Questions in Dialogue

Nicholas Asher
Department of Philosophy,
University of Texas at Austin,
Austin, Texas 78712,
USA
nasher@bertie.la.utexas.edu

Alex Lascarides
Centre for Cognitive Science and HCRC,
University of Edinburgh,
2, Buccleuch Place,
Edinburgh, EH8 9LW,
Scotland, UK
alex@cogsci.ed.ac.uk

April 25, 1996

Abstract

In this paper we explore how compositional semantics, discourse structure, and the cognitive states of participants all contribute to pragmatic constraints on answers to questions in dialogue. We synthesise formal semantic theories on questions and answers with techniques for discourse interpretation familiar from computational linguistics, and show how this provides richer constraints on responses in dialogue than either component can achieve alone.

1 Introduction

Understanding dialogue involves a complex interaction between the semantics of the utterances, the discourse context of which they are part and the cognitive states of the participants. In an effort to begin to spell out these connections precisely, we focus here on the role of questions and answers in dialogue—more specifically, *how*-questions. Like many researchers (e.g., Hobbs 1985, Polanyi 1985), we represent a discourse context as a structured object consisting of propositions related by discourse relations. *How*-questions yield an intuitive connection between a particular sort of discourse structure and the intentions of the participants, since answers to a particular type of *how*-question are plans.

We will model this interaction by drawing on recent plan recognition work in the analysis of dialogue by Grosz and Sidner (1990), Litman and Allen (1990), Pollack (1986, 1990), and others. They exploit domain knowledge about how to achieve actions in building the structure of the dialogue. This is an essential component of the task. But these theories haven't linked their accounts to the role of compositional semantics in building a discourse structure. We think this poses several problems: it forces them to rely on domain knowledge for information about intentional structure in certain cases where this isn't plausible; and it fails to fully analyse the pragmatic/semantic interaction—in particular, how discourse relations can affect information about *both* the participant's intentional states *and* the semantic content of the discourse—thus missing important generalizations about dialogue.

The exact nature of discourse structure is determined in large part by what information one wants to extract from it during dialogue processing. This in turn determines which knowledge representations are relevant for building it. We will argue that if discourse structure is to constrain the possible antecedents to anaphora, enable us to specify the semantic content of a sentence in the light of the way it connects to the previous discourse, and inform us about how the dialogue participants' cognitive states change as the dialogue proceeds, then domain knowledge about plans is an essential source of information for building discourse structure, but it's insufficient on its own.

We propose to use SDRT to analyze *how*-questions in cooperative dialogues. SDRT is a theory of discourse semantics designed to explore systematically the interface between semantics, pragmatics and discourse structure. To date it has been used to model the effects of discourse structure on pronominal and temporal anaphora in monologue (e.g., Asher 1993, Lascarides and Asher 1993). Asher and Lascarides (1994) examine how discourse structure, participant's intentions and semantic content interact in monologue, but SDRT still lacks an account of how discourse structure interacts with plans. We will combine the techniques in SDRT together with those in plan recognition in an extension of SDRT, so as to achieve an analysis of dialogue with the advantages of both. To motivate our analysis, we first consider some simple examples.

2 Motivating Examples

Very simple examples of questions and answers in dialogue reveal the complex interaction between discourse structure, participants' mental states and semantic content.

(1) a. A: How can I get to the treasure?
b. B: It's at the secret valley.
c. A: But I don't know how to get there.

An adequate analysis of (1) must provide answers to at least the following questions: How does A infer from (1b) that she should go to the secret valley? On what basis is (1b′) an odd response to (1a)? And why is the use of *but* coherent in (1a,b,c), but not in (1a,b,c′)?

(1) b′. B: ?Mary's hair is black.
c′. A: ?But I know how to get there.

On an intuitive level, answers to these questions involve reasoning about compositional semantics, discourse structure and cognitive states. For instance, the compositional semantics and discourse structure must represent the fact that (1a) is a question, (1b) is a response to it, and *it* in (1b) refers to the treasure. Moreover, a combination of the compositional semantics of (1b) (from which one learns that the treasure is at the secret valley) and A's cognitive state (more specifically, his knowledge about action and causation), should be sufficient to infer at this stage that A would achieve getting to the treasure by executing the action of going to the secret valley. This must contrast with the semantic contribution of (1b′), where intuitively A's cognitive state doesn't combine with the information in (1b′) to help A infer what he has to do to get to the treasure. On the other hand, the semantics of (1c) and the way it attaches to (1a,b) in the discourse structure should allow B to infer that A still can't execute a plan of getting to the treasure, even though he knows it's at the secret valley. Typically, B must then revise his model of A's cognitive state (i.e., revise his assumption that A can execute a plan to go to the secret valley).

We believe that an adequate analysis of dialogues like (1) lies outside the scope of formal semantic and extant AI theories of dialogue. Plan recognition techniques alone don't provide a precise semantics of cue words like *but* and so they don't detect a difference in acceptability between (1c) and (1c′). One exception is Green and Carberry (1994), but their plan-based semantics of *but* requires a taxonomy of plans and plan operators that isn't motivated independently of dialogue analysis. Consequently, providing a general axiomatisation of how one reasons with the various types of plans required is problematic. On the other hand, formal semantic theories of questions and answers (e.g., Groenendijk and Stokhof, 1986) do not invoke the complex cognitive states or discourse structures that, as argued by Bromberger (1962), are necessary for analysing the semantic content of *how-* and *why*-questions and the pragmatic constraints on answerhood.

Dialogue (2) below exemplifies a truism that is nevertheless difficult to capture with extant theories of dialogue: people often use the compositional semantics of an utterance to learn different ways of doing an action, when the first attempt to do it failed.

(2) a. A: How do I install the assembly?

b. B: Typically, you put the assembly on before tightening the screws on the assembly.

c. A: But I tried that and it was too difficult.

d. B: So let's tighten the screws on the assembly first then.

The domain knowledge in (2b) is ultimately explicitly ruled out as being irrelevant to achieving the action. If one cannot use a detailed semantics of the utterances to construct the discourse structure, then it's unclear how to predict in precise ways exactly how the participants' domain plan for installing the assembly can change as the dialogue proceeds. B's suggestion in (2d) need not be based solely on an existing plan in his plan library. Rather, he has actively constructed it via the semantics of the dialogue. Discourse structure, though sometimes built up by exploiting domain knowledge, also often affects and even disregards domain knowledge. So one cannot use domain knowledge alone to build plans for executing speech acts or discourse structure; formal semantics is a crucial source of information too.

Dialogues (3) and (4) show that constraints on interpretation given by compositional semantics, in this case concerning anaphora, can refine those given by intentional structure.

(3) a. A: How can I get to 6th Street?

b. B: You can get there by asking someone Downtown.

c. A: ?What's his name?

c$'$. A: Oh yeah? And could you tell me his name?

(4) a. A: How can I get to 6th Street?

b. B: There's someone Downtown that you could ask.

c. A: What's his name?

Intuitively, one cannot use the pronominal reference *his* in (3c) in (3a,b,c), whereas it's acceptable in (3c$'$) in (3a,b,c$'$) and in (4c). It is difficult to see how the plan recognition techniques for dialogue analysis could predict this. This is because the domain level plan alluded to in both (3) and (4) are extremely similar, if not exactly the same. In both cases, the relevant plan is to ask directions for 6th Street from someone who has a sufficient knowledge of the local area. It is conceivable that one could obtain a distinction in the domain level plans along the following lines: that the one underlying (3) involves going Downtown and choosing the person to ask *randomly*, whereas with (4) the person isn't chosen randomly. But it is difficult to motivate this, because it requires an axiomatisation over very fine grained distinctions among the plan operators during dialogue processing. It would put a huge load on this axiomatisation. Moreover, assuming that (3b) and (4b) have different underlying speech acts also lacks independent motivation, especially if B's cognitive state in each case is the same (e.g., in both cases B has a *de dicto* belief that someone knows the way). Therefore, recognising the speech acts doesn't distinguish (3) and (4) either. At any rate, reasoning about the participants' cognitive states is in general difficult and uncertain, and so it would be better to predict constraints on pronominal reference from monotonic procedures for building compositional semantic representations whenever it is possible to do so.

Here, the difference between (3b) and (4b) lies in the scope relations between the existential quantifier (*someone*) and the modalities (*know* and *can*). The difference between (3c) and (3c′) lies in the fact that they connect in different ways to the context (3a,b); there's some aspect of *contrast* in connecting (3c′) to (3b), which is lacking in (3c)), resulting in different constraints on the available antecedents to anaphora (Asher, 1993).[1] We need to make use of a semantics in which these scope differences in (3b) vs (4b), and the rhetorical differences in (3a,b,c) vs (3a,b,c′) affect anaphoric possibilities. More generally, the surface similarities between (3b) and (4b) and (3c) vs (3c′) show how small changes in the surface structure of a sentence should sometimes have profound effects on discourse structure, so that the changes in anaphoric possibilities that result from these small changes in the sentences are recorded. Using compositional semantics to build discourse structure helps achieve this.

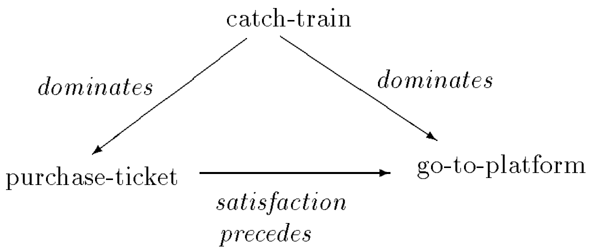
3 Previous Approaches to Dialogue

Many researchers in AI have demonstrated that plans are an important source of information about discourse structure (e.g., Grosz and Sidner 1990, Litman and Allen 1990, Green and Carberry 1994). But plans are not always the most useful clue for constructing discourse structure and content. Moore and Pollack (1992), for example, demonstrated that there isn't a one to one mapping between dialogue and plan structures. Here, we take this further, and argue that a semantic-based theory of discourse structure like SDRT is needed to assess when exploiting plans is appropriate.

Grosz and Sidner (1990) and Litman and Allen (1990) emphasise the distinction between a domain level plan and a discourse level one. In other words, one can intend to do things in the domain—such as eat dinner—and one can intend to perform a speech act—such as persuade the interpreter that a particular proposition is true. The two structures aren't always isomorphic. For example, in dialogue (5), at the domain level, the relevant plan is to catch the train.

(5)
a. A: I need to catch the 1:20 to Philadelphia.
b. Where's it leaving from?
c. B: Platform 7.
d. A: Where do I get a ticket?
e. B: From the booth at the far right end of the hall.

One can assume that by world knowledge, this plan has the structure represented below: catching a train involves (or, in Grosz and Sidner's terminology, *dominates*) two subactions; buying a ticket and going to the platform, where the former satisfaction-precedes the latter (assuming that you must have a ticket before you board the train):



[1]Note that replacing *And* in (3c′) with the explicit contrastive item *But* doesn't change the overall content of this utterance, thereby providing evidence that (3c′) contrasts in some way with the context.

On the other hand, at the speech act level, A has requested information about where to buy a ticket *after* he requests information about the platform, and so the former speech act doesn't satisfaction-precede the latter. Hence the plan for performing speech acts in (5) isn't isomorphic to the domain level plan.

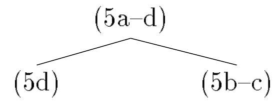
Grosz and Sidner (1990) and Lochbaum (1995) rightly argue that commonsense plans and discourse level ones are closely related, and are often isomorphic. They use techniques discussed by Perrault (1990) to compute the speech act that underlies the utterance currently being processed. Then, using a conversational default rule, they infer from it that the participants want to achieve a particular shared plan. This plan is partially specified, and filling it out involves complex inference about domain knowledge and what gets said in the discourse. They rely on commonsense plans from the plan library to reason about how this shared plan relates to previous ones that were inferred from previous utterances, via the conversational default rule. The resulting intentional structure is then used to structure the dialogue into segments. They argue that the natural segmentation of a discourse is isomorphic to this intentional structure. In other words, a segment that's embedded in another in the linguistic structure means that the discourse segment purpose (DSP) or speech act of the former segment is dominated by that of the latter in the intentional structure. And when one segment follows another, the DSP of the first satisfaction-precedes that of the second.

In the particular case of dialogue (5) this appears to yield unintuitive results, when attaching (5d) to the preceding discourse. Applying their principles in the most straightforward way produces the following segmentation of the discourse segment (5a,b,c).

a		A:	I need to catch the 1:20 to Philadelphia.
	b		Where's it leaving from?
	c	B:	Platform 7.

This is because the conversational default rule yields from (5a) that A and B intend to achieve a shared plan that A catch the 1:20 train to Philadelphia; and from (5b) it yields that A and B intend to achieve a shared plan that B inform A which platform the train is leaving from; (5c) continues this discourse segment purpose (DSP); and in the above domain level plan, the dominance relation between catching a train and going to the platform allows one to conclude that the DSP for (5a) dominates that for (5b,c). So we get an embedding in the discourse segmentation as shown above, because of the isomorphism between linguistic structure and intentional structure.

Processing (5d) proceeds along similar lines, and the conversational default rule is used to infer that A and B mutually believe that A desires to achieve a shared plan that B informs A of where the ticket booth is. According to the above commonsense plan, locating the ticket booth is dominated by catching the train. So as before, one can use this aspect of the domain level plan to infer that (5d)'s DSP is dominated by that of (5a), and hence (5d) is embedded in the top level discourse segment in the above linguistic structure. However, using the above commonsense plan in this way to reason about how (5d)'s DSP is related to that of (5b,c) will produce a counterintuitive linguistic structure. Locating the ticket booth satisfaction-precedes going to the platform in the commonsense plan. If we continue to assume that the relation between DSPs are determined by the structural relations between the actions in the commonsense plan, then (5d)'s DSP will satisfaction-precede (5b)'s, and consequently, by the isomorphism of intentional structure and linguistic structure, the linguistic structure is one where (5d) is not on the right frontier; rather (5b-c) are:



This yields the wrong predictions about possible antecedents to anaphora for utterances that continue the dialogue (for Grosz and Sidner assume a right frontier constraint, in line with Polanyi (1985), Webber (1991) and others).

One might think that this counterintuitive result could be avoided, if instead of appealing to the structural relationship between the *actions* in the commonsense plan about catching trains, one appealed to the dominance and satisfaction-precedence relations between *knowing* how to catch a train, *knowing* where to buy a ticket, and *knowing* which platform the train is leaving from. Although knowing how to catch a train dominates knowing where to buy a ticket and knowing where the platform is, there are no satisfaction-precedence relations between the latter two pieces of knowledge. So if the relationship between (5a), (5b) and (5d) was determined by this epistemic level of the commensense plan about trains, rather than by the domain level actions themselves, then one wouldn't obtain a counterintuitive relationship between (5b) and (5c) in the discourse structure.

