Cognitive States, Discourse Structure and the Content of Dialogue

Alex Lascarides
Division of Informatics,
University of Edinburgh
alex@cogsci.ed.ac.uk

Nicholas Asher
Department of Philosophy,
University of Texas at Austin
nasher@bertie.la.utexas.edu

1 Introduction

In this paper, we examine anaphora in dialogue. Anaphora resolution is extremely complex, so for the sake of clarity we will focus on just two puzzles, illustrating them with a 'minimal pair' of examples, in order to bring out the major features of the framework. The first puzzle is that constraints on anaphora are not only sensitive to contextual information, but also to the way it's presented: compare (1ab′c), where the anaphoric expression *his name* sounds odd, to (1abc) and (1ab′c′), which are acceptable.

(1) a. A: How can I get to sixth street?
  b. B: There's someone Downtown that you could ask.
  b′ B: You can ask someone Downtown.
  c A: What's his name?
  c′ A: Oh yeah? And what's his name?

Second, communicative intentions affect anaphora. For example, *2pm* in (2e) can't be 2pm on Saturday, even though Saturday is referred to in the previous utterance, because this is incompatible with the underlying intention: given (2d), why would A ask if he can meet at 2pm on Saturday? Furthermore, something more than intentions are needed, since *2pm* can't be on Sunday either.

(2) a. A: How about meeting next weekend?
  b. B: That sounds good.
  c. How about Saturday afternoon?
  d. A: I'm afraid I'm busy then.
  e. ??How about 2pm?

We'll argue for three criteria for dialogue interpretation. First, one needs rhetorical relations such as *Elaboration* and *Contrast*. Second, the framework must incorporate several levels of representation and reasoning. In particular, discourse content (both semantic and pragmatic) must be represented at a different level from cognitive states; and the logic for computing discourse content must have only restricted access to that content and to cognitive states. And third, one must account for complex interactions between the discourse and cognitive levels.

Most current theories of dialogue analysis fail to fully meet these criteria, and their accounts of (1) and (2) are problematic. We'll show here how framework of discourse semantics known as SDRT meets our requirements. Asher (1993) and Lascarides and Asher(1993) show how SDRT augments DRT with rhetorical relations and is modular in the way required, and Asher (1999) has added a separate module for limited reasoning about cognitive states and Gricean implicatures. In this paper, we defend the view that separate but interacting theories of discourse structure and cognitive states are essential to the analysis of dialogue. We illustrate our approach by first using principles of rationality to derive principles for computing discourse content—in particular the rhetorical relations that hold, and then exploiting these cognitively-motivated but discourse-level axioms to account for the constraints on anaphora observed in (1) and (2).

2 Dialogue and Mental States

Many researchers have emphasised cognitive modelling in analysing dialogue (e.g., Grosz and Sidner 1986, 1990, Litman and Allen 1990, Hobbs *et al.* 1993, Sperber and Wilson 1986). In these theories, the compositional semantics of an utterance is incorporated as a new belief, and this leads to further new beliefs or changes to old ones. These updates are viewed as pragmatic effects. So interpretation is belief revision, and cognitive states and discourse content are represented in the same module.

Moore and Paris (1993) and Asher and Lascarides (1998a) argue against this approach. First, the step from utterances to attitudes can only be made by default (Sparck Jones, 1989); interpreters have no direct access to other speakers' cognitive states. Further, as the *full* semantics of the utterance goes into a belief, the language for characterizing cognitive states must be at least first order. A serious problem, then, is that this view has a tractable model neither for belief revision nor even for belief update; this is because both will involve tasks equivalent to consistency checks over extensions of first order default theories. On this view, dialogue interpretation goes beyond what's recursively enumerable.

Equating discourse interpretation with cognitive effects also leads to empirical problems: one can't explain why small surface changes affect anaphora, e.g., as in (1). One can't distinguish the intentions of the utterances (1abc) and (1ab′c) on independent grounds, because one can't with confidence distinguish the domain level plan that underlies (1abc) vs (1ab′c), or the underlying discourse level plan of speech acts.

