

# Pragmatic Reasoning, Defaults and Discourse Structure

Nicholas Asher and Madison Williams  
Department of Philosophy,  
University of Texas at Austin,  
`nasher@la.utexas.edu`

August 23, 2003

## 1 Introduction

In this paper we investigate the rational foundations of pragmatic reasoning. For specificity, we'll use a particular theory of discourse interpretation that combines an account of rhetorical structure with dynamic semantics. In particular we argue for a rich, *linguistic* notion of discourse context which we compute by means of simple defeasible rules in a nonmonotonic propositional logic. These defaults are linguistic in nature but have access to nonlinguistic domains of world knowledge that are encoded in a more complex nonmonotonic logic. Our approach takes as basic the idea that compositional and lexical semantics produce an underspecified logical form and that there's a level of *linguistic* pragmatic reasoning that resolves underspecifications where possible, thus producing a more complete logical form for interpretation. We argue here that this assumption is reasonable from the standpoint of a general theory of rationality and we investigate particular justifications for the various pragmatic principles that the theory adopts using a variety of game theoretic techniques.

## 2 The Problem

The problem we want to concentrate on has to do with the very nature of pragmatics. Many pragmatic phenomena—e.g., the preferred resolution of anaphoric expressions or implicatures—depend on many factors that can't easily be analyzed in a modular fashion. For example, if we consider a standard implicature like *some* implicates *not all*, it is well known that various contextual factors defeat this inference—e.g., the rhetorical function of the existential clause, background knowledge or other explicitly given information in the discourse. This contrasts with phenomena analyzed by compositional and (some varieties of) lexical semantics; for instance, the meaning of a quantifier like *most* is something that can be stated independently of the context and of the meanings of its arguments—not so for many pragmatic phenomena, indeed all the pragmatic phenomena with which a theory like SDRT is concerned. The analysis of pragmatic content and reasoning must use laws that are true proximally and for the most part, which we think are suitably formalized in some form of nonmonotonic logic—viz. as default rules in default logic (e.g., Reither 1980) or as conditionals that only defeasibly support modus ponens in a modal, nonmonotonic logic (e.g., Asher and Morreau 1991). Drawing interesting consequences from such defaults in nonmonotonic logic makes use of a notion of nonmonotonic consequence, according to which, roughly, conclusions follow from premises in all the “normal cases.” However that notion is ultimately to be analyzed (and it's immaterial for our purposes here how it is spelled out), nonmonotonic consequence makes the justification of individual axioms not an easy matter. For many defaults in a nonmonotonic logic typically interact together to define what are the relevant “normal cases” for any defeasible generalization in the theory. This is especially true in a system like that of Asher and Morreau (1991) in which the consequence relation is sensitive to the logical specificity of the antecedents of conditional defaults when these defaults conflict. By this we mean that if we have a theory  $T$  in which we have two defaults (read  $A > C$  as *if A then normally C*):

- $A > C$
- $(A \wedge B) > D$

and  $\{C \wedge D\} \cup T$  is classically inconsistent, then:

- $T, A \wedge B \not\approx D$

In such a system a normal case for the first default must be one in which we have  $A \wedge \neg B$ . That is, the second default tells us in fact what the normal cases are for the first default. We can imagine a theory with a set of ever more specific defaults, thus making the definition of what a normal case is for any one of the defaults quite complex and dependent on the way the theory analyzes other contextual parameters besides those actually stated in the default. So the empirical adequacy of one default may not be calculable in isolation from any of the others. Such pragmatic rules have at best a very indirect connection with the data they are supposed to secure.

Were such default theories merely instrumental and the individual generalizations therein played no explanatory role except to generate certain implications, in our case components of discourse interpretations, then it might not matter which individual defaults we chose as long as the whole theory behaved according to expectations. But recall that pragmatic reasoning is at least affected by many sources of information—including general world knowledge. So if all of pragmatic reasoning is really just instrumental, then we can't isolate a linguistic pragmatics out of one big Quinean interconnected set of beliefs about the world (Hobbs et al. 1993). We think that picture is wrong and that there is a separate and interesting theory of linguistic pragmatics. So we think there ought to be a rational basis for which defaults speakers assume. Without such a basis, it would also be difficult to understand how a linguistic community would come to adopt the pragmatic principles they do, as they're not explicitly taught or even part of our conscious awareness when we interpret discourse.

