to Preferences`. On the other hand, via the logic of CP-nets we have shown how conditional preference statements affect how the conversation will proceed.

Now let's consider how agents decide what dialogue move to make next, in the light of their knowledge of their own preferences and those of their interlocutors. The logic CL is now set

[18] In SDRT, we argue that ellipsis is a matter of picking up an antecedent at the level of discourse logical form. See Asher (1993), Asher et al. (2001) and also Hardt (1992).

up to ensure that there is a dynamic interaction between information about cognitive states and dialogue moves. For example, let's examine R and C playing the putdown game in three scenarios that vary on how partial (or complete) R's and C's knowledge of each other's preferences are. First, suppose R and C have complete (and accurate) knowledge of each others preferences, which are those in Figure 1. Then by `Preferences to Intentions` R will intend $\neg p_r$ (i.e., pat C on the back), and similarly C will intend $\neg p_c$ (i.e., pat R on the back). By `Intent to Ground` both intentions will become also mutual intentions of R and C. And both have a rational expectation for how the verbal exchange will go.

Now consider the case where R's preferences are those in Figure 1 but R does not know if C is a jerk or not. On the other hand, C does know R's preferences (and in fact C assumes the CP-net from Figure 1). Then R may not yet have formed an intention with respect to the goal, since he has no information on C's preferences or intentions. But C will act as above and thus R will learn about C's intentions, assuming that C's actions are rational and follow his intentions.[19] That is, on observing C perform $\neg p_c$ R will know that C intended it and by `From Commitments to Intentions to Preferences` she will update her model of C's preferences with $\neg p_c \succ_c p_c$. This now allows her to use the CP-net so-constructed to make the move that maximises her preferences—i.e., $\neg p_r$.

Finally, consider the case where neither R nor C know anything about each other's preferences. They meet for the first time, as it were. If R is to make the first move, then unlike the prior case R cannot use C's actions to influence her move. Instead, she must reason by 'cases', using each CP-net that is compatible with her own preferences. Suppose that R's preferences are those in Figure 1, and furthermore, R knows C to be either a non-jerk (as in Figure 1) or a jerk (making C's CP-net simply $p_c \succ_c \neg p_c$). Then R can reason as follows. If C is a non-jerk, then R infers that C prefers $\neg p_c$ on condition that R performs a $\neg p_r$ (reasoning as before), making R's best move $\neg p_r$. On the other hand, if C is a jerk, then R infers that C prefers p_c regardless, making R's best move p_r. R would therefore require further strategies for deciding which of p_r vs. $\neg p_r$ to prefer. For instance, R might 'hope for the best' and perform $\neg p_r$. In any case, where all that is involved is an insult, R may consider it better to potentially receive an insult and know about C's desires than to behave like a jerk herself. An extension of the CP-net could model these additional preferences.

This simple example illustrates how an agent can use her own CL model of the discourse to reason strategically about what subsequent moves would be optimal (both for her, and her interlocutors). She can engage in this reasoning even when she lacks complete information about the other agents mental state. In such cases, she can essentially invoke a 'generate-and-test' process, where among the candidate things she could say and the candidate fully specific mental states that are compatible with her partial information, she uses CL to compute the effects, and determines from this which speech act would optimise satisfying her intentions (which in turn reflect an optimised solution for satisfying her preferences).

[19] For details on the inferences from dialogue moves to intentions, see Asher and Lascarides (2003).