To see this, let's consider the domain level plans first. One might try to distinguish them by assuming that the domain plan underlying (1ab′c) involves going Downtown and choosing someone *randomly*, whereas the choice isn't random in the plan underlying (1abc). But this places a huge burden on the axiomatisation of such fine-grained plan operators. One would have to assume plan recognition axioms that translate very small differences in surface form into dramatic effects on the plan operators that are invoked in the domain plan. We believe such a model would be highly impractical. At any rate, even if it were feasible, the differences in the anaphoric possibilities that are generated by (1b) vs (1b′) has a simpler, *monotonic* explanation, which bypasses domain plans altogether. We discuss this in section 3.

Now consider the discourse level plan of speech acts. Traditionally, these are computed via inferences about the speaker's beliefs, on the basis of his utterance (e.g., Perrault 1990, Grosz and Sidner 1990). But this means that one can't claim with confidence that the speech acts underlying (1b) vs. (1b′) are distinct, because one can't claim with confidence that B's beliefs which lead him to utter (1b) vs (1b′) are distinct. For example, it's possible that in each case, his belief that someone Downtown knows the way is *de dicto* (i.e., he doesn't have a particular person in mind). The fact that the belief could be *de dicto*, even if he utters (1b′), also undermines the previous conjecture, that the domain level plan for (1ab′c) may involve asking a particular person Downtown, as opposed to a random person.

Overall, an independently motivated axiomatisation of plans and intentions is too coarse-grained to provide the rich discourse structures that are necessary for constraining anaphora. So they overgenerate possible interpretations and predict discourse coherence where there isn't any.

3 Dynamic Semantics

How can these shortcomings be avoided? Dynamic semantics can help. In dynamic semantics, computing discourse content isn't equated with cognitive effects. Rather, one defines *discourse update*, i.e., the way the current utterance updates a representation of the meaning of the discourse context (e.g. DPL (Groenendijk and Stokhof, 1991), DRT (Kamp and Reyle, 1993) and SDRT (Asher 1993, Lascarides and Asher 1993)). Unlike Hobbs *et al.* (1993), discourse content is represented at an 'intermediate' level between compositional semantics and the attitudes: it expresses more content than compositional semantics since discourse update may resolve underspecified semantic conditions arising from the grammar, but it does not in general record cognitive effects. Indeed, cognitive modelling isn't traditionally addressed within dynamic semantics at all.

Dynamic semantics provides the basis for analysing (1). The difference in scope between the quantifier *someone* and the modality *can* in (1b) vs (1b′) yields, monotonically, the differences between (1abc) and (1ab′c) (Kamp and Reyle, 1993).

Moreover, one difference between (1ab′c) vs. (1ab′c′) is that there's a *Contrast* between (1c′) and (1b′) which is lacking in (1b′c) (note that intuitively, inserting *but* in (1c′) doesn't affect its semantic contribution to the dialogue). This rhetorical difference leads to different constraints on anaphora in SDRT (Asher, 1993), which predict that *his* refers successfully in (1c′). To see this, let's examine SDRT's analysis of *Contrast*. Roughly put, two propositions that are connected with *Contrast* must have a partial isomorphism between their hierarchical DRS-structures,[1] and at least one pair of nodes in this mapping must have opposite polarities or 'contrasting themes'. The details for computing contrasting themes are given in Asher (1993). But this isn't our main concern here. Rather, we focus on how the 'isomorphic' mapping affects anaphoric possibilities. Again, roughly put, there's often a choice of partial isomorphisms, and the highest ranked mapping is the one that's closest to a total isomorphism. Furthermore, a discourse referent that is introduced in a sub-DRS D_1 of the first proposition in the *Contrast* relation can act as an antecedent to an anaphor in the sub-DRS D_2 of the second proposition that D_1 is mapped to. This means *Contrasts* allows elements in embedded DRSs to be antecedents to certain anaphora, since D_1 may be embedded, and this is contrary to the general constraint of accessibility in DRT (and indeed, the fact that embedded DRSs are generally inaccessible explains why (1ab′c) is odd, as the discourse referent introduced by *someone* is in a DRS that's embedded within the scope of the modality *can*).