One might also think that the incoherence of the following dialogue could be explained by appealing to this epistemic level of the commonsense plan. However, this way out of the problem won't work in general, because the order in which one knows things doesn't usually matter. Dominance relationships at the epistemic level are not unusual. But satisfaction precedence relationships are. It is difficult to think of cases where knowing one thing satisfaction precedes knowing something else. There are some examples in mathematical proofs where this occurs, but this is quite a specialised context. Consequently, there isn't much structure to the epistemic level of a commonsense plan. And so appealing to this alone won't always provide sufficient information on which to determine discourse structure. In essence, satisfaciton precedence would have to take a back seat in Grosz and Sidner's theory, thereby weakening their constraints on antecedents to anaphora.

Indeed, in some cases where plan structure can be used to epxlain a dialogue phenomenon, something else could be used instead. For example, one could explain the incoherence of (6) via reasoning about the commonsense plan:

(6) a. A: I need to catch the 1:20 to Philadelphia. Where is it leaving from?
b. B: Platform 7.
c. A: ?Where can I get information about trains to Philadelphia?

Regardless of whether one uses the underlying commonsense plan at the action level or epistemic level, the DSP underlying (6c) is associated with an item in the commonsense plan that dominates the DSP underlying (6a,b). Under the assumption that a latter utterance dominating a former one is an illegal discourse structure, the isomorphism between intentional structure and discourse structure predicts that (6) is odd. But although the incoherence of (6) can be explained by appealing to plan structure, it needn't be. It could also be explained by appealing to Grice's maxim of Quantity: (6c) violates this maxim, because given the context of this utterance, we can assume that A knows something about how to get information about trains to Philadelphia; otherwise, why did he ask B for information about this? So A violates the maxim of Quantity, by asking a question that B can presume A already knows the answer to. Note how A's utterance is much improved in (6a,b,c$'$):

(6) c$'$. A: Where can I get more information about trains to Philadelphia?

This doesn't appeal to plans about how to catch trains at all, but instead appeals to Gricean-style pragmatic maxims.

Commonsense plans are to be sure an essential source of knowledge in dialogue processing. But we need a way of deciding when we can use the commonsense plan to reason about how DSPs are related to each other, and when we can't. This is currently lacking in Grosz and Sidner's (1990) and Lochbaum's (1995) analyses. We need to predict automatically when the intentional structure of the discourse is isomorphic to the commonsense plan, and when it isn't. By giving a formal theory of the integration between plans, constraints on discourse structure and compositional semantics, we can come closer to this goal.[2]

Intuitively, the dynamic view of compositional semantics familiar from current formal linguistic theories could contribute to this decision about the relevance of the commonsense plan in structuring discourse. In particular, the *compositional semantics* of (5a) informs us that this sentence describes A's intention, and that of (5b) informs us that this is a question.[3] This semantic information, which can be built via *monotonic* procedures, could be used instead of the above domain plan structure, to infer that A asked the question because he doesn't have enough information to achieve the intention. In other words, the monotonic semantic information is used to nonmonotonically infer, via the fact that the intention and the question are in juxtaposed utterances, that the intention prompted the question. Thus (5b) is seen as subordinate to (5a) in the discourse structure because we reasoned via semantics that the question was prompted by the intention, rather than reasoning that one subordinates the other because this is what happens in the domain structure.

However, the commonsense plan would be useful in constraining cooperative responses to (5b), given the context in which A asked it: Any coherent response should elaborate a plan that will enable A to achieve the intention that prompted the question in the first place. So a coherent answer to (5b) must elaborate a plan for catching the 1:20pm train to Philadelphia; thus one must check that (5c) elaborates the above plan for catching a train. A similar line of reasoning would underly the analysis of (5d) and its response (5e)—dynamic semantics, rather than plan structure, would be the central clue for reasoning about how (5d) relates to (5a) and (5c), but then the commonsense plan must be used to ensure (5e)'s an appropriate response, given A's purposes. We will spell out this analysis precisely in section 6.2.

Grosz and Sidner's analysis of dialogue would benefit from a precise formal representation of semantics, since semantic information could help one select exactly when the commonsense plans are relevant to structuring dialogue, via a model of inference that links cognitive states, discourse structure, *and* semantics. The same holds for all AI theories of dialogue that use plan structures to interpret discourse. There are four problems that remain because such inferences aren't modelled. First, one can't attack discourse segmentation. Since discourse segmentation is inferred via plan structure, we need inferences about which plan is most plausibly the one underlying the discourse, so that we can rank choices. This leads to the second problem. Since a lack of inference procedure precludes one from choosing among alternative intentional structures, one can't motivate the discourse structure constructed in dialogue analysis, except by theory bound intuitions. Thirdly, without a model of inference from discourse structure to semantic effects, one can't use the discourse structure to learn more about the semantic content of segments in the dialogue, such as the temporal order of events described in dialogues such as (7), or the inference that the House Bill 1711 is bad for big business in (8) (taken from Moore and Pollack 1992).

(7) A: John's late for work.
B: He must have missed the bus.
C: That was because he had to help with the baby.
B: Yeah, his wife hasn't been feeling well lately

(8) George Bush supports big business. He's sure to veto House Bill 1711.

[2] One could test the following hypothesis: the intentional and commonsense plan structures match as long as it doesn't violate constraints on discourse structure, such as the most recent utterance being on the right frontier.

[3] Also note that the dynamic semantics helps resolve *it* in (5b) to the train (cf. the account of anaphora resolution in Kamp and Reyle, 1993).

4 Our General Strategy

We use SDRT to combine compositional semantics, discourse structure and information about the participants' intentional states in order to over come the inferential problems encountered in the AI plan recognition work. We'll do this for simple dialogues involving the question/answer alternation, with the hope that this yields a general methodology for dealing with other patterns of dialogue.

SDRT has four main advantages. First, the way discourse structure affects and is affected by semantic content has already been studied extensively in this framework (e.g., Asher 1993, Lascarides and Asher 1993). Second, the basic semantic framework which underlies SDRT (Discourse Representation Theory or DRT, Kamp and Reyle 1993), specifies constraints on anaphora resolution that are sensitive to the relative scope of quantifiers such as *someone* and modalities such as *know* and *can*, and we argued earlier that this would be helpful for distinguishing the acceptability of (3) vs (4). Third, DRT has already been used to model cognitive states (e.g., Asher 1986, 1987, Kamp 1990, Singh and Asher, 1993), and so SDRT is ideally suited for folding discourse structure, semantic content and mental states together, so that coherence constraints on dialogue exploit information from all these levels. Finally, a distinctive feature of our analysis is that the rhetorical relations which are used to represent a dialogue impose constraints on multiple levels, not only on the semantic content of the constituents being connected together, but also on the discourse structure surrounding those constituents and on the cognitive states of the participants.

Our plan of attack is first to specify a formal semantics of questions and answers in DRT. Second, we will extend SDRT so that it represents dialogue as well as monologue. We will then use the formal semantics of questions and answers in DRT to help constrain responses to questions in dialogue in SDRT.[4] SDRT will impose constraints on the information flow between discourse structure, semantics and information about the intentional states of the participants in the dialogue. These will augment the existing coherence constraints already specified in the theory. For example, the new constraints on questions and answers and the existing constraints on *Contrast* (and hence the cue word *but*), will predict why (1a,b,c) is better than (1a,b,c′).

4.1 A Crash Course in SDRT

SDRT, which is an extension of DRT, is a theory of discourse structure and content. An NL discourse is represented by a segmented DRS (SDRS), which is a recursive structure of DRSs that represent clauses with rhetorical relations linking them. The rhetorical relations are modeled after those proposed by Hobbs (1985), Polanyi (1985) and Thompson and Mann (1987) and others, and include *Explanation*, *Contrast* and *Continuation*, among others. The model-theoretic interpretation of SDRSs is given in Asher (1993).

SDRT pays particular attention to the way discourse structure affects the content of a discourse, understood in a dynamic fashion. A distinctive feature of SDRT is that the rhetorical relations constrain the semantic content of the constituents they connect together, and these constraints can trigger *modifications* to the content of the constituents as discourse is processed. If information that is used to establish the coherence of a rhetorical relation is not already there, then it's added in a constrained manner. Intuitively, these modifications record the *accommodation* that's necessary to interpret utterances correctly. Asher (1993) shows how encoding these modifications explicitly is useful in assessing how subsequent portions of discourse relate to the previous ones. Our analysis of (1) will involve these tools for accommodation:

[4] We encode these constraints in SDRT rather than the simpler DRT, because SDRT contains essential information about discourse coherence constraints on relations like *Contrast* (and hence cue words like *but*), which is missing in DRT.

(1) a. A: How can I get to the treasure?
b. B: It's at the secret valley.
c. A: But I don't know how to get there.
c$'$. A: ?But I know how to get there.

As we extend SDRT to analyse dialogue, we'll extend the tools for accommodation so that they can be used to modify the cognitive states of the participants, as well as the semantic content of the constituents. In particular, attaching (1b) to the discourse context (1a) with a rhetorical relation will trigger modifications to (1b) and the cognitive state of the interpreter, via formally precise reasoning. Roughly put: (1b) in this context will be interpreted as: *The treasure is at the secret valley, you can get to the treasure by going to the secret valley, and I assume that you know how to go to the secret valley.* These modifications will then affect how the representation of (1c) attaches, because the coherence constraints on the rhetorical relation *Contrast*, as signalled by the cue word *but*, are dependent on the way (1b) is interpreted.

SDRSs are built up incrementally, and building an SDRS from NL discourse involves computing which rhetorical relation should be used to attach the representation of the current clause to the discourse structure built so far. One of the main features of SDRT is the underlying axiomatic theory DICE (Discourse in Commonsense Entailment) which allows us to infer rhetorical relations, using semantic content and the interpreter's background knowledge as clues (Lascarides and Asher, 1993). These clues are represented as default rules, and they enable the interpreter to come to nonmonotonic conclusions about which rhetorical relation to use to bind the constituents in an SDRS together. DICE also contains axioms concerning the semantic effects of rhetorical relations.

The rules in DICE are default conditionals of the form $P > Q$, which means *If P, then normally Q.* For example, `Background` states: If β is to be attached to α with a rhetorical relation, where α is part of the discourse structure τ already (i.e., $\langle \tau, \alpha, \beta\rangle$ holds), and β describes a state (i.e., $state(e_\beta)$ holds), then normally, the rhetorical relation is *Background*. This captures the intuition that states normally provide background information. All rhetorical relations are accompanied by coherence constraints. For example, the `Axiom on Background` states that constituents related by *Background* describe eventualities that temporally overlap; there's also an axiom which states they have a distinct common topic.

- `Background`: $(\langle \tau, \alpha, \beta\rangle \wedge state(e_\beta)) > Background(\alpha, \beta)$
- `Axiom on Background`: $Background(\alpha, \beta) \rightarrow overlap(e_\alpha, e_\beta)$

Since these conditionals are axioms, they are taken to be true everywhere in the model. DICE exploits a nonmonotonic notion of validity ($\approx\!\!|$) with several nice properties, which we have discussed in detail elsewhere (Lascarides and Asher 1993). Four of these are relevant for our purposes. First, $|\!\approx$ validates Defeasible Modus Ponens (DMP): when the default laws whose antecedents are verified all have consequents that are consistent with the KB and with each other, then all the consequents are nonmonotonically inferred. Second, $|\!\approx$ validates Specificity: when conflicting default rules apply, the consequent of the most specific default rule (if there is one) is inferred. Third, $|\!\approx$ is robust in that if $\Gamma|\!\approx\phi$ then ϕ will survive as a consequence of the premises Γ augmented with logically independent information. And finally, for each deduction $P|\!\approx Q$ there is a corresponding embedded default in the object language (that is, a formula in which one occurrence of the connective $>$ occurs within the scope of another) which links boolean combinations of the formulae P and Q, and which is verified to be true. This amounts to a weak deduction theorem, but we are assured of such embedded defaults only for quantifier free languages. It turns out that for analyzing the interaction of cognitive states and discourse structure, we need quantification. To be able to express $|\!\approx$ within the object language at least partially, we will add to the underlying logic a rule that enables us to derive new defaults from the theory. We will call it the `Weak Deduction Rule`:

- **Weak Deduction Rule:**
  Let T be a theory in L_{DICE}, and suppose $T, P \mathrel{|\!\approx} Q$ but $T \mathrel{|\!\not\approx} Q$ and $T \not\models \neg(P > Q)$.
  Then $T \mathrel{|\!\approx} P > Q$.

Weak Deduction preserves consistency but results in a nonconservative extension of DICE, as discussed in Lascarides and Asher (1993). We will use **Weak Deduction** in section 8 to deduce default axioms of DICE from more general principles about how beliefs, desires and intentions interact.

To see how DICE works to compute rhetorical relations, consider (9). DMP on **Background** and then Modus Ponens on **Axiom on Background** will correctly predict that (9b) provides background information for (9a) and temporally overlaps with it.

(9) a. Max opened the door.
 b. The room was dark.

On the other hand in (10), even though the antecedent to **Background** is verified, one can't attach the constituents with *Background* because the constraint that there must be a distinct common topic would be violated, and hence the consequent of **Background** is inconsistent with the KB. Furthermore, the other rules in DICE for computing rhetorical relations fail to supply an alternative, for either the antecedent of the relevant default rule isn't verified, or coherence constraints on the rhetorical relation are violated. So a rhetorical relation cannot be inferred, leading to incoherence.

(10) a. Max opened the door.
 b. ?Mary's hair was black.

Hobbs *et al.* (1993) also use a form of nonmonotonic reasoning on rules that represent the reader's background knowledge, to infer rhetorical relations as discourse is parsed. However, there is a difference in emphasis on their theories and ours. Whereas the approach taken in SDRT is more formally developed, and we wish to ensure that our theory integrates fully with current theoretical linguistic frameworks, Hobbs *et al*'s approach has been shown to integrate with current NLP technology, and there are working implementations of the system in limited domains. Our main concern in this paper is to develop an understanding of the requirements on the formalism underlying an adequate theory of dialogue interpretation. In future work, we will investigate the impact of this research on implementation.

4.2 Extending SDRT to Dialogue

We wish to use SDRT and the theory DICE to model dialogue, so as to take advantage of the way they model the information flow between compositional semantics and discourse structure.

In general, the participants A and B of a dialogue have different cognitive states. And since the SDRS is determined in DICE via the KB of the interpreter, the SDRSs they construct in general differ. So as discourse analysts, we must build at least two SDRSs; one for A and one for B. In fact, A and B must also do this, since they each have models of each other's cognitive states. We analyse misunderstandings between interpreters who understand the words in the dialogue, as occurring when the two SDRSs are not isomorphic. Dialogue participants detect such misunderstandings when their own SDRS does not match the SDRS they construct from their model of the other participant's KB.[5]

[5] We remain agnostic on whether a common ground SDRS, following the lines of (Stalnaker, 1978) and (Cohen and Levesque 1990, Lochbaum 1993, Green and Carberry 1994, and others) should be built as well.

To model the interaction between an agent's cognitive states and discourse structure, we will assume that agents believe not only all the monotonic consequences of their beliefs but the non-monotonic consequences as well. The axiom `Believing Inferences` ensures this.