5 Conclusions

This paper presents just the first steps towards a comprehensive formal semantic theory of dialogue interpretation. Giving a detailed cognitive logic which encapsulates general principles of rationality and cooperativity is beyond the scope of this paper (although see Asher and Lascarides (2003) for an early version of such a logic). We argue in this paper that the logic in question should extend dynamic logics of public announcement (Baltag et al., 1999, Kooi, in press, Gaudou et al., 2006). The extensions must (a) allow one to reason by default, and (b) allow one to link public announcements and other observable actions on the part of agents to their intentions. For example, we showed in Section 4 that a simple axiom of Sincerity can validate inferences that grounding normally leads to mutual belief, but it is not always so. In the example (1), the consequences of our Sincerity axiom cannot be inferred, and so the public announcements made there yield no conclusions about private belief. Following the work in Asher and Lascarides (2003), we intend in future work to show that such a logic, extended with further axioms of cooperativity, can be used to prove some of the axioms from Section 3 (see Asher and Lascarides (forthcoming) for details). In this way, our theory of dialogue interpretation will ultimately provide a link between agent rationality and cooperativity on the one hand and dialogue interpretation and grounding on the other.

References

J. Allen and D. Litman. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11(2):163–200, 1987.

L. Amgoud. A formal framework for handling conflicting desires. In *Proceedings of ECSQARU 2003*, 2003.

N. Asher. *Reference to Abstract Objects in Discourse.* Kluwer Academic Publishers, 1993.

N. Asher. From discourse micro-structure to macro-structure and back again: The interpretation of focus. In H. Kamp and B. Partee, editors, *Context-Dependence in the Analysis of Linguistic Meaning.* Elsevier, 2004.

N. Asher. Dynamic discourse semantics for embedded speech acts. In S. Tsohatzidis, editor, *John Searle's Philosophy of Language*, pages 211–244. Cambridge University Press, 2007.

N. Asher and D. Bonevac. Free choice permission is strong permission. *Synthese*, pages 22–43, 2005.

N. Asher, D. Hardt, and J. Busquets. Discourse parallelism, ellipsis and ambiguity. *Journal of Semantics*, 18(1), 2001.

N. Asher and A. Lascarides. Indirect speech acts. *Synthese*, 128(1–2):183–228, 2001.

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

N. Asher and A. Lascarides. A cognitive logic for dialogue interpretation. Draft copy available from the authors., forthcoming.

N. Asher and M. Morreau. Commonsense entailment. In John Mylopoulos and Raymond Reiter, editors, *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pages 387–392, Los Altos, California, 1991. Morgan Kaufmann.

Nicholas Asher and Elise Bonzon. Extraire et modéliser des préférences à partir dun dialogue. In L. Cholvy, editor, *Journes d'Intelligence Artificielle Fondamentale*, pages 8–15, Paris, 2008.

P. Balbiani, H. van Ditmarsch, A. Herzig, and T. de Lima. A tableau method for public announcement logics. In *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX)*, 2007.

A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. Technical Report SEN-R9922, Centrum voor Wiskunde en Informatica, 1999.

S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In D. Hackermann and A. Mamdani, editors, *Proceedings of the 9th International Conference on Uncertainty in Artificial Intelligence*, pages 411–419. Morgan Kaufmann, 1993.

E. Bonzon. *Modélisation des Interactions entre Agents Rationnels: les Jeux Booléens.* PhD thesis, Université Paul Sabatier, Toulouse, 2007.

C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos, and David Poole. Cp-nets: A tool for representing and reasoning with conditional *ceteris paribus* preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

C. Boutilier, R.I. Brafman, H.H. Hoos, and David Poole. Reasoning with conditional ceteris paribus preference statements. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 71–80, Stockholm, 1999.

Herbert H. Clark. *Using Language.* Cambridge University Press, Cambridge, England, 1996.

A. Copestake, D. Flickinger, I. Sag, and C. Pollard. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2–3):281–332, 2005.

Carmel Domshlak. *Modeling and Reasoning about Preferences with CP Nets.* PhD thesis, Ben Gurion University, 2002.

M. Egg, A. Koller, and J. Niehren. The constraint language for lambda structures. *Journal of Logic, Language, and Information*, 10:457–485, 2001.

B. Gaudou, A. Herzig, and D. Longin. Grounding and the expression of belief. In *Proceedings of the 10th International conference on Principles of Knowledge Represetnation and Reasoning (KR'06)*, pages 221–229, Riva de Garda, Italy, 2006.