So, how does *Contrast* effect the anaphora in (1ab′c′)? Well, the embedded DRS D_1, that has a discourse referent introduced by *someone*, is mapped to some part of the DRS structure of (1c′). Now, (1c′) contains a λ-abstract introduced by *what*, and therefore the discourse referent introduced by *his* is in a DRS D_2 that's embedded within the scope of this λ operator. Indeed, *Contrast* here induces a total isomorphism, with D_1 mapped to D_2, since these are both embedded one step down the top level DRS. So the *Contrast* makes *someone* accessible to *his*. As we mentioned, it's inaccessible in (1ab′c), because of the different rhetorical relation that holds between (1b′c).

Dialogue (1) demonstrates that the quantifiers, modalities and rhetorical relations constrain anaphora. Dynamic semantics can define such constraints. But we must extend it to account for dialogue. Grice (1975) and the above AI work demonstrate that the interplay between what people say and the goals that motivate them is vital. Dynamic semantics must model this, and dialogue (2) demonstrates why. On the one hand, if (2e)'s underlying goal—to meet next weekend but not on Saturday afternoon—plays no role in constraining (2e)'s content, then *2pm* could be interpreted as 2pm on Saturday. On the other hand, if *only* the goals matter, then *2pm* would refer to 2pm on Sunday. The fact that *2pm* fails to refer demonstrates that both linguistic and cognitive constraints are necessary.

We draw several conclusions. First, since rhetorical relations constrain anaphora, our theory should include them and their effects on content. Second, if interpreting dialogue is to be reliable and computable, then one must represent dialogue content at a separate level from cognitive states. Moreover, given that computing rhetorical connections involves reasoning with partial information (Lascarides and Asher 1993, Hobbs *et al.* 1993), the logic that models this can't have full access to the logic of dialogue content or the logic of cognitive states, as this would make discourse interpretation undecidable. And finally, an adequate account of anaphora requires a precise model of the interaction between the discourse and cognitive levels.

Most frameworks of dialogue analysis fail to meet our three criteria. The cognitive approaches mentioned in §2 aren't modular in the way required. Grosz and Sidner (1986) analyse discourse at several levels, but they don't distinguish *Contrast*s from other coordinating moves and they don't distinguish the constraints on anaphora imposed by (1b) vs. (1b′). Standard dynamic semantic theories fail in other respects. They ignore rhetorical relations and the influence of goals and intentions on dialogue content, and therefore fail to account for (2).

Our main point is that a satisfactory interpretation of dialogue requires both dynamic semantics and cognitive modelling in *separate*, but communicating modules. To this end, we have extended SDRT with a module that links what people say to what prompted them to say it (Asher 1999).[2] we'll use general principles of rationality in this theory to compute axioms for inferring discourse structure. We'll then demon-

[1] This mapping is independently motivated on the grounds of VP ellipsis, gapping, and other ellided constructions.

[2] We introduced links between cognitive states and discourse structure in Asher and Lascarides 1994, and Asher and Lascarides 1998a but the modules were not distinguished.

strate that the resulting theory models the interplay between linguistic and cognitive information in the interpretation of (2).

4 Cognitive Modelling in SDRT

SDRT has two parts. First, there's a formal language for representing discourse content (Asher, 1993). This is an extension of DRT: a discourse is represented as an SDRS, which is a recursive structure of labelled DRSs, with rhetorical relations between the labels. Second, there's a theory known as DICE, which is used to compute rhetorical relations in the SDRS of a discourse from the compositional semantics of its clauses (Lascarides and Asher, 1993). DICE consists of a modal propositional language augmented with $>$ ($A > B$ means *If A then normally B*). Rules in DICE are of the form in (3), where $\langle\tau, \alpha, \beta\rangle$ means β is to be attached to α with a rhetorical relation (α and β label SDRSs), where α is part of the discourse context τ, *some stuff* is information about α and β, and R is a rhetorical relation:

(3) $(\langle\tau, \alpha, \beta\rangle \wedge \text{ some stuff}) > R(\alpha, \beta)$

DICE doesn't have full access to the language of SDRSs. If it did, then discourse interpretation wouldn't be computable. Rather, the conditions in an SDRS K_α get translated in DICE into predicates of the propositional variable α, where α labels K_α. So DICE has access to the *form* of information content, but not to all its entailments.