- `Believing Inferences:` $\mathcal{B}_A\phi$ *if and only if* $KB_A \mathrel{|\!\!\approx} \phi$

Further, as we said above, agents construct SDRSs both for themselves and for the other participants. The participants in a discourse mutually believe that they are all doing this. So we will assume that the set of agents $\mathcal{P}$ in the dialogue mutually believe the axioms of DICE; `Mutually Believing DICE` stipulates this. Finally, we formalize as a rule a reflection principle about mutual belief: `Mutual Belief` states that if A and B mutually believe ϕ, then A believes ϕ and also believes they mutually believe it, and similarly for B.

- `Mutually Believing DICE:` $MB_{\mathcal{P}}\text{DICE}$
- `Mutual Belief:` $MB_{A,B}\phi \rightarrow (\mathcal{B}_A(\phi \wedge MB_{A,B}\phi) \wedge \mathcal{B}_B(\phi \wedge MB_{A,B}\phi))$

These rules enable A and B to use the axioms in DICE, to believe that a discourse has the discourse structure they have computed, and to come to conclusions about the nature of the other participants' SDRSs. We shall examine this claim in detail in section 7.

In addition, it's often useful in dialogue to keep track of who said what. So we will ensure that each agent tags the constituents in his SDRS with who said it: $A{:}\alpha$ means that participant A has said that α. We will occasionally write α for $A{:}\alpha$, when there's no confusion.

SDRT for monologue predicts that a discourse is coherent for an interpreter if he can compute an SDRS for it, via the rules in DICE. This holds for dialogue too. But in addition, an interpreter is not only constructing his own SDRS, but also the SDRSs for the other participants from his model of their KBs, so that he can reason about what they think the dialogue means. Thus the definition of coherence for dialogue is more complex: the current utterance is coherent for an interpreter if he can compute not only a rhetorical relation that connects it to his SDRS, but also a rhetorical relation that he thinks the speaker uses. This latter condition captures the intuition that the utterance is coherent for the hearer only if he can compute what he thinks the speaker intended to convey by it. `Computability`, which is an axiom on what are possible SDRT attachment sites for β captures this.

- `Computability:`
  $\langle \tau, \alpha, B{:}\beta \rangle$ *only if for some* $R \in$ *rhet-relns*, $KB_A, \langle \tau, \alpha, B{:}\beta \rangle \mathrel{|\!\!\approx} \mathcal{B}_B(R(\alpha, \beta))$.

In words, `Computability` states that the hearer A can assume B's utterance is to be attached to α only if he has sufficient information in his KB KB_A to come to a default conclusion about what the speaker B believes is the rhetorical connection between his utterance and the context.[6] That is, the hearer must have a default idea about what the speaker was trying to contribute to the dialogue. Intuitively, this corresponds to the hearer's assumption that the speaker is cooperative.[7] We'll see that `Computability` is violated when attaching (1b′) to (1a), leading to incoherence. But in contrast, it's satisfied for (1b).

In monologue it is natural to think of rhetorical relations in SDRT as linking structured propositions. But in dialogue, a new terminology for describing the discourse structure is possible: speech acts (from traditional philosophy of language) or discourse plans (from AI, e.g., Perrault 1990). Though

[6] This is a very weak assumption, given that one can take some of the rhetorical relations in *rhet-relns* to be very vague or general semantically, and that the above rule doesn't block A from being able to infer several relations between α and β; he simply has to infer at least one.

[7] We ignore the analysis of dialogues in non-cooperative contexts in this paper.

the planning terminology is suggestive, little in the way of inference peculiar to planning rather than standard nondynamic logic seems to be applied in extant theories of dialogue (e.g., Perrault 1990, Grosz and Sidner 1990, Lochbaum 1995, Pollack 1990). So one can think of the discourse structure in SDRT as a discourse plan, where the simple actions in the plan correspond to speech acts (cf. Litman and Allen, 1990).

Suppose one views SDRSs in this way. Then, instead of viewing rhetorical relations in SDRT as linking structured propositions, they link speech acts. This makes each relation correspond to a *speech act type*, where the second term of the relation is a speech act of the appropriate type relative to its discourse context, which of course includes the first term of the relation. For example, *Explanation* is a particular type of speech act, and in *Explanation*(α, β), β is the speech act, performed by the person who uttered β, of asserting an *Explanation*, relative to the discourse context α. This is a richer terminology for speech acts than in traditional philosophy, because speech acts form a recursive structure, and the speech act associated with an utterance is always defined relative to the discourse context in which it was uttered.

For our purposes, the important aspect of viewing relations in SDRT as linking speech acts, rather than structured propositions, is that this emphasises that we must define their semantics at least in part in terms of the cognitive states of the participants. Viewing an SDRS as a discourse plan also makes the SDRT approach to dialogue compatible with Grosz and Sidner's, but with SDRT paying more attention to the interaction between compositional semantic content and discourse structure. In essence, SDRSs are viewed now as intentional structures since they are recursive speech acts, and so discourse structure and intentional structure are isomorphic, just as in Grosz and Sidner's theory. But in contrast to Grosz and Sidner's theory, SDRT incorporates a rich interaction between lexical and compositional semantics in building discourse structure. The gaping hole in SDRT's analysis of dialogue is an examination of the interaction between this semantic information and mental states. Our aim is to do this for the question-answer alternation.

5 The Semantics of Questions and Answers

Since in building discourse structure we wish to exploit monotonic information about compositional semantics whenever it is possible to do so, we require a compositional semantic analysis of questions and answers. As we mentioned, we will concentrate on *how*-questions. Extant formal semantic theories have largely ignored these. We propose an analysis of *how*-questions and their answers that builds on the previous formal semantic work on questions.

An elegant picture of the formal semantics of questions comes from Belnap (1985), Bennett (1979), Karttunen (1977) and Groenendijk and Stokhof (1982, 1986). For them, the meaning of a question is a function $[\![q]\!]$ from worlds to propositions, where propositions are sets of worlds; $[\![q]\!](w)$ is glossed as the true, exhaustive answer at w. All questions can be expressed by lambda expressions with the lambdas binding variables. For Groenendijk and Stokhof, all questions bind a propositional variable. In addition, *wh*-questions bind variables that are the translation of the trace of the moved *wh*-phrase. So the kind of variables they bind will depend on the type of question. Thus the question *Who came to the party?* binds a variable of the same type as a generalised quantifier. For Kartunnen, and Groenendijk and Stokhof, an answer to a question is its extension at an index. So, if the true answer is *John and Mary*, then this amounts to the assertion that John and Mary came to the party, and no one else.

As Ginzburg (1995) points out, this analysis in terms of *exhaustive* answers poses problems when one tries to explain how a quantified expression like *Several politicians* contributes as an answer to the question *Who came to the party?* One might argue that *several politicians* contributes a partial answer, and pragmatics may relax the constraint of exhaustiveness for answers. But in agreement with Ginzburg (1995), we think that this notion of "relaxation" will be quite difficult to capture unless we take the semantics of questions, viz. the *exhaustiveness* requirement on answers, to be some sort of default. But this last assumption seems to get the pragmatics/semantics interaction backward. Ideally, semantics should give monotonic information about content while pragmatics supplements this with defeasible content. It seems rather more likely that exhaustiveness is a pragmatic effect of various Gricean maxims.

Consequently, in line with Ginzburg, we will allow non-exhaustive answers to play a direct role in specifying the semantics of a question. For us, the set of propositions $[\![q]\!](w)$ corresponding to the meaning of a question q at a world w is the set of all true direct answers to the question, including non-exhaustive ones.

We stipulate that the meaning of a question is associated with *true* answers as opposed to just answers, because otherwise the semantics of the same types of questions are indistinguishable. In light of this semantics, it's important to stress that analysing questions this way doesn't preclude false answers from forming coherent responses in dialogue in our analysis. We will see in section 6 that our analysis of dialogue predicts, correctly, that the questioner A in (11) finds B's response coherent, even if he knows that this response isn't true:

(11) A: Who does Bill like?
B: Mary.

So true answers aren't a necessary requirement for coherence, in spite of the role that true answers play in the semantics of questions. The two concepts truth and coherence are obviously related, however. We'll spell out the role truth plays in determining the structure of the dialogue in section 6. Briefly, the rhetorical relation A uses to attach B's response to his question will be affected by whether he has reason to believe the response is false. And this difference in rhetorical structure can serve to influence what A should say next.

We will suppose that the logical form of a *wh*-question is a λ abstract where the bound variable occupies the trace position left by the *wh*-element.[8] *Who loves Mary?* has the logical form $?\lambda x loves'(x, mary')$ (we use the ? to avoid lambda abstracting over propositional variables and to simplify notation). The set of direct answers associated with this question at a world w are all of the form $\alpha(\lambda x loves'(x, mary'))$, such that $w \models \alpha(\lambda x loves'(x, mary'))$, where α is any expression of the type that takes a property as argument and returns a proposition (typically a generalized quantifier denoted by an NP).[9] For us these propositions will be DRSs.[10] If the DRS representing *Several politicians love Mary* is in $[\![?\lambda x(loves'(x, mary'))]\!]$, then this means that several politicians (whoever they are), and perhaps some others, love Mary.

Analysing questions in terms of (perhaps) non-exhaustive answers receives support for embedded *how*-interrogatives, for (12) is true so long as John knows one way of getting to the treasure:

(12) John knows how to get to the treasure.

If the semantics of a *how*-question was defined in terms of exhaustive answers, then John would have to know *all* possible ways of getting to the treasure, in order to know how to get there. However, *who*-embedded interrogatives such as (13) appear to have an exhaustiveness requirement,

[8] In contrast to Groenendijk and Stokhof, we don't λ abstract over the index.

[9] Note that because there is the requirement that the λ-reduced expression is satisfied at the world w, this is a *semantic* characterisation of answerhood.

[10] We do not follow Ginzburg (1995) who analyses questions in situation theory, with an underlying ontology that includes not only propositions, but situations and states of affairs as well. In our analysis, however, propositions are more fine grained than in situation theory. In any case, we want to use SDRT because constraints on anaphora resolution are already defined there, and we feel that answers to certain *wh*-questions—*why*-questions in particular—must be represented in terms of rhetorical relations. An answer to $?why\ \alpha$ is an explanation, or more formally, a proposition of the form $Explanation(\alpha, \beta)$. In contrast to situation theory, representing such rhetorical information in a DRT environment has been studied extensively (Asher, 1993).

which we can't capture via our semantics of *who*-questions alone, since their meaning doesn't constitute the set of exhaustive answers.

(13) John knows who came to the party.

(13) isn't true if John knows only some of the people that came to the party, but not all of them. We will assume that the exhaustiveness associated with (13) arises as a conventional implicature from *know* and *who*. Evidence for this comes from Ginzburg (1995). Compare (13) with (14):

(14) John wonders who came to the party.

Consider a context where John is wondering only about whether some particular set of people were at the party. Then (14) is true, even though John isn't wondering who constitutes the complete set of people that were at the party. So, the difference between (13) and (14) indicates that the exhaustive requirement on the former must be a consequence of the *combination* of *know* and *who*, rather than a consequence of *who* alone. However, we gloss over how the conventional implicature might arise from the lexical entries for *know* and *who* and their syntagmatic co-composition, since embedded interrogatives are not our main concern here.

5.1 The Semantics of How-Questions

We distinguish two kinds of *how*-questions: *restricted how*-questions are those that are accompanied by an adverbial of quantity such as *much*, *often*, *small* and *many* (e.g., (18–20)); *unrestricted how*-questions aren't accompanied by such an adverbial (e.g., (15–17)).

(15) A: How can I get to the treasure?
B: By going to the secret valley and digging under the biggest palm tree.

(16) A: How do you steal pie out from under Mom's nose?
B: Quietly.

(17) A: How is Susan?
B: Fine.

(18) A: How much do you love Susan?
B: A lot.

(19) A: How often do you visit Susan?
B: Twice a week.

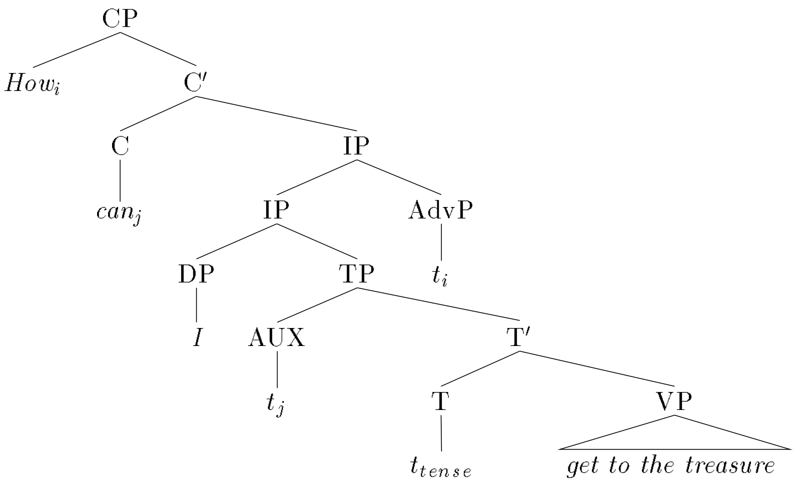
(20) A: How many children do you have?
B: Two.

In some languages, restricted *how*-questions are lexicalised differently from the unrestricted ones. For instance in French, *how much* would be expressed by *combien*, whereas *how* is *comment*.

We take the variables bound by *how*-questions to be adverbials, where the type of the adverbial will depend on the type of the *how*-question. For unrestricted *how*-questions, the adverbial is one of manner. For restricted ones, it's a modifier of an event that realises the event type. So B's responses in (15–20) are all direct answers to the questions A poses.

It's not always clear where the adverbial attaches in the syntactic structure. Indeed, for unrestricted *how*-questions, the scope seems to vary. For instance, the answer in (16) is naturally understood as a predicate of the event and this would suggest that the variable of manner in the question should be adjoined to the VP. On the other hand, this analysis is much less firmly grounded for examples like (15). Given the treatment of gerunds in Asher (1993), the answer in (15) contains an IP, whose translation yields a fact or possibility, not a predicate of events. Given the bottom up construction procedure for DRT in Asher (1993), it makes more sense to attach the manner adverbial to the IP and to relate the fact expressed by the PP with the fact expressed by the main clause. Accordingly, we will treat the adverb in the question as adjoint to the IP. So the syntactic analysis of (1a) (or equivalently, (15)) is given below.

(1) a. How can I get to the treasure?



The trace t_j is caused by moving the auxillary can_j to form a question. But we assume, as is standard, that it leaves a variable of AUX type in its trace. This will aid building the compositional semantics, as we will show shortly. The trace t_i would have to be filled by an adverbial of manner expressed as a *by*-phrase containing a gerund phrase of some sort. Filling it with something like *by ϕ-ing* produces a direct answer to (1a). A direct answer to (1a) is : I can get to the treasure by going to the secret valley.

In asking (1a), A doesn't require a description of every possible way of getting to the treasure. Nor does our direct answer supply one. Indeed, it's unclear how one could construct the complex adverbials that the semantics of a *how*-question would demand, if one analysed them as referring to exhaustive answers. A direct answer may specify all the possible ways of doing some action—the adverbial of manner may be a complex *by ϕ-ing or ψ-ing*. But according to our analysis, this isn't necessary. Moreover, as we mentioned before, our analysis captures the intuition that (12) is true if John knows just one way of getting to the treasure.