J. Ginzburg. Resolving questions part i. *Linguistics and Philosophy*, 18(5):459–527, 1995a.

J. Ginzburg. Resolving questions part ii. *Linguistics and Philosophy*, 18(6):567–609, 1995b.

H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Synax and Semantics Volume 3: Speech Acts*, pages 41–58. Academic Press, 1975.

J. Groenendijk. Questions and answers: Semantics and logic. In *Proceedings of the 2nd CologNET-ElsET Symposium. Questions and Answers: Theoretical and Applied Perspectives*, pages 16–23, 2003.

J. Groenendijk and M. Stokhof. Semantic analysis of wh-complements. *Linguistics and Philosophy*, 5(2):175–233, 1982.

B. Grosz and C. Sidner. Plans for discourse. In J. Morgan P. R. Cohen and M. Pollack, editors, *Intentions in Communication*, pages 365–388. MIT Press, 1990.

C. Hamblin. *Imperatives.* Blackwells, 1987.

D. Hardt. Vp ellipsis and semantic identity. In Chris Barker and David Dowty, editors, *Proceedings of SALT II*, number 40 in OSU Working Papers in Linguistics, pages 145–162, OSU, Columbus, Ohio, 1992. Department of Linguistics.

B. Kooi. Expressivity and completeness for public announcement logics via reduction axioms. to appear in *Journal of Applied Non-Classical Logics*, in press.

J. Lang, L. van der Torre, and E. Weydert. Hidden uncertainty in the logical representation of desires. In *Proceedings IJCAI 2003*, pages 685–690, 2003.

A. Lascarides and N. Asher. Agreement and disputes in dialogue. In *Proceedings of the 9th SigDial Workshop on Discourse and Dialogue (SIGDIAL)*, pages 29–36, 2008.

David K. Lewis. *Convention: A Philosophical Study.* Harvard University Press, Cambridge, Massachusetts, 1969.

B. Lipman. Language and economics. In M. Basili, N. Dimitri, and I. Gilboa, editors, *Cognitive Processes and Rationality in Economics*, pages 103–145. Routledge, London, 2003.

K. Lochbaum. The use of knowledge preconditions in language processing. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1260–1266, San Francisco, 1995. Morgan Kaufmann.

J.D. MacKenzie. Question begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–233, 1979.

P. McBurney and S. Parsons. Dialogue games in multi-agent systems. *Informal Logic. Special Issue on Applications of Argumentation in Computer Science*, 22(3):257–274, 2002.

Jerry L. Morgan. Some interactions of syntax and pragmatics. In Peter Cole, editor, *Syntax and Semantics 3: Speech Acts*, pages 289–303. Academic Press, New York, 1975.

M. Poesio and D. Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3), 1997.

D.F. Sally. On sympathy and games. *Journal of Economic Behaviour and Organization*, 44 (1):1–30, 2001.

J. Searle. *Speech Acts.* Cambridge University Press, 1969.

S. Shieber. *An Introduction to Unification-based Approaches to Grammar*. CSLI Publications, 1986.

D. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Computer Science Department, University of Rochester, 1994.

D. Traum and J. Allen. Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94)*, pages 1–8, Las Cruces, New Mexico, 1994.

J. van Benthem, S. van Otterloo, and O. Roy. Preference logic, conditionals, and solution concepts in games. In *Festschrift for Krister Segerberg*. University of Uppsala, 2005. also available on the ILLC prepublication repository: PP-2005-28.

D.N. Walton and E.C.W. Krabbe. *Commitment in Dialogue*. SUNY Press, 1995.

F. Wolf and E. Gibson. Representing discourse coherence: A corpus-based analysis. *Computational Linguistics*, 31:249–288, 2005.

M. Wooldridge. *Reasoning about Rational Agents*. MIT Press, 2000.