We turn now to model interactions between dialogue content and cognitive states. We'll first correlate attitudes with utterances, and then derive DICE axioms from basic principles of rationality. Dialogue utterances are often associated with beliefs but also what we call *speech act related goals* (SARGs): SARGs reflect the intuition that people say things for a particular purpose (Asher 1999). Two rhetorical relations where beliefs and goals play a central role are: *IQAP* (Indirect Question Answer Pair) and *Q-Elab*(Question Elaboration). $IQAP(\alpha, \beta)$ holds if α is a question, and β together with other information the questioner has access to allows him to derive a correct answer to α (Asher and Lascarides 1998a). Note that this subsumes direct answers (so $IQAP(2a, 2b)$ and $QAP(2a, 2b)$ hold, since (2b) is a direct answer to (2a)). Second, $Q\text{-}Elab(\alpha, \beta)$ holds if β is a question whose answers all specify part of a plan to bring about an SARG of α (e.g., $Q\text{-}Elab(2b, 2c)$ holds, since the SARG of (2b) is to find a time to meet next weekend, and any answers to (2c) will reduce the search for those times) (Asher 1999). We'll shortly derive axioms for inferring *IQAP* and *Q-Elab*.

Cognitive modelling has a problem analogous to the one for computing discourse content: just as the logic for computing discourse content must be shallower than the logic of discourse content itself, the logic for cognitive modelling must be shallower than the logic of cognitive states. This is for analogous reasons: We want the cognitive modelling that's required to perform dialogue interpretation to be computable. And as we mentioned before, consistency checks are needed, because default reasoning is necessary as a result of the lack of direct access we have to other people's cognitive states. So the underlying monotonic logic must have a validity problem that's decidable at worst, or reasoning becomes uncomputable. And this means that cognitive states must be represented in a language that is strictly less expressive than first order, and it must have a more impoverished notion of validity than first order logic.

We achieve this as follows: like DICE, the language of cognitive modelling will be a modal propositional one augmented with $>$. This is much less expressive than the language which represents full cognitive states; according to Asher (1986) and Kamp (1990) this should be at least as expressive as the language of information content, and so in this case it would be SDRSs. However, the less expressive language is 'linked' to cognitive states and to discourse content, because the propositional variables in this language get indexed to the labels for SDRSs that appear at these discourse and cognitive levels. So, let π label the SDRS K_π, and let the proposition variable p_π be indexed to π. Then an interpretation of the language for cognitive modelling is *admissible* only if the worlds assigned to p_π are the ones that make K_π true. Modal operators $\mathcal{B}$ (believes) and $\mathcal{I}$ (intends) will then operate over p_π, and because of admissibility, $\mathcal{B}_A p_\pi$ corresponds to A believing the content represented in the SDRS K_π. We'll assume that whenever an agent intends something he does not already believe that it is true; in symbols, $\mathcal{I}_A\phi \rightarrow \neg\mathcal{B}_A\phi$.

We'll now provide some axioms of rationality and cooperativity. First, consider Cooperativity: an agent B is cooperative with agent A if he adopts A's goals. If B doesn't do this for whatever reason, then being cooperative means he will normally indicate this to A (Asher 1999). These principles are represented as a pair of axioms:

- `Cooperativity`
  (a) $\mathcal{I}_A\phi > \mathcal{I}_B\phi$
  (b) $(\mathcal{I}_A\phi \wedge \neg\mathcal{I}_B\phi) > \mathcal{I}_B\mathcal{B}_A\neg\mathcal{I}_B\phi$

Part (a) is clearly default: B may have conflicting goals that stop him from adopting A's goals, for example. If he doesn't adopt A's goals, then part (b) means he should make it evident to A. But this is default too: B may not, for various reasons, want A to know. For now, *pace* Sperber and Wilson (1986) who claim that cooperativity is epiphenomenal, we assume `Cooperativity` is a basic principle.

`Cooperativity` models goal transfer, but it doesn't yield an agent's goals from the evidence of what he said. In general, this is hard to do! But certain types of speech acts have goals or SARGs which one can compute in a relatively straightforward manner. It's these goals that *prima facie* we have an obligation to adopt via `Cooperativity`. For instance, questions have SARGs which will call *question related goals* (QRGs): the QRG of a question is to know an answer (A :?α means that A said ?α, and ?α means α is a question).