In building up the DRS for a question, we must decide how the manner adverbial is incorporated into it. In view of our discussion of gerunds and PPs that contain them, our analysis of manner adverbs like that in (15) is distinct from the typical, Davidsonian or neo-Davidsonian account found for instance in Kamp and Reyle (1993). Our hypothesis is that the semantics of the auxillary exploits the contribution of the adverbial. The modal auxillaries *can* and *do* intuitively take the adverbial to be a condition on the realization of the eventuality described by the VP. We realize this intuition technically by interpreting the modal auxillary as connecting a restrictor P which is filled in by a manner adverbial to the nuclear scope Q which is filled in by the tensed VP.

$Can(\phi(x), \psi(x))$ means roughly x *Can* ψ *by* ϕ. Its semantics should capture two intuitions. First, if x actually does ϕ and intends to bring about ψ, he will normally succeed in doing so, eventually.

Second, x can ψ by doing ϕ only if x can ϕ. We can motivate this second intuition about *can* via the following simple example:

(21) Alfred can go to the moon by taking a spaceship.

Intuitively, (21) is false if Alfred can't take a spaceship for one reason or another; e.g., because there aren't any spaceships. Ensuring that the semantics of *can* entails that Alfred can do the manner in the restrictor captures this. Formally, we express these two intuitions in the semantics of *Can* below, where a semantics for x's intention that ϕ, written $\mathcal{I}_x(\phi)$, is assumed to be like the DRT-based one given in Singh and Asher (1993).

- `Semantics of Can :`
  $M, w \models_g Can(\phi(x), \psi(x))$ iff
  (a) $M, w \models_g (\phi(e, x) \wedge \mathcal{I}_x(\psi(x))) > \exists e'(e' \succeq e \wedge \psi(e', x))$ and
  (b) $M, w \models_g \mathcal{I}_x(\phi(x)) > \exists e''(\phi(e'', x))$

Clause (a) captures the first intuition and clause (b) the second. Clause (b) does this because by the above semantics of *Can*, $Can(\top, \phi(x))$ (or in words, x can ϕ) is true iff (a) $\mathcal{I}_x(\phi(x)) > \exists e'(e' \succeq e_\top \wedge \phi(e', x))$, and (b) $\mathcal{I}_x(\top) > \top$. Clause (b) and $e' \succeq e_\top$ are true trivially. So $Can(\top, \phi(x))$ is true iff $\mathcal{I}_x(\phi(x)) > \exists e''(\phi(e'', x))$ is true, and this is clause (b) in the above definition of $Can(\phi(x), \psi(x))$. Hence x can ψ by ϕ entails x can ϕ.

$Can(\phi(x), \psi(x))$ doesn't entail that $\phi(x)$ or $\psi(x)$ actually hold. On the other hand, the past tense auxillary *Did* seems to have the force of an existential. $Did(\phi(x), \psi(x))$—which should be read as *x did ψ by ϕ*—entails that $\phi(x)$ and $\psi(x)$ both happened in that order in the past and that ϕ was the cause of ψ. We won't give the semantics for *Did*, but it should be obvious (provided one has a semantics for causal connections!) how to do so. We suppose that all auxillaries have either an existential interpretation like *Did* or modal one like *Can*.[11]

The compositional semantic representations of questions are built from the syntax using the "bottom up" DRT construction procedure of Asher (1993). The example (1a) that we are concerned with here is one where the manner adverbial that *how* binds is expressed by a prepositional phrase that includes a gerund phrase. If the manner adverbial is expressed by a subjectless ACC-ing or POSS-ing gerund as in *by going to the secret valley*, then according to the construction procedure in Asher (1993), the gerund phrase will introduce a one-place property with the argument place of this property to be filled in by the controlling subject of the main clause. This property must be an action of the subject, and will be in the restrictor of the modal auxillary *Can*. Assuming that *Can* applies to a tensed VP would imply that the nuclear scope must also be a one-place property of individuals. Thus, the correct logical form of the auxillary *can* should be: $\lambda Q \lambda x \lambda P\, Can(P(x), Q(x))$, where the individual variable x is to be filled in by the partial DRS derived from the subject DP, Q is to be filled by the tensed VP, and P is to be filled by the subjectless gerund phrase.

The DRS construction procedure is shown in Figure 1. We note a few points for clarification. First, we have assumed that the tensed VP *get to the treasure* combines with an AUX variable Δ left by the trace of *can*, but an alternative would be to interpret *can in situ*, prior to movement; the results would be equivalent. This variable Δ is then identified with the semantic contribution of *can*. Second, we have assumed that (1a) is in the present future, which is why the condition $t \prec n$ is in the partial DRS representing the tensed VP. Finally, as is usual in DRT, definite descriptions, indexicals and proper names are promoted to the top level of the DRS, in line with the mechanisms for accommodating presuppositions in DRT which are presented in van der Sandt

[11] One might argue that the above truth conditions of $Can(\phi(x), \psi(x))$ overgenerates. If x can ϕ and x can ψ, then $Can(\phi(x), \psi(x))$ is true as long as x intends to ϕ. This overgeneration results from the fact that the current truth conditions don't record a causal relation between ϕ and ψ; the relation given in clause (a) is somewhat weaker than a causal one. For the sake of simplicity, we have ignored this problem here. The above truth conditions are adequate for the natural language examples that we analyse here.

(1992). The discourse referents i and n stand for the indexicals *I* and *now* respectively. Thus the final representation of (1a) is (1a$'$).

(1) a. How can I get to the treasure?

a$'$. $$?\lambda P \begin{array}{|l|}\hline i, n, x \\ \hline treasure(x), \\ \\ Can\left(\begin{array}{|c|}\hline \\ \hline P(i) \\ \hline \end{array}, \begin{array}{|l|}\hline e, t \\ \hline get\text{-}to(e, i, x), \\ holds(e, t), \\ n \prec t \\ \hline \end{array} \right) \\ \hline \end{array}$$

Similarly, the DRS construction procedure produces a logical form of (22a) that's the same as (1a$'$) except that $n \prec t$ is replaced with $t \prec n$ (because the tense is past) and the auxillary *Can* is replaced with *Did*.

(22) a. How did I get to the treasure?

P must range over adverbials of manner in both of these. So (1a$'$) refers to a set of DRSs, where each element of the set replaces P with a one-place property, such as the representation of the subjectles gerund *by going to the secret valley.* (23) is thus a direct answer to the question (1a):

(23) You can get to the treasure by going to the secret valley.

Its syntactic analysis yields the DRS (23$'$):

(23$'$) $$\begin{array}{|l|}\hline u, n, x, y \\ \hline treasure(x), \\ secret\text{-}valley(y), \\ \\ Can\left(\begin{array}{|l|}\hline e', t' \\ \hline go\text{-}to(e', u, y), \\ holds(e', t'), \\ n \prec t' \\ \hline \end{array}, \begin{array}{|l|}\hline e, t \\ \hline get\text{-}to(e, u, x), \\ holds(e, t), \\ n \prec t \\ \hline \end{array} \right) \\ \hline \end{array}$$

The semantics of *Can* yields several desirable entailments from (23$'$). It entails first you can go to the secret valley, and second that if you go to the secret valley and you intend to ge t to the treasure, then normally you will succeed in doing so.

6 From Semantics to Discourse Structure

We now look at how the above compositional semantics of questions and answers contributes to the discourse structure of dialogue. We will do this by embedding this semantics in SDRT, and we'll investigate how it interacts with the discourse coherence constraints already defined there. We'll explicitly mark questions and their answers by a new rhetorical relation—the *Question Answer Pair* or *QAP*. We'll then use this relation to define the constraints on responses that aren't direct

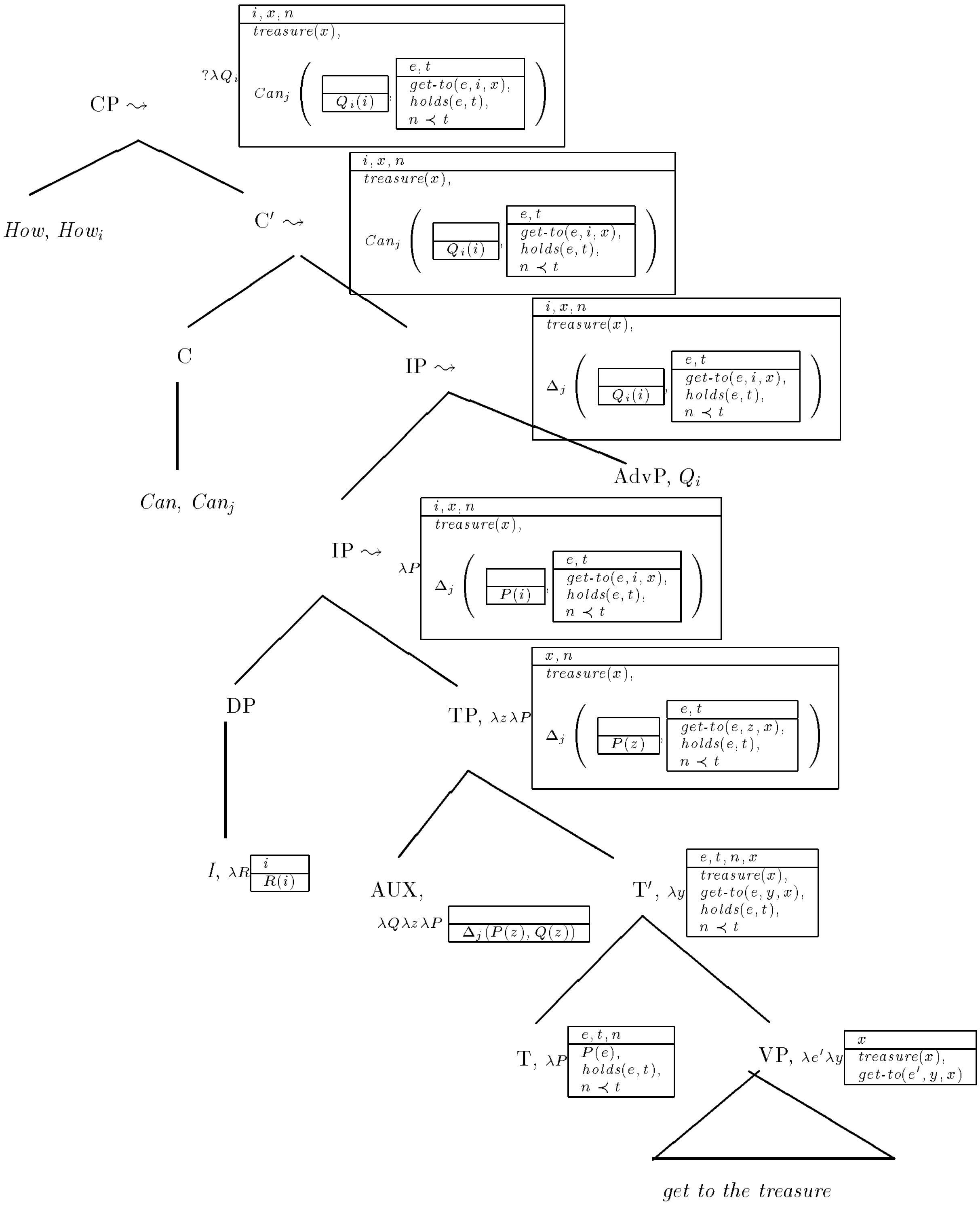


Figure 1: The Bottom up DRS Construction of *How can I get to the treasure?*

answers, an example of which is (1b). QAP is a rhetorical relation that brings in very directly considerations about why participants ask questions and pragmatic constraints about how to respond to them. These pragmatic considerations will not only be important to the construction of SDRSs but also to understanding the information flow between discourse structure and models of the participant's cognitive states.

6.1 The Question Answer Pair

We assume $QAP(A{:}\alpha, B{:}\beta)$ can hold only if α is a question and β is a direct answer to that question, as determined by the compositional semantics of questions and answers specified above. So if α is an unrestricted *how*-question, then $QAP(A{:}\alpha, B{:}\beta)$ can hold only if β specifies an adverbial of manner. In fact, these aren't only necessary conditions, but they are normally sufficient too. More formally, let $Sat\text{-}QAP(\alpha, \beta)$ mean that α and β satisfy the coherence constraints we've imposed on the semantics of $QAP(\alpha, \beta)$. That is, α is a question and β is a true direct answer to that question.[12] Then if β is to be attached to α, and α and β satisfy the coherence constraints imposed by QAP, then normally α and β are a question answer pair:

- `Question Answer Pair (QAP):`
  $(\langle \tau, \alpha, \beta \rangle \wedge Sat\text{-}QAP(\alpha, \beta)) > QAP(\alpha, \beta)$

We think of $QAP(\alpha, \beta)$ as subordinating β to α, so that follow up questions can be related directly to the previous question α and hence the content of the questions can mutually influence each other. If $QAP(\alpha, \beta)$ were treated as a coordinating relation, then the follow up questions couldn't attach to α since it wouldn't be on the right frontier. It's important to stress that it may be difficult or even impossible to infer $Sat\text{-}QAP(\alpha, \beta)$ with certainty for the agent that asks the question. But we note that there may be many methods of nonmonotonically inferring $Sat\text{-}QAP(\alpha, \beta)$, so that the QAP axiom fires. For instance, we might suppose that if A asks a *how* question α and β gives an adverbial of manner that is a doxastically possible direct answer to α–that is, is a direct answer to α in some world compatible with A's beliefs– then that is normally sufficient for B to conclude $QAP(\alpha, \beta)$.

With even this minimal account of QAP in SDRT, we can now exploit the compositional semantics and predict the differences between (3) and (4).

(3) a. A: How can I get to 6th Street?
b. B: You can get there by asking someone Downtown.
c. A: ?What's his name?

(4) a. A: How can I get to 6th Street?
b. B: There's someone Downtown that you could ask.
c. A: What's his name?

By the monotonic procedure for DRS construction, the representation of (3b) and (4b) are respectively (3b′) and (4b′)

[12] It is important to note there may be other coherent responses to a question besides a true direct answer. Indeed, a coherent response may even be another question. We will attach such coherent responses with a rhetorical relation other than QAP, to reflect the fact that they have a different relation to the question than that of a true direct answer. The analysis of coherent responses that aren't true direct answers is examined in detail in section 6.2.