With some pragmatic reasoning principles, we can use standard techniques in game theory to show that these principles constitute a game theoretic optimal equilibrium, conditional upon certain assumptions (Asher, Sher and Williams 2002). But in general we'll argue here that traditional game theoretic techniques can't help us; the problem of converging on a common set of defaults for pragmatic reasoning translates into a game theoretic problem of incomplete information, for which standard techniques don't work. While some people like Skyrms (1996) or van Rooy (in press) have used standard game theoretic techniques to justify general signalling conventions or particular pragmatic principles, we don't think that such justifications can be given in general for many rules of pragmatic reasoning. Almost all such justifications are dependent upon certain descriptions of the initial problem. Rather than a cause for despair, though, we think this descriptive sensitivity can serve to our advantage. In particular this suggests that our pragmatic

defaults are quite finely attuned linguistically; further, investigations into focal points by, e.g., Bacharach and Bernasconi (1997) can help us understand in a more precise way coordination in the absence of optimization.

We'll proceed by laying out a background for a particular theory of the pragmatic/semantic interface, SDRT, and then go into details about the nature of the pragmatic rules in that theory. Though our discussion will be specific to SDRT we think many of the points we make about pragmatic reasoning should carry over to other frameworks. We'll then introduce the problem of justifying pragmatic rules game theoretically in a traditional way. Finally, we'll introduce focal point theory and examine how this helps us justify particular default rules of SDRT.

### 3 SDRT Background

The general problem we posed in the previous section is very abstract and needs to be made more concrete. Just what sort of pragmatic reasoning do we have in mind? This section provides an answer.

Over a decade of research has made it quite plausible that a theory of discourse interpretation that combines an account of rhetorical structure with dynamic semantics has linguistically and philosophically interesting contributions to make concerning various contributions discourse structure makes to discourse content. Particular areas of application for one such theory, SDRT, have included the spatiotemporal structure of texts (e.g., Lascarides and Asher 1993, Asher and Bras 1994, Asher et al. 1996, Wu 2003), pronominal anaphora (Asher 1993), VP ellipsis (e.g., Asher 1993, Asher, Hardt and Busquets 1997, Hardt and Asher 1998), presupposition (Asher and Lascarides 1998), lexical disambiguation (Asher and Lascarides 1995), plural quantification (Asher and Wang 2003, Wang, Mcready and Asher 2003), modal subordination (Asher 2002, Mcready 2003), speech acts and implicatures (Asher and Lascarides 1998b, 2001, Asher 1999). By integrating dynamic semantics' attention to semantic detail and the syntax/semantics interface with a broader AI vision of what discourse interpretation should try to accomplish suitably formalized with the help of a nonmonotonic logic, SDRT has contributed to the analysis of various semantic and pragmatic phenomena.

Fundamental to SDRT is a view of the pragmatics/semantics interface that has at least some currency in other pragmatic theories: pragmatics supple-

ments semantic content in discourse interpretation. SDRT implements this idea by having lexical and compositional semantics provide an underspecified logical form (ULF) within which certain metalinguistic variables require pragmatic information to fill them in so as to produce a fully specified logical form. ULFs describe in effect sets of completely specified logical forms; pragmatic reasoning then determines a set of preferred fully specified logical forms, and their interpretation yields the linguistic content a discourse conveys to a competent interpreter.<sup>1</sup>

Another element on which SDRT and many approaches to the pragmatic/semantic interface agree is that the pragmatic supplementation of semantic meaning exploits a notion of context. There are many different notions of context: Kaplanian contexts, Stalnakerian contexts, dynamic semantic contexts. Kaplan's notion of context is perhaps the simplest in that it only fixes values for deictic and indexical expressions. It can be understood as a fixed, background assignment to a particular set of variables or discourse referents. More complex are dynamic notions of context, according to which the context changes as discourse proceeds. Stalnakerian contexts and dynamic semantic contexts are examples of contexts of this kind. Dynamic semantic contexts are typically formalized as assignment functions to discourse referents that various linguistic devices can reset or extend; each new sentence in a discourse then defines a relation between an input and an output assignment. The interpretation of logical operators and quantifiers yield constraints as to which values of assignments pass on and which remain within a "local" environment, and this in turn provides semantic constraints on what values anaphoric expressions can pick up.