- `Question Related Goals (QRG):`
  $(A :?\alpha \wedge QAP(\alpha,\beta)) > (\neg\mathcal{B}_A\beta \wedge \mathcal{I}_A\mathcal{B}_A\beta)$

So asking a question can lead to an inference about at least one of its SARGs.

We've so far linked goals and utterances. Grice's maxim of Quality links beliefs and utterances. It stipulates that one should say only that which one thinks is true, and avoid saying things for which one lacks adequate evidence. `Sincerity` represents the first part of this maxim (see Perrault, 1990):

- `Sincerity`: $A{:}\phi > \mathcal{B}_A\phi$

Like `QRG`, `Sincerity` links what one says to one's attitudes. Grice's second part to Quality provides a rule for belief transfer. In essence, it stipulates that interpreters assume by default that an agent who volunteers some information is *competent* with respect to it, i.e., what he says is true:

- `Competence`: $\mathcal{B}_A\phi > \phi$

Lewis (1969) argues persuasively that such a default must form the basis of cognitive modelling if one is to explain why linguistic conventions occur. It works as a principle of belief transfer in the following way. Suppose B hears A say that ϕ. Then by `Sincerity` B infers $\mathcal{B}_A\phi$ and by `Competence` he infers ϕ. So B must believe ϕ or succumb to Moore's (1912) paradox *ϕ but I don't believe ϕ.*

Finally, intentions, belief and action all interact. In Asher and Lascarides (1998a), we argued that Aristotle's Practical Syllogism is useful for modelling this: normally, people intend to do things that normally will eventually fulfill their goals. We encode this principle below: if (a) B intends ψ and believes $\neg\psi$ and (b) B believes that he can nonmonotonically infer ψ if ϕ is true, and ϕ is B's choice for achieving ψ, then (c) B intends ϕ:

- `Practical Syllogism (PS):`
  (a) $(\mathcal{I}_B\psi\wedge$
  (b) $\mathcal{B}_B(\phi > eventually(\psi)) \wedge choice_B(\phi,\psi))$
  (c) $> \mathcal{I}_B\phi$

We will often use `PS` and an additional axiom that actions are intentional to infer an agent's attitudes from his behaviour. For example, suppose we observe a speech act performed by A. Then we'll infer, given absence of evidence to the contrary, that A intended this action. According to `PS`, A's intention resulted from a particular mix of goals and beliefs: by fixing either one we'll abduce the other. Abduction boils down to deduction in this framework: we rearrange the antecedent and consequent of `PS` to form new axioms, but for reasons of space we don't spell out all permutations of (a), (b) and (c) here. `PS` as written above is used to infer expectations about the speech act B will perform, given some prior knowledge of his goals and beliefs.

5 Computing Discourse Structure

Recall that $IQAP(A{:}?\alpha,\beta)$ holds only if A can use β to compute a direct answer to α:

- `Axiom on IQAP`:
  $IQAP(A{:}\alpha, \beta) \rightarrow \exists\gamma(QAP(\alpha, \gamma) \wedge \mathcal{B}_A(\beta > \gamma))$

This is a necessary condition for *IQAP*, but what are the sufficient conditions? Cognitive modelling supplies an answer.

Suppose A asks $?\alpha$. Then B uses `QRG` to infer that A wants to know an answer. Now, since A and B mutually believe the axioms in §4, A believes that B will deduce this. Moreover, by `Cooperativity`, B adopts A's goal and A infers B does this. So A and B both believe clause (a) of `PS` is verified, where ϕ is substituted with "A believes an answer to α". Now, A observes B utter β. By the axiom that actions are rational, A assumes this was intentional and it was B's choice for achieving ϕ. So A believes (and B believes) that clause (c) of `PS` is verified, where ψ is substituted with "B utters β". So A can abduce clause (b): B thinks β by default leads to the goal. That is, β provides A with enough information to infer an answer to α. So A infers that B believes β is attached to α with *IQAP*. Given `Competence`, A can also infer $IQAP(\alpha, \beta)$, as long as the constraint is met—namely, that A can indeed derive an answer to α from β. Thus the following is valid:

- `PS`, `Cooperativity`, `Competence`, $\langle \tau, ?\alpha, \beta \rangle$
  $\models\!\approx IQAP(\alpha, \beta)$

The consequence relation $\models\!\approx$ has a weak deduction property such that the above makes `IQAP` valid:

- `IQAP`: $\langle \tau, ?\alpha, \beta \rangle > IQAP(\alpha, \beta)$

Note that the only condition for this rule applying is that the surface speech act of α is a question.