(3) b′.

$$\begin{array}{|l|}\hline \mathit{you}, d \\ \hline \mathit{downtown}(d), \\ \\ \mathit{Can}\left(\begin{array}{|l|}\hline e', t', y \\ \hline \mathit{ask}(e', \mathit{you}, y, p), \\ p = \mathit{?}, \\ \mathit{location}(y, d) \\ \hline \end{array}, \begin{array}{|l|}\hline e, t, x \\ \hline \mathit{get\text{-}to}(e, \mathit{you}, x), \\ \mathit{holds}(e, t), \\ n \prec t, \\ x = \mathit{?} \\ \hline \end{array} \right) \\ \hline \end{array}$$

(4) b′.

$$\begin{array}{|l|}\hline \mathit{you}, y, d \\ \hline \mathit{downtown}(d), \\ \\ \mathit{Can}\left(\begin{array}{|l|}\hline e', t' \\ \hline \mathit{ask}(e', \mathit{you}, y, p), \\ p = \mathit{?} \\ \hline \end{array}, \begin{array}{|l|}\hline e, t, x \\ \hline \mathit{get\text{-}to}(e, \mathit{you}, x), \\ \mathit{holds}(e, t), \\ n \prec t, \\ x = \mathit{?} \\ \hline \end{array} \right) \\ \\ \mathit{location}(y, d) \\ \hline \end{array}$$

In both cases, the conditions $x = ?$ and $p = ?$ stem respectively from the anaphor *there* and the fact that the verb *ask* requires a question that is asked. These anaphors are resolved by reasoning that (3b)/(4b) are related to (3a)/(4a). Indeed, the definition of availability in SDRT means that p and x must pick their antecedents from those that are accessible in DRT terms (i.e., not in embedded DRSs) in the open constituent in the discourse structure (i.e., constituents on the right frontier of that structure). There is only one open constituent: (3a)/(4a). So regardless of the rhetorical relation that connects them, the only candidate for x is *6th Street*, and the only candidate for p is the question *How can I get to 6th Street?*

Different syntactic structures in (3b) and (4b) lead to the discourse referent y (which represents the person to be asked) being available in SDRT terms for anaphoric reference in (4c), but not in (3c). We haven't had to reason with A's or B's cognitive state to predict this. We didn't exploit open ended nonmonotonic inference procedures with cognitive states, such as those proposed in Perrault (1990), to produce the representations of (3b) and (4b). And so we haven't had to consider issues such as: Is B's belief that someone knows how to get to 6th Street *de dicto* or *de re*? Rather, the procedure was monotonic. Thus, modelling the way information is presented in a sentence can refine the constraints on pronominal reference supplied by the more open ended plan recognition techniques. Obviously, DRS structure alone isn't sufficient to define all the relevant constraints on pronominal reference (see Asher (1993) for detailed arguments). But it can play a useful part in the story, because without it one would have to obtain all constraints on anaphora from the intentional structures behind the dialogue. This requires a distinction among plan structures that is too fine grained to be motivated independently.

6.2 Indirect Answers

Some responses aren't direct answers, but nevertheless contribute in coherent ways as responses to the question.[13] This occurs in (1), where (1b) isn't an adverbial of manner.

[13] Indeed, on a search of two million words of the BNC corpus, 14 questions starting *How can...* were found, and none of these had direct answers in response (that is, the answers weren't adverbials of manner). However, apart from the two examples (24b) and (24c) cited below, they were all responses like that in (24a), where an adverbial of manner can be inferred.

(1) a. A: How can I get to the treasure?

b. B: It's at the secret valley.

Such a response won't meet the coherence constraints on *QAP*. However, it is an indirect answer, in the sense that we can compute the direct answer (23) from it, given the discourse context and what we know about the world.

(23) You can get to the treasure by going to the secret valley.

Assume for now that (23) is true (in particular, that A can go to the secret valley). Then the DRS representing (23) can attach to the representation α of (1a) with *QAP*. So in general, a response will be an indirect answer if the interpreter can infer from it a proposition γ such that $QAP(\alpha, \gamma)$ holds. Predicting when this is possible thus requires complex reasoning about the discourse context and cognitive state of the interpreter. For example, (1b) is an indirect response to (1a), because A's knowledge about locations, spatial displacements, and actions allows him to compute the direct answer (23) from (1b).

In contrast to (1b), A can't compute the direct answer from (1b′) that B intended to convey, without making some very unusual contextualised assumptions about what B thinks A believes.

(1) a. A: How do I get to the treasure?

b′. B: ?Mary's hair is black.

So, A can't support a belief that B thinks she's providing an indirect answer, because to believe this she must assume B thinks A has sufficient knowledge to compute an adverbial of manner for getting to the treasure from (1b′), and this assumption is untenable. Indeed, A can't work out what B intended by (1b′) at all, and so the response is incoherent for A.

We now formalize these intuitions. We represent the relationship between indirect answers and questions at the discourse structure level via the relation *Indirect Question Answer Pair* (*IQAP*), and the following coherence constraint captures the above intuition about questions and indirect answers: namely, the questioner must be able to nonmonotonically infer a direct answer γ from his KB augmented with the indirect answer β. So by the axiom `Weak Deduction` given in section 4.1, this means that A must simply be able to infer the conditional $\beta > \gamma$, where he can also infer $QAP(A{:}\alpha, \gamma)$ too. This constraint is captured in **`Axiom on IQAP`**.

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

The axiom on *IQAP* entails that direct answers are also indirect answers; *IQAP* generalizes *QAP* because $\beta > \beta$ is an axiom on DICE and A believes these axioms.[14] More importantly, this axiom links information about A's cognitive state to compositional semantics and the rhetorical relation

(24) a. A: How can we help reduce this kind of accident?
B: For a start, novice pilots should have some experience of aerotowing in single seaters before converting to types which are not fitted with a nose hook, and gliders intended for inexperienced pilots should always be manufactured with a nose hook for aerotowing.

b. A: How can I get rid of head lice?
B: It's very important to get treatment for lice and nits as soon as you find them.

c. A: How can I do that?
B: That is in the hands of the government.

[14] Notice that given our axiom indirect answers and direct answers don't exhaust the set of answer-like responses that we might encounter in dialogue. But we don't intend our classification to be exhaustive.

used in discourse structure. On the one hand, the conditional $\beta > \gamma$ could be believed by A because he inferred γ nonmonotonically from his KB augmented with β, and then used the axiom **Weak Deduction** and **Believing Inferences** to believe $\beta > \gamma$. So A may believe this conditional because of his background knowledge, including his plan library. On the other hand, the rhetorical structure $QAP(A{:}\alpha, \gamma)$ is true only if γ is a direct answer to α, as defined by the compositional semantics of questions and answers. So our framework is rich enough to express rules which relate three levels of analysis: compositional semantics, rhetorical structure and cognitive states.

This axiom is quite a weak condition on indirect answerhood, because it stipulates only that A be able to nonmonotonically infer a direct answer γ from his *whole* knowledge base, giving him the opportunity to exploit plans and intentions as well as propositions. For example, if the response to a question is an instruction, as in (25), then A can use his knowledge of plans to infer the direct answer *the platform that the porter takes you to*:

(25) a. A: What platform is the 1:20 to Philadelphia leaving from?
b. B: Follow that porter. He'll put you on the train.

Thus in spite of the fact that (25b) features a directive speech act, which is distinct from the declarative speech act that direct answers to (25a) have, (25b) nevertheless satisfies the above constraint on indirect answerhood. In this sense, speech acts take a back seat in this theory.

Like the constraint on QAP, the necessary condition for $IQAP$ is normally sufficient too. So if $Sat\text{-}IQAP(A{:}\alpha, B{:}\beta)$ means that α and β satisfy the above coherence constraint on $IQAP$ (that is, the consequent of the **Axiom on IQAP** holds for α and β), then the following specifies that in such a situation, one normally attaches the constituents together with $IQAP$:

- **Indirect Question Answer Pair (IQAP):**
  $(\langle \tau, \alpha, \beta \rangle \wedge Sat\text{-}IQAP(\alpha, \beta)) > IQAP(\alpha, \beta)$

IQAP implies that as a byproduct of inferring a rhetorical relation between α and β, the agent computes a direct answer. In (25), for example, A can attach (25b) to (25a) with $IQAP$ only if he can compute the direct answer *the platform the porter takes you to* from his KB and the instruction (25b), presumably via his plan library. On the other hand, in a dialogue such as (26), the 'indirect' answer B gives doesn't enable the hearer to work out a direct answer, and both A and B know this:

(26) a. A: How do I get Downtown?
b. B: Well, you get there either by taking I-79 or I-279, but I'm not sure which.

According to our definition, partial answers like that in (26b) aren't indirect. In fact, (26b) implicates that B doesn't know a direct answer to (26a). In examples like (25) and (26), if A doesn't have the necessary information in his KB to compute a direct answer from B's response, then he can't attach with $IQAP$; we'll shortly see what relation he uses instead.

Let's see in formal detail how these rules enable us to analyse (1).

(1) a. A: How can I get to the treasure?
b. B: It's at the secret valley.

Given our semantics of *how*-questions, the DRS representing (1a) is α below. The one representing (1b) is β:

$$(\alpha) \qquad ?\lambda P \begin{array}{|l|}\hline i, n, x \\ \hline treasure(x), \\ \\ Can\left(\begin{array}{|c|}\hline \\ \hline P(i) \\ \hline \end{array}, \begin{array}{|l|}\hline e, t \\ \hline get\text{-}to(e, i, x), \\ holds(e, t), \\ n \prec t \\ \hline \end{array} \right) \\ \hline \end{array}$$

$$(\beta) \qquad \begin{array}{|l|}\hline e', y, z, t', n \\ \hline at(e', z, y), \\ secret\text{-}valley(y), \\ z = ?, \\ holds(e', t'), \\ n \subseteq t' \\ \hline \end{array}$$

Our task now is to connect them together with a rhetorical relation. The only antecedents accessible to the anaphor z in β are those in α and β. Given the syntactic and semantic agreement constraints on z, the only possible antecedent is x—the treasure. So we can assume that the representation of (1b) is β_1 before we attempt to compute the rhetorical relation.

$$(\beta_1) \qquad \begin{array}{|l|}\hline e', x, y, t \\ \hline treasure(x) \\ secret\text{-}valley(y) \\ at(e', x, y) \\ hold(e, now) \\ \hline \end{array}$$

To complete the SDRS for (1), we must compute the rhetorical relation between α and β_1 using DICE. We will consider two cases: one where A knows that he can go to the secret valley, and one where he can't but B thinks he can. Let's examine the first case first. Let's also assume, as intuitions would dictate, that A has a general, domain level plan in his library which represents the following information: the action of going to the location y of an object x generates the action of getting to the object x (where *generates* is assumed to be defined as in Pollack (1990)). Then using techniques suggested by Pollack (1990) and Lochbaum (1993), one can fill in the x parameter of this plan with the treasure. Moreover, the semantics β_1 of (1b) permit A to fill in the parameter y with the secret valley. So this information obtained from the compositional semantics β_1 of (1b), together with the above very general domain plan from the plan library, allow A to construct a generates relation between actions (which could be added to the plan library): going to the secret valley generates getting to the treasure. Goldman (1970) and Pollack (1990) suggests that the generates relation between actions in plans is closely connected to natural language *by*-phrases. In our terms, this is tantamount to the assumption that if $\alpha_1(x)$ generates $\alpha_2(x)$ in a domain plan and x can do α_1, then $Can(\alpha_1(x), \alpha_2(x))$ holds—that is, x can do α_2 by doing α_1. So, assuming this link between the modality *Can* in our extension of DRT and the definition of generation in plans which Pollack alludes to, A uses this generates relation between actions that he has inferred via the semantic content of β_1, to infer nonmonotonically the direct answer (23) to (1a), as represented in the DRS (23′) given earlier. Let's rename (23′) as γ. Then this discussion about A's KB demonstrates that A can use the `Weak Deduction` axiom in DICE to infer a new condition: $\beta_1 > \gamma$.

Since we are for the moment assuming that A can go to the secret valley and that A believes this (thus assuming γ is true), A can infer $(\beta_1 > \gamma) \wedge QAP(\alpha, \gamma)$ holds. So by `Believing Inferences`, he believes this. Hence the coherence constraints on $IQAP(\alpha, \beta_1)$ are satisfied. Thus the antecedent to `IQAP` is verified. So is the antecedent to `Background` verified, since β_1 describes a state. These default rules are the only ones that apply, and both consequents are consistent with A's KB. Therefore by defeasible modus ponens (DMP), A nonmonotonically infers $IQAP(\alpha, \beta_1)$ and $Background(\alpha, \beta_1)$.

It is important to stress that inferring this discourse structure in this example crucially depended on the assumptions both that A can go to the secret valley and that A believes this. If A can't go to the secret valley, by clause (b) of the semantics of *Can*, γ is false, and so can't constitute part of an answer to the question (since $QAP(\alpha, \gamma)$ entails γ is true). This restriction matches intuitions. B's attempts to perform a speech act of providing a correct answer to α fail if the associated direct answer is false. So if A doesn't believe that he can go to the secret valley, then he cannot accept an SDRS that stipulates that the speech act type *IQAP* has been performed.

This raises the question: what discourse structure does A compute in this second case to be considered? And on what grounds is (1b) a coherent response, given that A can't compute that it's an answer, indirect or otherwise? As we've mentioned in section 4.2, (1b) is coherent for A only if he can compute how β_1 attaches to α in his SDRS, and how it attaches in his model of B's SDRS—as demanded by the attachment constraint `Computability`. Let's deal with A's model of B's SDRS first. The axiom `Answer` below captures the intuition that normally, a respondent to a question thinks he's answering it:

- `Answer`: $\langle \tau, A{:}?\alpha, B{:}\beta \rangle > \mathcal{B}_B(IQAP(?\alpha, \beta))$

`Answer` describes cooperative behaviour during communication. It is a specific instance of a more general Gricean maxim of cooperativity. We will derive `Answer` from this more general principle in section 8. There are exceptions to `Answer` such as (26) and (27), but they're usually linguistically explicit.

(27) A: How do I get to the treasure?
B: I'm afraid I can't answer that question.

Examples like (26) and (27) would be accounted for by conflicting, more specific default rules than `Answer`, which would enable the participants to infer via Specificity a different rhetorical relation from *IQAP*.

`Answer` permits A to satisfy the `Computability` constraint when attaching (1b) to (1a): He uses it to infer that B's SDRS is $IQAP(\alpha, \beta_1)$. But what about his own SDRS? As we've stated, in the case we're considering here where A can't go to the secret valley and believes this, A can't attach β_1 to α with *IQAP* in his SDRS, even though he can infer that this is what B does. So intuitively, β_1 doesn't provide enough information for A to construct a true answer to the question α. So A should use the relation *NEI* (*Not Enough Information*) to relate β_1 to α. From the standpoint of monologue *NEI* might appear to be an odd discourse relation, because it seems to assert paradoxically the lack of a coherence relation between two constituents. But this would be a misunderstanding of the relation. NEI is an acceptable discourse relation for dialogue, because it reflects and codifies in the SDRS a coherent and common move in dialogue–one person's answer to another's question is signalled as not being sufficient.

NEI is inferrable in DICE with the help of the following axiom `NEI`. It states: if B's response β is to be attached to A's question α, and $IQAP(\alpha, \beta)$ doesn't hold although B believes it does, then normally β is to be attached to α with *NEI*. That is, B's response doesn't provide enough information for the questioner A:

- `NEI`:
$(\langle \tau, A{:}\alpha, B{:}\beta \rangle \wedge Question(\alpha) \wedge \neg IQAP(\alpha, \beta) \wedge \mathcal{B}_B(IQAP(\alpha, \beta))) > NEI(\alpha, \beta)$

So A has constructed different SDRSs for himself and for B. He uses *IQAP* via **Answer** for B, but *NEI* for himself because he knows γ (or (23′)) is false. A similar analysis would apply to (25), if B believes that the porter will take him to the wrong platform, for example. We'll show in section 7 that the DICE rules we've given capture the intuition that the difference in the SDRSs stemmed from B making incorrect assumptions about A's cognitive state, and we'll see how A then uses (1c) to get B to correct those assumptions.