While dynamic semantic contexts have proven to have many uses, by themselves they cannot do justice to the sort of phenomena that have concerned SDRT. For example, purely compositional and lexical semantic constraints simply don't suffice in many languages to fix the temporal structure of discourse (Lascarides and Asher 1993, Asher and Bras 1994, Wu 2003). In addition dynamic semantic contexts are not equipped to account for the resolution of ambiguities of rhetorical connection, which SDRT has made a central concept in the determination of many aspects of linguistic content. Thus, SDRT uses a much more complex representation of context. Each clause yields a labelled ULF that must attach to an available attach-

---

<sup>1</sup>We mean to separate out linguistic content from speaker content. The former is what every competent speaker can glean from a discourse, while the latter depends not only on linguistic content but on nonlinguistic background beliefs of the interpreter's concerning the speaker or other elements of the context.

ment point in the discourse context which is a graph consisting of labels of previously processed clauses linked by different rhetorical relations. The choice of rhetorical relation determines which of the labels in the graph are available attachment points for new information. Once an attachment point is chosen, we have to compute a means of attachment—i.e., we have to figure out what the rhetorical role of the given clause is, if we can. The update of the contextually given graph with the new information yields a new graph. These graphs, known as SDRSs, translate naturally into formulas that have dynamic semantic interpretations, but the main point of SDRSs is that they constrain and help us calculate rhetorical connections and those in turn help determine many aspects of the preferred interpretation of discourses. In effect, SDRT refines the clause linking transition ‘;’ (dynamic conjunction) of dynamic predicate logic into a variety of different transitions, each of which is a different discourse relation. Examples of these relations are: Explanation, Elaboration, Narration, Commentary, Contrast, Parallel, Question-Answer-Pair (QAP).

Different types of rules combine to yield inferences concerning rhetorical connections. Sometimes lexical semantics determines the matter, e.g. when a discourse particle (Knott 1995) is present:<sup>2</sup>

- (1) John fell. And then Mary pushed him.
- (2) John fell. Mary pushed him.

This minimal pair shows that while we prefer to interpret the second clause in the absence of any discourse connector as providing an explanation for why John fell, the presence of the phrase *and then* forces us to connect the two clauses in a way that we have a narrative sequence of two events. This difference in rhetorical connection also affects the temporal interpretation of the two clauses.

While such discourse connectors are very helpful when present, it is nevertheless clear that in the absence of such particles, competent speakers also largely agree on the rhetorical connection between clauses in many cases. SDRT takes the position that this reasoning is a matter of linguistic mastery; it’s part and parcel of constructing a logical form for a discourse. The way SDRT is set up, this reasoning has the task of filling in metalinguistic

---

<sup>2</sup>Other examples of English discourse particles are *too*, *but*, *because*, *as a result*, *then*, *etc.*.

variables in logical form. The question is, what are the axioms or postulates that such reasoning makes use of?

SDRT uses several distinct types of rules or axioms for computing discourse structure. First there are defeasible rules for computing discourse relations that look like (3), where  $\alpha$ ,  $\beta$  and  $\lambda$  are metavariables over SDRS-labels:

$$(3) \quad (?(\alpha, \beta, \lambda) \wedge Info(\alpha, \beta, \lambda)) > R(\alpha, \beta, \lambda)$$

In words, if  $\beta$  is to be attached to  $\alpha$  with a rhetorical relation as yet unspecified (hence marked with a '?') and the result is labelled  $\lambda$ , and information  $Info(\alpha, \beta, \lambda)$  about  $\alpha$ ,  $\beta$  and  $\lambda$ , that represents (the appropriate) information from sources like the lexicon, domain knowledge and cognitive states holds, then normally, the rhetorical connection is  $R$ . These rules allow us to reason defeasibly about inferring a particular discourse relation given a suitable attachment point. To choose among attachment sites, SDRT employs a monotonic rule that picks from all possible updates of a given SDRS with new information those that maximize the strength of rhetorical relations globally and minimize underspecifications; this rule, **Maximize Discourse Coherence** (MDC), maximizes the informativeness of a given message.