Now consider *Q-Elab*. Informally, $\textit{Q-Elab}(A{:}\alpha, \beta)$ holds only if (a) β is a question, and (b) for any γ such that $IQAP(\beta, \gamma)$ holds (i.e., γ's an indirect answer to β), γ provides information from which A can specify a plan p for achieving the SARG of α, and this information wasn't derivable from what A and B mutually knew on the basis of the discourse so far. Given this constraint on *Q-Elab*, we can use the cognitive principles from §4 to derive an axiom for inferring *Q-Elab* at the discourse level:

- `Q-Elab`: $\langle \tau, \alpha, ?\beta \rangle > \textit{Q-Elab}(\alpha, \beta)$

This axiom stipulates that by default, questions attach to their antecedents with *Q-Elab*. That is, the default role of a question is to help achieve a goal that prompted some previous utterance. Clarification questions (e.g., (4ab)) are a particular kind of *Q-Elab*, since one often can't achieve the SARG of α until one is clear on the content of α (in the case of (4a) the SARG is that A get to a particular location):

(4) a. A: I need to be in Cambridge tomorrow.
b. B: Cambridge UK or Cambridge USA?
b′. B: Is your hair grey?

Q-Elab also connects (2c) to (2b), because any answer to (2c) will further specify a plan to find a time to meet: an answer will either eliminate Saturday afternoon from the search (as (2d) does) or it will restrict the search to it. However, (4ab′) is predicted to be incoherent, because although the antecedent to `Q-Elab` is verified, its consequent can't be inferred because there are answers to (4b′) which don't specify plans to achieve the SARG of (4a) and therefore $\textit{Q-Elab}(4a, 4b')$ can't be true. Since the DICE axioms give no other candidate, a rhetorical connection between (4a) and (4b′) can't be computed.

`Q-Elab` is derived from the cognitive principles as follows. Suppose A utters α with SARG ϕ. Then by `Cooperativity`, B adopts ϕ. Hence clause (a) of `PS` is satisfied. A observes B ask a question β, and by the rationality of action, he assumes this was intentional. So A knows that clause (c) of `PS` is satisfied too. So as before, A (and we) can infer clause (b): B believes that his asking β will normally lead to ϕ. Now, β has an SARG, which by `QRG` is that B know an answer. How does this goal fit in with B's other goal ϕ? B must construct a plan for achieving ϕ; similarly for the QRG. And by the default link given in clause (b) of `PS`, the plan for achieving the QRG is a subplan of that for achieving ϕ. Hence B thinks that the answers to β will specify a plan that will lead to the fulfilment of A's SARG ϕ. By `Competence`, A (and we) conclude by default that B's question does have these properties. And this is part of what the constraint on *Q-Elab* specifies. The constraint also stipulates that the relevant information isn't available to A and B already. But if it were available, then B's question would be moot and `QRG` wouldn't fire. So the default `QRG` implies that answers to questions are informative, and this yields the second part of the *Q-Elab* constraint. Hence the cognitive axioms yield `Q-Elab` as an axiom of DICE.

Having illustrated how cognitive principles lead to discourse ones, let's return to (2), and in particular the content of *2pm*. Let's assume A and B have a goal $\mathcal{G}$, which is to gain *de re* knowledge of a time at which they can meet. This yields at least two SARGs for (2a). First, there's the goal $\mathcal{G}$. (2a)'s compositional semantics are consistent with this SARG, because any answers to (2a) will either add *de re* knowledge that they can meet next weekend, or it will add knowledge that they can't. Either way, the possible times to meet are reduced. Second, by `QRG`, an SARG of (2a) is that A know an answer.