But first let's contrast the above analysis of (1a,b) with that of (1a,b′).

(1) a. How do I get to the treasure?
    b′. ?Mary's hair is black.

Let's assume, as intuitions would dictate, that in contrast to (1b), A is unable to assume that B has a model of A's KB which includes a domain plan that permits A to compute an adverbial of manner from (1b′) for getting to the treasure. For otherwise, A has a very unusual model of what B thinks A can do. Then, A must assume that B believes the consequent of the **Axiom on IQAP** is false. Therefore, by the logic of belief, A assumes B doesn't attach (1b′) to (1a) with *IQAP*; for otherwise A is committed to B having an inconsistent KB. Therefore, although the antecedent to **Answer** is verified, its consequent can't be inferred, because it's inconsistent with A's KB—specifically his model of B's KB. Indeed, A can't come to any conclusions about which discourse relation B used to attach (1b′) to (1a). For just as in (10), although the antecedent to **Background** is also verified, it's consequent isn't inferrable because the constraint that there is a distinct common topic would be violated. Therefore, the SDRT attachment constraint **Computability** is violated. A cannot infer, even nonmonotonically, how B attached (1b′) to (1a), and this inability on A's part to compute what B intended yields incoherence.

Another example of a failure to provide even an indirect answer to a question occurs in (2). Just as in dialogue (1), B infers *IQAP*(2a, 2b) and A infers *NEI*(2a, 2b) in (2):

(2) a. A: How do I install the assembly?
    b. B: Typically, you put the assembly on before tightening the screws on the assembly.
    c. A: But I tried that and it was too difficult.
    d. B: So let's tighten the screws on the assembly first then.

The analysis proceeds along similar lines to those above. B's response isn't an adverbial of manner, and so it doesn't satisfy the constraints for attaching (2b) to (2a) with *QAP*. However, it refers to a plan which specifies an order between action types. On the assumption that A successfully infers that the plan referred to is part of a plan for installing the assembly, A can infer an adverbial of manner from the response B gives in (2b): Install the assembly by putting the assembly on before tightening the screws on the assembly. But as in (1), this direct answer is false, because he knows that he can't do the adverbial of manner suggested and so the axioms on the semantics of the modality *do* are violated. Nevertheless, as in (1), he uses DMP on **Answer** to infer that B has *IQAP*$(2a, 2b)$ in his SDRS, and he uses DMP on **NEI** to infer *NEI*$(2a, 2b)$ for himself. Having detected the discrepancy in their SDRSs, A must tailor his next utterance to informing B that there is this discrepancy.

Our approach also helps relieve the pressure on domain knowledge for constructing discourse structures for dialogue. To see this, let's reconsider (5).

(5) a. A: I need to catch the 1:20 to Philadelphia. α
    b. Where's it leaving from? β
    c. B: Platform 7. γ
    d. A: Where do I get a ticket? δ
    e. B: From the booth at the far right end of the hall. ϵ

The compositional and lexical semantics of (5a,b) inform us that A states an intention and then requests information. Since (5a,b) each give rise to a constituent in SDRT—call them α and β respectively—we need to compute a rhetorical relation between them. Intuitively, the agent is signalling with his question that he does not have enough information to carry out the intention in α. We mark this again with the discourse relation *Not Enough Information* or *NEI*. The DICE axiom `NEI II` below thus reflects this intuition:

- `NEI II`:
  $(\langle \tau, A{:}\alpha, B{:}\beta\rangle \wedge \mathit{Intention}(\alpha) \wedge \mathit{Question}(\beta)) > NEI(\alpha, \beta)$

We understand $\mathit{NEI}(\alpha, \beta)$ as subordinating β to α, to reflect the fact that a coherent response to β must help A produce some action relevant to the realization of the intention stated in α. There may be several occurrences of the *NEI* relation in dialogue, but eventually we will 'pop' out from these, once the questioner has enough information from all his queries to bring about his desire stated in α.

In dialogue (5), both A and B infer $\mathit{NEI}(\alpha, \beta)$ via DMP on `NEI II`. And so the discourse structure at this stage looks as follows:

$$\begin{tikzcd} \alpha \arrow[d, "\mathit{NEI}"'] \\ \beta \end{tikzcd}$$

This contrasts with the plan recognition analysis of (5a,b), where commonsense plans are used to compute the relation in the discourse structure. We have used compositional semantics as the clue for computing *NEI*, rather than a commonsense plan, since the antecedent of `NEI II` is verified by compositional semantic information.

Now, we can compute via the compositional semantics of questions and answers that the constituent γ derived from B's response (5c) is a direct answer to the question β. Moreover, the relation $\mathit{NEI}(A{:}\alpha, A{:}\beta)$ can serve to constrain coherent responses γ to the question β. Intuitively, γ is coherent (for A) only if it helps him achieve the intention in α that prompted the question β. In other words, γ must elaborate a plan which when executed achieves the intention—that A catch the 1:20 train to Philadelphia.

We can formally specify this coherence constraint on γ in DICE. Let *Plan-Elab* be the rhetorical relation Plan Elaboration, so $\mathit{Plan\text{-}Elab}(\phi, \gamma)$ means that γ describes part of a plan which, when executed, results in ϕ. Let $\mathit{Info}(\phi, \gamma)$ summarise all the monotonic information given by ϕ and γ. Then `Constraint on Answers` encapsulates the relevant constraint. It states the following. If β is a question that A asked to gain information on how to achieve his intention ϕ which is stated in α (as asserted by $[\mathcal{I}_A(\phi)](\alpha) \wedge NEI(\alpha, \beta)$); and γ is a direct answer to β (as asserted by $QAP(A{:}\beta, B{:}\gamma)$); then A must be able to use the background knowledge in his KB (which includes the plan library), augmented with the information given in ϕ and γ, to nonmonotonically infer that the answer γ elaborates a plan for achieving ϕ.

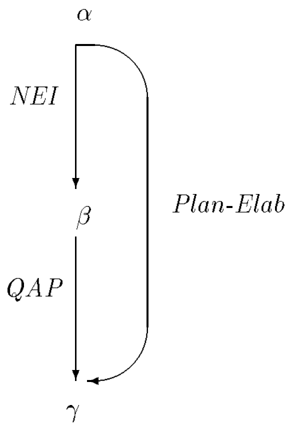
- `Constraint on Answers`:
  $([\mathcal{I}_A(\phi)](\alpha) \wedge \mathit{NEI}(\alpha, \beta) \wedge QAP(A{:}\beta, B{:}\gamma)) \rightarrow \mathcal{B}_A(\mathit{Info}(\phi, \gamma) > \mathit{Plan\text{-}Elab}(\phi, \gamma))$

So in dialogue (5), the direct answer γ must satisfy `Constraint on Answers`, given that the question β stands in the relation *NEI* to α. In other words, A must believe the conditional

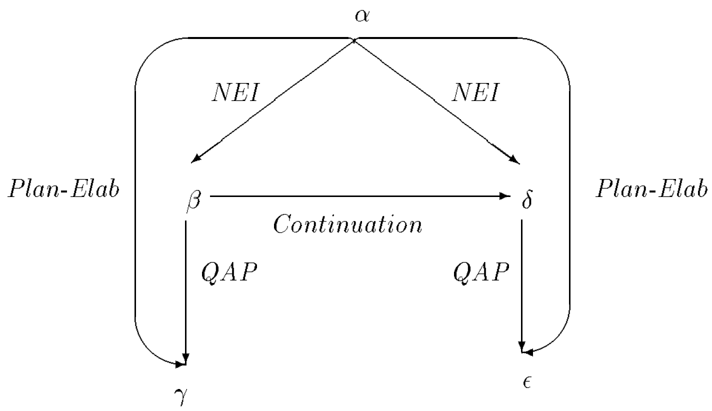
Info$(\phi, \gamma) >$ *Plan-Elab*(ϕ, γ). He can believe this as a result of using the `Weak Deduction` axiom, so long as his KB (which includes background knowledge) augmented with the information in ϕ and γ allows him to nonmonotonically infer that γ elaborates a plan for achieving ϕ. So, to check the coherence of γ, A needs to appeal to domain knowledge or a plan library. But this should be no surprise: it is the same strategy as already found in DICE for inferring *Elaborations* (Lascarides and Asher, 1993). Intuitively, given A's domain knowledge about how to catch a train (specifically, the plan we specified in section 3), the information in γ is part of what is needed to catch a train. More formally, the information in γ is a discourse-permissible subtype ($Subtype_D$) of the information needed to catch a train because of this commonsense plan. So by the default rule given below (cf. the axiom `Elaboration` in Lascarides and Asher 1993), A is able to nonmonotonically infer *Plan-Elab*(β, γ) from his KB when processing dialogue (5), as required by `Constraint on Answers`.

- `Plan Elaboration:`
  $(\langle \tau, \alpha, \beta \rangle \wedge Subtype_D(\beta, \alpha)) >$ *Plan-Elab*(α, β)

Thus, γ successfully attaches to the available constituents on the right frontier: the relations are *QAP*(β, γ) and *Plan-Elab*(ϕ, γ), where ϕ is *catch-train*(A) and is part of the information in α:



QAP is a subordinating relation, and so the follow up question (5d) can attach to γ, β or α— they're all on the right frontier. The representation δ of (5d) attaches to α in the same way as β did because of `NEI II`. Moreover, we can encode a rule in DICE which models the intuition that when two questions are both asked to achieve a particular intention in the discourse structure, then these questions are related by *Continuation*. Finally, B's response ϵ yields similar inferences to those used to attach γ. So the SDRS that A and B build for dialogue (5) is as follows:



This analysis shows that in DICE, domain knowledge and a plan library are still useful when inferring elaborations. With elaborations in monologue, we may have to use domain knowledge to conclude that one event is possibly part of another event; here, we use the plan library to check that the information that A receives is part of what is needed to carry out a plan to catch a train. However, semantics lets us use this knowledge selectively. We don't always need to use commonsense plans as clues for structuring discourse. The structure of the commonsense plan was used to infer the *Plan-Elab* relations in the above, but not the *NEI*, *QAP* or *Continuation* relations, which also contribute to the overall discourse structure.

These examples demonstrate several things. First, they show the advantages of using compositional semantics to help provide precise coherence constraints to responses to questions in dialogue. It is unclear how one defines incoherence in the analyses of dialogue which use plan recognition techniques alone (e.g., Lochbaum 1993). Clearly, (1b$'$) fails to fill in any of the parameters in plans that are in the library and relevant to getting to the treasure. However, this alone cannot be grounds for incoherence, because if it were then it would not be possible to learn new plans from dialogue. Rather, the interpreter at this stage should construct a partial plan which has getting to the treasure as its goal, and which has a partially specified action as a part which utilises the colour of Mary's hair. Of course, this latter action would be almost completely unspecified, because A can't work out how to utilise Mary's hair colour to achieve getting to the treasure. So the question then arises: at what stage is such a plan so partial, that the dialogue sounds odd? It is not clear what the answer should be if all one has recourse to is the structure of plans and the cognitive state of the participants. However, using the compositional semantics of questions and answers supplies a precise definition of this threshold: it is 'too partial' if it fails to allow A to compute what direct answer to the question B intended to allude to. In the case of *how*-questions, therefore, the plan constructed is "too partial" if it fails to support an adverbial of manner by which to bring about the event in the question. Thus the compositional semantics of questions and answers complements the analyses of dialogues that exploit reasoning about plans, by providing precise specifications of when the plan constructed through interpreting dialogues yields incoherence.

Second, these examples show how the DICE axioms we've given so far model two compelling intuitions. The first is that they allow the questioner A to learn something about what the respondent B thinks A thinks. For suppose A uses `Answer` to infer his model of B's SDRS, and to ensure `Computability` is satisfied. Then by the `Axiom on IQAP`, A believes that B believes that his response β was one which A could use to derive a direct answer γ. So A learns something about what B thinks A thinks. $\mathcal{B}_B\mathcal{B}_A(\exists\gamma(\beta_1 > \gamma) \wedge QAP(\alpha,\gamma)))$ will now hold in A's KB, as a consequence of using `Answer` to compute what B's SDRS is. A could even take things further, and reason about exactly what direct answer B thinks A will compute, by assuming B has modelled A's knowledge of geometry.

The second compelling intuition that DICE validates is that when B thinks he's provided an answer to a question, he thinks that the questioner A thinks so too. Of course, the respondent B may use axioms and theorems other than `Answer` to infer $IQAP(\alpha,\beta_1)$ in his model of the dialogue—he may use `IQAP`, for example. But at any rate, whichever way B infers $IQAP(\alpha,\beta_1)$, the constraints on $IQAP$ given in `Axiom on IQAP` mean that whenever B infers it, $\mathcal{B}_A(\exists\gamma(\beta > \gamma) \wedge QAP(\alpha,\gamma)))$ must be in B's KB. Moreover, this belief affects the way B builds his model of A's SDRS. DICE is mutually believed, and so B can compute from his model of A's KB, which contains $\exists\gamma(\beta > \gamma) \wedge QAP(\alpha,\gamma))$, that A uses the default DICE axiom `IQAP` to nonmonotonically infer $IQAP(\alpha,\beta)$. Hence B nonmonotonically concludes $\mathcal{B}_A(IQAP(A{:}\alpha, B{:}\beta))$. This nonmonotonic inference followed directly from $IQAP(A{:}\alpha, B{:}\beta)$ being in B's model of his own SDRS. So, by `Believing Inferences`, $\mathcal{B}_B(IQAP(A{:}\alpha, B{:}\beta))$ entails $\mathcal{B}_B\mathcal{B}_A(IQAP(A{:}\alpha, B{:}\beta))$ in DICE. That is, whenever B believes he's giving an indirect answer to A's question, he believes that A believes this too. Thus our axioms that specify how to compute rhetorical relations, and very general axioms concerning belief, model how hearers can use what's said in the dialogue as clues to infer things about what the speakers believe about themselves and the hearers.

7 Interactions between the Discourse Relations

In the previous section, we saw how the discourse relations *QAP* and *IQAP* serve to specify constraints on the coherence of responses to questions and information about the cognitive state of the participants in the dialogue. Here we consider how these relations and the constraints they engender interact with other discourse relations like *Contrast*. We will show how the dialogue setting enables us to use *Contrast* to learn about the participant's cognitive states as well.

7.1 Informed Responses in SDRT

We used `Answer` in section 6 to capture an intuition about cooperativity: a cooperative respondent thinks he's answering the question. Moreover, the DICE axioms validate that whenever B believes he's answering A's question, he believes that A believes this too. This is an instance of how A and B use what's said in the dialogue as clues to infer things about what they believe about themselves and each other. They enable an interpreter to detect discrepancies between his beliefs and the speaker's. This is essential, if one is to tailor subsequent utterances to the needs of the participants (Moore and Paris, 1993). This interaction between the SDRSs and cognitive model will lead us to rethink how rhetorical relations may operate in the dialogue. We will focus on *Contrast* and the role of *but* in our examples (1) and (2).