## 4 A Closer Look at Defaults in SDRT

The rules that SDRT has developed enjoy some empirical support as they capture intended rhetorical relations in many examples in a precise way. But our question here is whether they have any deeper and more particular justification. Asher and Lascarides (2003) show that some of the instances of (3) derive from other defeasible rules linking a speaker's intentions and beliefs with his utterances that form a simplified cognitive model of the speaker used to infer Gricean implicatures (Asher 1999, Asher and Lascarides 2003) and other pragmatic information. And at least some of these principles of cognitive models in turn find a basis in a broader, game theoretic view of rationality (Asher, Sher and Williams 2002). To give a feel for the sort of argument we gave, we formalized two Grice-like constraints about belief as:<sup>3</sup>

- Competence (for the hearer):  $\mathcal{B}_S\phi > \mathcal{B}_H\phi$

---

<sup>3</sup>In Asher and Lascarides (2003), we actually give Sincerity a slightly different formulation, but the difference is immaterial here.

- Sincerity (for the speaker):  $Says_S\phi > \mathcal{B}_S\phi$

Given certain assumptions, Asher, Sher and Williams (2002) show that these defaults describe a unique, strict Nash Equilibrium for speakers and hearers in a particular game represented by  $4 \times 4$  matrix in which the simple strategies for hearer (believe the speaker) and the speaker (tell the truth) depend on a parameter  $\alpha$ , which stands for the proposition that it is in  $S$ 's interest to tell the truth in his current situation. The basic strategies for the speaker, for instance, are: tell the truth given  $\alpha$  and tell the truth given  $\neg\alpha$ ; lie given  $\alpha$  and lie given  $\neg\alpha$ ; tell the truth given  $\alpha$  and lie given  $\neg\alpha$ ; and lie given  $\alpha$  and tell the truth given  $\neg\alpha$ . Provided we make quite precise assumptions about the costs of complexity and the probability of  $\alpha$ ,<sup>4</sup> we were able to show that the simple defaults Competence and Sincerity would be preferable to more complex defaults that take factors like the probability of  $\alpha$  explicitly into account. We also showed a similar result within an evolutionary game theoretic framework: we showed that a population will converge on the population state corresponding to our two defaults from any point in the interior of the game space, provided we make certain minimal assumptions about payoff structure that most dynamics for evolutionary games obey and our assumption about costs and the probability of  $\alpha$  hold.

Sincerity and Competence offer a particularly nice symmetry and set of alternative strategies that a straightforward game theoretic justification can exploit. A similar sort of argument applies to defaults that characterize cooperativity about intentions. Those few rules, one of which we give below, for computing rhetorical relations that we can derive from the principles of cognitive modelling like Sincerity and Competence and from common-sense principles about, e.g., the goals people have when they ask questions (because typically they want to know answers to them) have a secure and rational basis (for the derivations of these rules see Asher and Lascarides 2003). Below  $IQAP(\alpha, \beta, l)$  says that the formula  $IQAP(\alpha, \beta)$  holds in the constituent labelled  $l$  and  $IQAP(\alpha, \beta)$  says that  $\beta$  provides sufficient infor-

<sup>4</sup>To be completely precise, the assumption is that:

$$p(\alpha) > \max\{1/2, 1 - \frac{C(T, F) - C(T, T)}{\bar{u}(-\alpha, F) - \bar{u}(-\alpha, T)}\} \quad (4)$$

where  $p(\alpha)$  is the probability of  $\alpha$ ,  $C(T, F)$  is the cost of the more complex strategy of telling the truth when it's advantageous to doing so for the speaker  $S$  and lying when it's advantageous to do so,  $C(T, T)$  is the cost of the simple tell the truth always strategy, and  $\bar{u}(-\alpha, F)$  and  $\bar{u}(-\alpha, T)$  are the utilities of lying when it is in  $S$ 's interest to lie and of telling the truth when it is in  $S$ 's interest to lie.



mation to derive a direct answer to  $\alpha$  in the discourse context. The formula  $\alpha :?$  says that  $\alpha$  has an interrogative form.

- IQAP:  $(?(\alpha, \beta, l) \wedge \alpha :?) > IQAP(\alpha, \beta, l)$

For our other rules, our argument is more involved. First, we argue that the general character of the system of defaults for computing rhetorical relations is preferable on rational grounds. Second, we investigate the grounds for adopting particular defaults.