A and B must now attach the representation of (2b) to the representation of (2a) with a rhetorical relation. Assuming the compositional semantics of *How about* questions taken from Sag and Ginzburg (in press), answers are adjectives. Therefore, it's consistent to attach (2b) to (2a) with *IQAP*. So this is inferred via `IQAP`. Furthermore, the content of (2b) helps specify a plan to achieve the other SARG $\mathcal{G}$ of α, since it provides new *de re* knowledge that A and B can meet next weekend. A rhetorical relation *Plan-Elab*—which is analogous to *Q-Elab* but where the second utterance is an assertion instead of a question—is inferred in SDRT under these circumstances:[3] If *Plan-Elab*(α, β) holds, then β specifies a plan q to achieve the SARG of α.

What's the SARG of (2b)? Well, it's plausible that normally, a plan q which leads to the SARG of α and which was derived from a cooperative response β is the SARG of β. We can formalise this as an axiom `SARGs of Cooperative Responses`, but we forego this here. At any rate, this means that the SARG of (2b) is to accomplish a plan q, to find *de re* knowledge of a time to meet *next weekend*.

The question (2c) can attach to either (2a) or (2b) since *IQAP* and *Plan-Elab* are subordinating relations. In both cases, the default `Q-Elab` would apply. However, `Cooperativity` leads to a default that new information attaches to the lowest site in the context: if it attaches higher up, then the SARG of the lower attachment site is never addressed, but `Cooperativity` commits the participants to addressing all goals by default. Therefore, (2c) attaches to (2b) (and so the elided question resolves to *How about meeting on Saturday afternoon next weekend?*).

The discourse structure so far allows (2d) to attach to (2c), (2b) or (2a). If (2d) attaches to (2c), then *then* resolves to *Saturday afternoon*. If it attaches to (2b) or (2a), it resolves to *next weekend*. There are two principles in SDRT for preferring the attachment to (2c). First, there's the default for low attachment we just mentioned. And second, attaching (2d) to (2b/a) produces a less coherent discourse, on the grounds that A's goal for asking the question (2a) would be moot. So a general default that discourse interpretations which maximise coherence are preferred (Asher and Lascarides 1998b) means (2d) attaches to (2c). Since (2d) attaches to (2c) with *IQAP*, *then* resolves to Saturday afternoon. Moreover, the plan q which this answer to (2c) gives rise to is to find a time to meet which is next weekend, but not on Saturday afternoon. By `SARGs of Cooperative Responses`, q is the SARG for (2d).

We now come to (2e). This can attach to (2d) or (2a). There are two reasons why it doesn't attach to (2a): first, the default for low attachment blocks it; and second, one can't resolve the anaphoric expression *2pm* with this attachment, because *next weekend* provides two alternative bridging relations to *2pm*, violating the constraint on uniqueness given by Clark (1977). So (2e) must attach to (2d). This means that the only available antecedent that *2pm* can resolve to is Saturday afternoon. The antecedent to `Q-Elab` is verified. However, its consequent can't be inferred because the constraints on *Q-Elab* are violated. This is because given that (2e) means *How about 2pm on Saturday?*, answers to this question fail to specify a plan to know of a meeting time that's *not* on Saturday afternoon. But given the SARG of (2d), (2e) must provide such a plan if it attaches with *Q-Elab*. And so (2e) can't attach to (2d) with *Q-Elab*. But no other default rules are verified. So no relation can be inferred for attaching (2e), resulting in incoherence.

6 Conclusion

Principles of rationality and cooperativity can account within SDRT for some of the links between cog-

[3] In fact, one can use the axioms in §4 to derive an axiom for inferring *Plan-Elab*, but we gloss over this here.

nitive and conventional information that are needed in an adequate account of anaphora in dialogue. We used the cognitive principles to derive axioms for inferring rhetorical relations at the discourse level and to give these relations a semantics that constrains the resolution of anaphora.

The cognitive axioms ensure that *IQAP* is the default way of attaching an assertion to a question, and questions are attached by default with *Q-Elab*. These defaults follow from `Cooperativity`: participants engage in a cooperative effort to solve the SARGs of utterances. These defaults also represent expectations about responses.