Since DICE enables the interpreter to infer things about the speaker's beliefs, he can utter things to rectify discrepancies that he detects. This is what happens in dialogues (1) and (2). Let's turn back to the analysis of (1a,b), in the situation where A can't go to the secret valley, and examine in more detail how A infers things about B's cognitive state from his utterance (1b).

(1) a. A: How can I get to the treasure?
    b. B: It's at the secret valley.

As we showed in section 6, A computes $NEI(\alpha, \beta_1)$ in his SDRS, and $IQAP(\alpha, \beta_1)$ in his model of B's SDRS. Thus the logical closure of belief and `Axiom on IQAP` mean that DDRS_B below is part of A's model of B's cognitive model—that is, $\mathcal{B}_A\mathcal{B}_B(\text{DDRS}_B)$ holds in A's KB—where α' is the DRS representing *I get to the treasure*:

$$(\text{DDRS}_B)\quad \boxed{\begin{array}{l} \\ \hline \mathcal{B}_A\left(\begin{array}{c}\beta_1,\\ IQAP(\alpha,\beta_1),\\ Can\left(\begin{array}{|c|}\hline \\ \hline P(A)\\ \hline\end{array}, \alpha'\right),\end{array}\right)\end{array}}$$

Let's now analyse how A and B reason about the content of (1c), given the SDRSs and models of their cognitive states that they have constructed so far.

(1) a. A: How do I get to the treasure?
    b. B: It's at the secret valley.
    c. A: But I don't know how to get there.

The DRS for (1c) is δ below (we've made the simplifying assumption that *don't know how to* is analysable as *can't*); this must be attached to β_1 with a discourse relation.[15]

[15] We have resolved the anaphor *there* to the secret valley, because this is the only candidate antecedent according to SDRT's definition of availability of antecedents, regardless of the rhetorical relation that's used to attach δ to β_1.

$$(\delta)\qquad \begin{array}{|l|}\hline i,n,y\\ \hline \mathit{secret\text{-}valley}(y), \\ \neg \begin{array}{|l|}\hline \\ \hline \mathit{Can}\left(\top, \begin{array}{|l|}\hline e,t\\ \hline \mathit{get\text{-}to}(e,i,y),\\ \mathit{holds}(e,t),\\ n\prec t\\ \hline\end{array}\right)\\ \hline\end{array}\\ \hline\end{array}$$

A has used the cue word *but*, which monotonically yields the relation *Contrast*(β_1, δ). This rhetorical relation, like all others, is accompanied by its own set of coherence constraints. These are defined precisely for monologue in Asher (1993). Roughly speaking, *Contrast*(α, β) in monologue requires a partial structural isomorphism and contrasting themes between α and β. These constraints are in the spirit of those defined for *Contrast* in Hobbs (1985). But they go further, in that they are able to check the coherence of elaborately structured constituents, such as propositional attitude reports like δ. These constraints capture the intuition that *Contrast*(β_1, δ) is coherent if δ indicates that some expectation arising from β_1 has been violated.[16] In the case of monologue, the expectations that arise from β_1 are essentially semantic and pragmatic entailments about the domain of discourse that β_1 refers to. But as we have seen, in the case of dialogue the expectations aren't simply recorded at the domain level. We have shown how the discourse action of uttering β_1 yields expectations about the content of the participants' cognitive states, via the axiom **`Answer`** (B thinks an adverbial of manner is inferrable for A from β_1); the axioms on *Can* (this adverbial of manner is the restrictor for the connective *Can*, and A must be able to do this manner for the adverbial to be an answer to the question); and the **`Mutual Belief`** axioms (A is aware of how B thought A should attach β_1 to α). So in order to check that δ violates an expectation arising from β_1, we must consider not only the content of β_1, but also the inferences we obtained from the utterance β_1 about the cognitive states of the participants.

We capture this by extending the extant coherence constraints in SDRT on *Contrast*. We assume that in the case of dialogue, a partial structural isomorphism and contrasting theme between δ and an expected cognitive state, given the utterance β_1, is sufficient for *Contrast*(β_1, δ) to be coherent. This is in contrast to monologue, where the partial structural isomorphism and the contrasting themes must hold between δ and β_1 itself.

7.2 Details Concerning Contrast in Dialogue

Let's see how this works. Asher (1993) defines structural isomorphism precisely, using the notion of an embedding tree (ET). For our purposes, one can assume that an ET of a DRS is the usual DRS structure. So, for example, the ET for δ above is the following three-level tree, where $\neg K$ is the negated DRS, and β' is the DRS representing *I go to the secret valley*:

$$(\mathrm{ET}_\delta)\qquad \begin{array}{ccc} & \delta & \\ & | & \\ & \neg K & \\ / & & \backslash \\ \top & & \beta' \end{array}$$

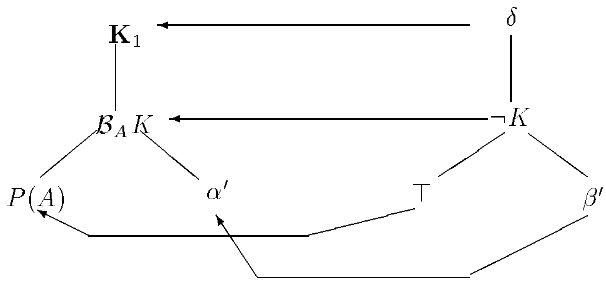
Our task is to ensure that this forms a partial isomorphism and contrasting theme either with the constituent β_1, or with some expectation of cognitive states that was based on his uttering β_1. In other words, we must find an SDRS or DRS which has an ET where the mapping between the nodes

[16] It is important to stress that this isn't a necessary condition, but intuitively it is a sufficient one.

of the ETs is as close to a bijection as possible, and where the nodes can be assigned opposite polarities because of their semantic content (e.g., one node contains the negated information of the node it's mapped to).[17] In our example, the constituent β_1 itself doesn't meet the coherence constraints. But the part $\mathbf{K}_1$ below of A's model of B's cognitive state—which is part of DDRS$_B$—does meet the requirements:

$$(\mathbf{K}_1) \qquad \boxed{\mathcal{B}_A\left(Can\left(\boxed{P(A)}, \alpha' \right)\right)}$$

It has an ET which, as shown below, forms a bijection with ET$_\delta$ (the node $\mathcal{B}_B K$ stands for the believed DRS K in $\mathbf{K}_1$, and $P(A)$ is the DRS representing the adverbial of manner):



This isomorphism also yields the contrasting theme we require: δ stipulates that A can't go to the secret valley; $\mathbf{K}_1$ stipulates that A can get to the treasure; and according to the Asher (1993) definitions for assigning polarities to the nodes in the structure, this is sufficient for the nodes $\mathcal{B}_A K$ and $\neg K$ to have opposite polarities (and the node pairs α' and β', and $P(A)$ and $\top$, agree in polarity). So the mapping of structures has produced the required contrasting theme, thereby ensuring that B successfully attaches δ to β_1 with *Contrast* in his SDRS.

But now, having successfully inferred *Contrast*(β_1, δ), B's model of A's cognitive state and δ contain contradictory elements. And yet δ is, nonmonotonically, part of B's model of A's beliefs via the following `Sincerity` axiom, which captures the intuition that people believe what they say:

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

B was assuming that A believed he could go to the secret valley. It is for this reason that B believed that his response (1b) could be attached with *IQAP* to (1a). But the above `Sincerity` axiom, when applied to δ, nonmonotonically yields that A believes he cannot. So, B must do some belief revision. A's discourse behaviour in uttering (1c) together with `Sincerity` has exposed to B that he got A's SDRS and his model of A's beliefs wrong. So, the DICE axioms have captured the following: A used B's utterance to infer what B thought A thought; detected a discrepancy; and so uttered (1c) which was specifically designed via the *Contrast* relation to enable B to detect his mistake and rectify it. Thus we have information flow from utterances to cognitive states and back again, via general axioms that encapsulate the semantics of rhetorical relations, cooperation and mutual belief.

Let's assume B drops the belief that A can go to the secret valley. Then $\mathbf{K}_1$ must be removed from B's cognitive state. This means that the coherence constraints on *IQAP*(α, β_1) aren't met

[17] The formal details of how to do this when an actual bijection isn't possible are to be found in Asher (1993;286–291). These details can be skipped over here, however, because in this case there is a bijection.

anymore. So B must drop $IQAP(\alpha, \beta_1)$ too. But now what discourse relation binds α to β_1 in B's revised SDRS? As B has inferred that A couldn't have attached β_1 to α with $IQAP$, he concludes that A must have used the axiom **NEI** to attach the constituents together with *NEI*. So we assume he does the same.[18] Now that B's model of A's cognitive state has been revised, the *Contrast* connection between β_1 and δ isn't tenable anymore either. Again, we assume that when *Contrast* has prompted the interpreter to revise an indirect answer to something that is "not enough information", the contrast relation is also replaced with *NEI*, to indicate that responses must address the lack of A's knowledge that's asserted in δ.

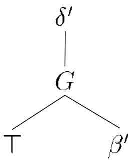
The SDRS structure underlying (1) has effects on the plans A intends to execute. These will be predicted in a precise way in section 8. But first compare our analysis of (1c) with (1c$'$). The representation of (1c$'$) is δ', which has the ET shown below, where β' is again the DRS that A goes to the secret valley:

(1) c$'$. But I know how to get there.

(δ')

$$\begin{array}{|l|}\hline y, n, i \\ \hline \textit{secret-valley}(y), \\ \\ \textit{Can}\left(\top, \begin{array}{|l|}\hline e, t \\ \hline \textit{get-to}(e, i, y), \\ \textit{holds}(e, t), \\ n \prec t \\ \hline \end{array}\right) \\ \hline \end{array}$$

(ET$_{\delta'}$)



This ET fails to produce a contrasting theme with either β_1 or with the above element $\mathbf{K}_1$ of B's cognitive state. This cognitive state doesn't support the use of *but* anymore, because one can't assign opposite polarities to the node pair $\mathcal{B}_A K$ and G, as one did to $\mathcal{B}_A K$ and $\neg K$, because $\mathcal{B}_A K$ and G contain the same information. So attachment of δ' to β_1 with *Contrast* isn't possible. Hence (1c$'$) is incoherent in the context of (1a,b), given the way we reasoned about what B believed, from observing his discourse behaviour of uttering (1b). We have not used plans and cognitive states alone to model the difference in acceptability between (1c) and (1c$'$); rather, we have demonstrated the utility of folding together compositional semantics with reasoning about cognitive states, the latter being familiar from AI theories of dialogue.

As we said earlier, dialogue (2) also contains a use of contrast that exploits the information about one of the participant's cognitive states. We analyze it analogously to (1).

(2) a. A: How do I install the assembly?

b. B: Typically, you put the assembly on before tightening the screws on the assembly.

c. A: But I tried that and it was too difficult.

d. B: So let's tighten the screws on the assembly first then.

As before, A reasons via `Answer` that B believes $IQAP(2a, 2b)$, and so A believes B thinks A can execute the plan described in (2b). A then B that he cannot do this plan and so sets up a contrast

[18] We gloss over the formalisation here of how revision is pursued so as to favor achieving SDRSs between the participants that agree in as many respects as possible.

with B's representation of what A can do. Just as in (1), there is a conflict via `Sincerity` between B's cognitive state and the information in (2c). In the subsequent response (2d), B revises both his SDRS and his representation of A's cognitive state and then comes up with an alternative plan to help A accomplish his goal. The discourse particle *so* indicates a result relation; the result of not being able to do the plan in β is to try the different plan in δ.

8 Reasoning with Cognitive States and Discourse

In this last section, we introduce some general techniques for reasoning about cognitive states from discourse contexts. One of the things we will do is derive `Answer` from these more general rules.

In Asher and Lascarides (1994), we used a general axiom which relates desires, beliefs and intentions, and which is known as Aristotle's Practical Syllogism, to model the information flow between semantic content and cognitive states in simple monologues. We use The Practical Syllogism here to model the information flow in dialogue. In particular, we can use the axiom to nonmonotonically infer that when A asks a question and B responds, then B thought that his response would enable A to achieve one of his *question related goal* or QRGs—that is, a goal g that A has which prompted him to ask the question (because he believes an answer will enable him to achieve g). In other words, we can infer that B believes he answered not only coherently, but appropriately too.

A's question asking is intentional: he has a goal in mind in asking the question. People ask questions for many purposes, but presumably one purpose that always *prima facie* applies in cooperative dialogue—which is the only kind of dialogue we consider here—is the purpose to get a correct answer to the question. A cooperative respondent should typically respond with a correct answer as far as he is able to judge. In the previous section, we saw how `Answer` describes awareness of this cooperative behaviour.

Searle (1969) classifies questions as a kind of request. Here, we link questions with desires, via general principles that link action, beliefs and intentions. In essence, we're attempting to furnish Searle's insights into speech acts with some formal detail, about how compositional semantics and cognitive states are linked.

Aristotle's Practical Syllogism states that normally, people intend to do things that they believe help them achieve their goals. The DICE axiom below describes this. If (a) B wants ψ and believes $\neg\psi$, and (b) B believes that he can nonmonotonically infer ψ from his KB augmented with ϕ, and moreover ϕ is B's choice for achieving ψ, then normally (c) B intends ϕ.

- `Practical Syllogism:`
  (a) $(\mathcal{W}_B(\psi) \wedge \mathcal{B}_B(\neg\psi) \wedge$
  (b) $\mathcal{B}_B((\phi > \mathit{eventually}(\psi))) \wedge \mathit{choice}_B(\phi)) >$
  (c) $\mathcal{I}_B(\phi)$

Now that we know what the `Practical Syllogism` is, let's return to the issue of the questioner's QRGs. Why would A ask the question? We cannot say unequivocally; there might be all sorts of reasons. But as we stated, at least *prima facie*, one reason for A asking $?\alpha$ is that A wants to know an answer, and doesn't have this knowledge now. We can formalize this in our logical framework as `QRG`:

- `QRG:` $A{:}?\alpha > \exists\gamma(QAP(\alpha,\gamma) \wedge \mathcal{W}_A\mathcal{K}_A(\gamma) \wedge \neg\mathcal{K}_A(\gamma))$

If we take this to be, defeasibly, a QRG of questions, then we see that A's asking, for example, a question α of the form *How can I ϕ?* entails defeasibly, if χ is a direct answer to α, that

$\mathcal{W}_A(\mathcal{K}_A(\chi))$ and $\neg\mathcal{K}_A(\chi))$ are true. Clause (a) of `Practical Syllogism` is thus verified with $\mathcal{K}_A(\chi)$ replacing ψ (and A replacing B), and if A believes that by asking a question α he can get an answer then clause (b) is verified too, with $A{:}?\alpha$ replacing ϕ. So we derive by `Practical Syllogism` that $\mathcal{I}_A(A{:}?\alpha)$. That is, we have derived within DICE A's intention to perform the speech act of asking a question.

We could go further and tie question answering behavior to further goals. One could suppose, for instance, that if an agent asks a *how*-question of the form *How can I ϕ?*, normally he wants to achieve ϕ but believes that $\neg\phi$. Suppose that B answers A's question with β, and that β has the effect B intended, which we modelled in the previous section via `Answer` and `Mutual Belief`. That is, β tells A that there is a plan ψ that A can do that will normally bring about ϕ. Once again the antecedent of the `Practical Syllogism` is verified. So, we can conclude that the consequent $\mathcal{I}_A(\psi)$; that is, A intends to execute the plan ψ given by the answer. Thus `Practical Syllogism` is used here to reason about how A's interpretation of the dialogue affects the actions he intends to do.

A concrete example of this line of reasoning surfaces in (1). We'll assume, in line with the above, that in addition to the QRG given by `QRG` (i.e., the goal of knowing a correct answer to (1a)), A also has the QRG given by the DRS α', which represents *A gets to the treasure*. Furthermore, let's suppose now that (1b) was appropriate. That is, A successfully computes the SDRS $IQAP(\alpha, \beta_1)$ for the DRS representations α and β_1 of (1a) and (1b) that we defined earlier. Hence A can infer from β_1 that for some ψ, $Can(\psi, \alpha')$ holds and therefore by the semantics of *Can*, he can do ψ, and doing it nonmonotonically will yield α', as long as he intends α'. So assuming nothing else about A, we infer in virtue of the `Practical Syllogism` that A intends ψ. Thus, we have the following fact: let P_a be the plan that specifies the manner a for the question q which is of the form *How can I ϕ?* Then: $(How\text{-}can\text{-}I\text{-}\phi(q) \wedge QAP(A{:}q, a)) \approx\!\!\!| \mathcal{I}_A(P_a)$ is valid.

The particular specification of the intended plan will depend on what plans were already in A's library, as we stated in section 6. These considerations are discussed in detail by Litman and Allen (1990) among others, and are beyond the scope of this paper. We would need to say a great deal more about plan generation and plan revision. But we still have shown how `Practical Syllogism` could be used to affect intentions.

Let's now use `Practical Syllogism` to look at how responses work. If B's cooperating, then he wants to do more than answer the question correctly. He also wants to help A achieve whichever other goals lay behind the question. In other words, if A has a QRG p—and so in particular, $\mathcal{W}_A p$ and $\mathcal{B}_A\neg p$ hold—then normally B cooperates by adopting p as his goal too. This is expressed by the axiom `Cooperation`:

- `Cooperation:` $(\mathcal{W}_A p \wedge \mathcal{B}_A\neg p) > (\mathcal{W}_B p \wedge \mathcal{B}_B\neg p)$

The axioms `Cooperation` and `The Practical Syllogism` play an essential part in the reasoning behind how responses to questions are interpreted in dialogue. They can be used to infer that B thought he was answering the question appropriately. For suppose that B reasons that A's QRG for α is p. Then $\mathcal{W}_A p$ and $\mathcal{B}_A\neg p$ form part of B's KB. So by DMP on `Cooperation`, B infers $\mathcal{W}_B p$ and $\mathcal{B}_B\neg p$; that is, B wants A to achieve his goals. Let ϕ in the `Practical Syllogism` stand for the discourse action $\langle A{:}\alpha, A{:}\alpha, B{:}\beta\rangle \wedge Info(\alpha, \beta)$, and ψ stand for p. Then because of the above inference from `Cooperation`, clause (a) in the antecedent of the `Practical Syllogism` holds. Moreover we assume that B's discourse action was intentional, and so the consequent (c) also holds. So as in Asher and Lascarides (1994), the antecedent (b) can be abduced: B believes that his discourse action nonmonotonically achieves p. That is, A will fulfil his goal p as a result of B's discourse action, by default.

We now return to `Answer`. The reasoning that we have just put in the head of B is reasoning that A can also do *about* B, and presumably would do about B assuming `Cooperation` and the fact that B responds to A's question with β. So we can now derive the following fact:

- `Practical Syllogism`, `Cooperation`, `QRG`, $\langle\tau, A{:}\ ?\alpha, B{:}\beta\rangle \approx \mathcal{B}_B(IQAP(A{:}?\alpha, B{:}\beta))$

Now, since `Practical Syllogism`, `Cooperation` are axioms of DICE and since we cannot derive nonmonotonically $\mathcal{B}_B(IQAP(A{:}?\alpha, B{:}\beta))$ if we do not assume $\langle\tau, A{:}\ ?\alpha, B{:}\beta\rangle$, we have by the `Weak Deduction` rule of CE given in section 4.1 that `Answer` is an axiom of DICE:

- DICE $\approx \langle\tau, A{:}?\alpha, B{:}\beta\rangle > \mathcal{B}_B(IQAP(?\alpha, \beta))$

Thus the cooperative axiom `Answer` that we introduced in section 6.2 is derivable in DICE from the more general axioms about how beliefs, desires and intentions interact, defined in the `Practical Syllogism` and `Cooperation`. These more general axioms have utility in all kinds of situations, not just the ones where dialogue is taking place.

8.1 Appropriacy and Contextual Relevance

One can think of (1b) or (2b) as *inappropriate* responses for A's purposes. Intuitively, A asked (1a) because he had a goal of getting to the treasure. An appropriate response should enable a questioner to achieve the goal he had in mind when he asked the question. Although (1b) was coherent, it was inappropriate for A's purposes because it provided insufficient information for A to achieve the goal behind his question—this was because A doesn't know how to get to the secret valley. This 'inappropriacy' was reflected in A's SDRS by the fact that β_1 attached to α with *NEI*.

Inappropriacy can occur in another guise. In (1b) or (2b), the response was inappropriate for A, but B was in ostensibly no position to recognize the inappropriacy. But inappropriacy can occur for other reasons, even when A, like B, successfully attaches the response to the question with *IQAP*. For example, consider the following dialogue (28c,d) in the contexts provided in (28a) and (28b); these are slightly modified examples of those given in Ginzburg (1995):

(28) a. [Context: A has just travelled by air]
b. [Context: A has just taken a five minute taxi ride]
c. A: Where am I?
d. B: Edinburgh.

Intuitively, (28d) is an appropriate response to (28c) in the context (28a), but not in the context (28b). Our analysis captures this intuition. Given `QRG`, A presumably wants to know a direct answer to his question that he does not already know. (28d) is a true direct answer to (28c) in both contexts. In contrast to (1), the SDRSs for A and B are the same as they both compute *QAP* between the constituents. However, what's odd about (28d) in the context of (28b) is that B should be aware that A knows what city he's in since he's only gone for a five minute taxi ride, and A assumes that B knows A knows which city he's in. So the answer B gives does not conform to what A thinks B would utter, if B was using the `Practical Syllogism` and `Cooperation`. For if B were using these axioms, then B's answer (28d) leads A to assume that B has odd assumptions about A's cognitive state, which aren't justified in this minimal context and so B is not being ideally rational. The alternative is that B is being deliberately uncooperative, which also leads us and A to wonder why. This deviation from the norm, with no explanation why, makes the dialogue sound odd.[19]

[19] The notion of appropriacy sketched here also seems to address some aspect of the problem of granularity of questions and answers. An answer to a question may not be appropriate because it is of the wrong granularity–for example, Rich Thomason (p.c.) has suggested that A told B what is in the basket if A says to B that an apple is in the basket. He hasn't told B who is in the next room when he says that a woman is next door. It seems to us that in the second case the answer may not be appropriate to B's qrg's, and that is why the answer doesn't seem helpful. Clearly, this issue requires alot more thought.

In contrast to the context (28b), it is more plausible that A doesn't know which city he's in after a long flight, and so (28d) doesn't violate `Cooperation` or `QRG` in the context of (28a). Like Ginzburg's analysis, our explanation of the difference between the two contexts makes use of some information about questioners' QRGs. But in addition, we tie inferences about questions to more general Gricean constraints on communication, which apply not only to question answer alternations, but other forms of dialogue as well.

9 Conclusions

We have shown how links between compositional semantics at the intra-sentential level, coherence constraints on rhetorical relations at the discourse level, and intentional reasoning at the cognitive state level, contribute to the interpretation of questions and responses in dialogue. A distinctive feature of our analysis was that one could compute what an utterance in dialogue meant via nonmonotonic inferences which exploited *both* compositional semantics and cognitive states. Furthermore, the way utterances are interpreted at the discourse level affected the cognitive states of the participants in the dialogue. For example, a questioner A learns what the respondent B thinks, by reasoning about the content of B's response. We modelled these cognitive effects via very general axioms which link belief, intention and desire, and which have utility beyond the dialogue situation.

These features of the analysis provided a convenient means for predicting when a contribution to a dialogue is coherent, and if coherent, when it is appropriate for achieving the participants' goals. At a general level, the axioms that we specified provide a first step towards linking formal semantic theories of intra-sentential analysis with discourse interpretation theories familiar from computational linguistics, where both rhetorical structure tools and plan recognition techniques have been used. Providing this combination in a single framework provides a richer environment for reasoning about the structure of dialogues, and their effects on the cognitive states of the participants. We demonstrated in our theory that even a small change in the surface syntax of a sentence can have a radical effect on the structure and meaning of the dialogue, in virtue of the communication links between compositional semantics at one level, and cognitive states at the other.

10 References

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

Nicholas Asher [1987] A Typology for Attitude Verbs, *Linguistics and Philosophy*, **10**, pp.125–197.

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

Nicholas Asher and Robert Koons [1993] The Revision of Beliefs and Intentions in a Changing World, in *Precedings of the* AAI *Spring Symposium Series: Reasoning about Mental States: Formal Theories and Applications*.

Nicholas Asher and Alex Lascarides [1994] Intentions and Information in Discourse, in *Proceedings of the 32nd Annual Meeting of the Association of Computational Linguistics (ACL94)*, pp.35–41, Las Cruces.

Nicholas Asher and Alex Lascarides [1995] Lexical Disambiguation in a Discourse Context, *Journal of Semantics*, **12**, pp.69–108.

Nicholas Asher and Pierre Sablayrolles [1995] A Typology and Discourse Semantics for Motion

Verbs and Spatial PPs in French, *Journal of Semantics*, **12** 2, pp.163–210.

Nuel D. Belnap, Jr. [1985], Questions and Answers in Montague Grammar, in Stanley Peters and Esa Saarinen (eds.) *Processes, Beliefs, and Questions*, pp.165–198, D. Reidel, Dordrecht.

Michael Bennett [1979] Questions in Montague Grammar, Indiana University Linguistics Club.

Sylvain Bromberger [1962] An Approach to Explanation, in Butler, R. (ed.) *Analytical Philosophy*, Oxford University Press, pp.72–105.

Philip Cohen and Hector Levesque [1990] Rational Interaction as the Basis for Communication, in in Philip R. Cohen, Jerry Morgan and Martha E. Pollack (editors) *Intentions in Communication*, pp.221–255. Cambridge, Massachusetts: Bradford/MIT Press.

Jonathan Ginzburg [1993] Resolving Questions, technical report HCRC/RP-46, Human Communication Research Centre, University of Edinburgh.

Jonathan Ginzburg [1995] Resolving Questions parts I and II, to appear in *Linguistics and Philosophy*.

Nancy Green and Sandra Carberry [1994] A Hybrid Reasoning Model for Indirect Answers, *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pp.58–65.

Jeroen Groenendijk and Martin Stokhof [1982] Semantic Analysis of *wh*-Complements, *Linguistics and Philosophy*, **5**.

Jeroen Groenendijk and Martin Stokhof [1986] *Studies on the Semantics of Questions and the Pragmatics of Answers*, PhD thesis, Centrale Interfaculteit, Amsterdam.

Barbara Grosz and Serit Kraus [1993] Collaborative Plans for Group Activities, in *Proceedings of the International Joint Conference in Artificial Intelligence (IJCAI93)*, Chambery, France, August 1993, pp.367–373.

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

Barbara Grosz and Candice Sidner [1990] Plans for Discourse, in Philip R. Cohen, Jerry Morgan and Martha A. Pollack (eds.) *Intentions in Communication*, MIT Press, pp.417–444.

Joseph Halpern and Yaram Moses [1984] Towards a Theory of Knowledge and Ignorance: Preliminary Report, in *Proceedings of the Non-Monotonic Reasoning Workshop*, AAAI Fall Symposium, October 1984, pp.125–143.

Jerry Hobbs [1985] On the Coherence and Structure of Discourse. Report No. CSLI-85-37, Center for the Study of Language and Information, October, 1985.

Jerry Hobbs, Mark Stickel, Douglas Appelt and Paul Martin [1993], Interpretation as Abduction, *Artificial Intelligence*, **63 (1–2)**, pp.69–142.

Hans Kamp [1990] Propositional Attitudes, in C. Anthony Anderson, Joseph Owens (eds.), *The Role of Content in Logic, Language, and Mind*, CSLI Lecture Notes **20**, University of Chicago Press, pp.27–90.

Hans Kamp and Uwe Reyle [1993] *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers.

Lauri Karttunen [1977] Syntax and Semantics of Questions, *Linguistics and Philosophy*, **1**.

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

Diane Litman and James Allen [1990] Discourse Processing and Commonsense Plans, in Philip R. Cohen, Jerry Morgan and Martha A. Pollack (eds.) *Intentions in Communication*, MIT Press, pp.365–388.

Karen Lochbaum [1993] A Collaborative Planning Approach to Discourse Understanding, Technical Report TR-20-93, Center for Research in Computing Technology, Harvard University.

Karen Lochbaum [1995] The Use of Knowledge Preconditions in Language Processing, in *Proceedings for the 14th Joint International Conference in Artificial Intelligence (IJCAI95)*, Montreal, August 1995.

William C Mann and Sandra Thompson [1986] Rhetorical Structure Theory: Description and Construction of Text, ISI technical report number RS-86-174.

Johanna Moore and Cecile Paris [1993] Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information, *Computational Linguistics*, **19** 4, pp.651–694.

Johanna Moore and Martha Pollack [1992] A Problem for RST: The Need for Multi-Level Discourse Analysis, *Computational Linguistics*, **18** 4, pp.537–544.

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

Livia Polanyi [1985] A Theory of Discourse Structure and Discourse Coherence. In *Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society*, pp. 25–27.

Martha Pollack [1986] *Inferring Domain Plans in Question-Answering*, Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Martha Pollack [1990] Plans as Complex Mental Attitudes, in Philip R. Cohen, Jerry Morgan and Martha A. Pollack (eds.) *Intentions in Communication*, MIT Press, pp.77–104.

Rob van der Sandt [1992] Presupposition Projection as Anaphora Resolution, *Journal of Semantics*, **9**, pp.133–377.

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

Munindar Singh and Nicholas Asher [1993] A Logic of Intentions and Beliefs, *Journal of Philosophical Logic*, **22**, pp.513–544.

Robert Stalnaker [1978] Assertion, in Cole, Peter (ed.) *Syntax and Semantics volume 9: Pragmatics*, Academic Press.

Bonnie Webber, B. [1991] Structure and Ostension in the Interpretation of Discourse Deixis, *Language and Cognitive Processes*, Volume **6** Number 2, pp.107–135.