But in some situations an agent doesn't give the expected response. For instance, `Cooperativity` may not apply, perhaps because B believes that A's SARG is already achieved, or he believes that there isn't any way of achieving the SARG. These exceptions arise from constraints on intentions adopted by Cohen and Levesque (1990), Koons and Asher (1994) and others. If B doesn't adopt A's SARG, then by part (b) of `Cooperativity` he is by default obliged to indicate this. So in these circumstances, B's question or assertion may have a different rhetorical role: instead of providing a *Q-Elab* or *IQAP*, it may be a contribution which is designed to imply that the SARG isn't adopted. In future work, we plan to explore such exceptions, and how one can use clues in the dialogue to infer them. We also plan to analyse indirect speech acts, and explain why utterances with similar meaning can have distinct SARGs; e.g., *You can say that again* vs. *You can repeat that.*

7 References

Asher, N. [1986] Belief in Discourse Representation Theory, *Journal of Philosophical Logic*, **15**, pp.127–189.

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

Asher, N. [1999] Discourse Structure and the Logic of Conversation. *Current Research in the Semantics Pragmatics Interface.*

Asher, N. and A. Lascarides [1994] Intentions and Information in Discourse, in *Proceedings of ACL94*, pp.35–41.

Asher, N. and A. Lascarides [1998a] Questions in Dialogue, *Linguistics and Philosophy*, **23.3**, pp237–309.

Asher, N. and A. Lascarides [1998b] The Semantics and Pragmatics of Presupposition, *Journal of Semantics*, **15**, pp239–299, OUP.

Clark, H. [1977] Bridging, in P. N. Johnson-Laird, P. C. Wason (eds.) *Thinking: Readings in Cognitive Science*, pp. 411–420, CUP.

Cohen, P. and Levesque, H. [1990] Rational Interaction as the Basis for Communication, in in P. R. Cohen, J. Morgan and M. Pollack (eds.) *Intentions in Communication*, pp.221–255, MIT Press.

Grice, H. P [1975] Logic and Conversation, in Cole, Peter and Morgan, Jerry L. (eds.) *Syntax and Semantics Volume 3: Speech Acts*, pp41-58. Academic Press.

Groenendijk, J. and M. Stokhof [1991] Dynamic Predicate Logic, *Linguistics and Philosophy*, **14**, pp39–100.

Grosz, B. and Sidner, C. [1986] Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, **12**, 175–204.

Grosz, B. and Sidner, C. [1990] Plans for Discourse, in P. R. Cohen, J. Morgan and M. Pollack (eds.) *Intentions in Communication*, pp.417–444, MIT Press.

Hobbs, J. R., M. Stickel, D. Appelt and P. Martin [1993], Interpretation as Abduction, *Artificial Intelligence*, **63 (1–2)**, pp.69–142.

Kamp, H. [1990] Propositional Attitudes, in C. A. Anderson, J. Owens (eds.), *The Role of Content in Logic, Language, and Mind*, pp.27–90, CSLI.

Kamp, H. and Reyle, U. [1993] *From Discourse to Logic*, Kluwer Academic Publishers.

Koons, R. C. and N. Asher [1994] The Revision of Beliefs in a Changing World, *Proceedings of TARK5*, pp. 321-340.

Lascarides, A. and N. Asher [1993] Temporal Interpretation, Discourse Relations and Commonsense Entailment, in *Linguistics and Philosophy*, **16**, pp.437–493.

Lewis, D. [1969] *Convention: A Philosophical Study*, Blackwells.

Litman, D. and J. Allen [1990] Discourse Processing and Commonsense Plans, in P. R. Cohen, J. Morgan and M. Pollack (eds.) *Intentions in Communication*, pp.365–388, MIT Press.

Moore, G.E. [1912] *Ethics*, London.

Moore, J. D. and C. Paris [1993] Planning Text for Advisory Dialogues, *Computational Linguistics*, **19.4**, pp.651–694.

Perrault, R. [1990] An Application of Default Logic to Speech Act Theory, in P. R. Cohen, J. Morgan and M. Pollack (eds.) *Intentions in Communication*, pp.161-186, MIT Press.

Sag, I. A., and Ginzburg, J. [in press] *Interrogatives*, CSLI publications.

Sparck Jones, K. [1989] Realism about User Modelling, in Kobsa, A. and W. Wahlster (eds.) *User Models in Dialog Systems*, pp341–363, Springer-Verlag.

Sperber, D. and Wilson, D. [1986] *Relevance*, Blackwells.