A striking feature of our rules for rhetorical relation and attachment site computation is their simplicity. Their simplicity plays a big factor in their optimality. They are used solely to fill in underspecified elements of the ULF and are expressed within a quantifier free language whose *nonmonotonic* consequence relation is decidable (see Asher and Lascarides 2003 for details). Further, the logic for computing discourse structure has only *restricted* access to the information within other information sources and their associated logics; for example, it has access only to *descriptions* of formulae in the logic of information content, but not to what those formulae entail (in the dynamic logic where logical forms for discourse are interpreted). The same goes for other information sources like world knowledge or the lexicon. This also means that our axioms will be simpler than what one might expect if one was looking for a formalization of general world knowledge.

In contrast, many AI approaches to discourse interpretation (e.g., Hobbs et al. 1993) emphasize the role of commonsense reasoning with non-linguistic information such as domain knowledge and cognitive states. But they make heavy use of complex formalizations of commonsense reasoning. Hobbs et al. (1993)'s reliance on general commonsense reasoning leads to a highly unmodular architecture, which misses certain generalizations. For example, one cannot express within Hobbs et al.'s system of weighted abduction, where information from grammar and from domain knowledge isn't distinguished at all, that the preferences for interpreting pronouns predicted by Centering Theory are overridden if the semantics of the rhetorical relation inferred from other information sources conflicts with them (see Stone and Thomason (2000) for motivation of such a rule). So not distinguishing a linguistic level of pragmatic reasoning will miss potentially important linguistic generalizations.

But we think that the simple and independent character of rules for specifying logical forms is also important for other reasons. First, we think that

reasoning about logical form, and about what a particular discourse conveys, should be a relatively simple matter, as opposed to evaluating what was conveyed, which may be very complex indeed. We're after what every competent interpreter of the language can glean as the basic message of a text, a message that can then be evaluated, integrated into the interpreter's beliefs, into a detailed cognitive model of the speaker, or even various forms of literary interpretation. We think this must be the case if linguistic signalling is to have an optimal value. In fact formal theories of signalling in game theory presuppose that signals should be *transparent* in the sense that the receiver must eventually be able to recover the message she is supposed to recover (the one that is intended or that is evolutionarily selected for). All known biological signalling systems other than language are designed this way, and language as a product of biological evolution should be no different.

Another way to think about the simplicity requirement is decision theoretically. Suppose we have two axiomatic systems, one of which,  $A_1$  uses a logic of worst case complexity that is higher than the other  $A_2$  (as in the case of weighted first order abduction vs. a propositional nonmonotonic logic). And suppose further that compositional and lexical semantics determine, as SDRT claims they do, a ULF with underspecified elements to be resolved by either  $A_1$  or  $A_2$ . Then at least as far as resolving these underspecified elements, if we can reliably estimate the probability that using  $A_1$  will have higher costs than using  $A_2$  given contextual information  $C$ —i.e. assign a value to  $P(c(A_1) > c(A_2)|C)$  such that  $P(c(A_1) > c(A_2)|C) > .5$ , then any payoff function that is monotonic with respect to costs will lead us to choose  $A_2$  as a more optimal than  $A_1$ . Moreover, we think this assumption about payoffs is completely reasonable when informativeness is not an issue. And informativeness is not an issue here since we have assumed that both  $A_1$  and  $A_2$  are performing the same task. To handle accuracy, we can mimic our earlier results for Charity and Sincerity by setting the parameter  $\alpha$  to  $A_2$ 's producing the correct result. Of course  $A_1$  may have other uses such as providing results of general commonsense reasoning. But there is no need to use a powerful tool for a simple task, and rationality dictates a simple logic for a simple task, provided the probability of  $\alpha$  is high enough (though what it will be will depend on the exact details of the payoff function).

The simplicity argument takes us still further in a way that Mill foresaw in his defense of rule utilitarianism over act utilitarianism. One might think of using decision theory itself to infer the rhetorical connections and attach-

ment points, as well as resolving other underspecified elements in logical form by assigning probabilities as well as assigning a payoff to each possible resolution. We don't think there's anything wrong with this, but once again this seems more complicated than need be. Our defaults for calculating rhetorical connections are simple rules of thumb that abstract away from the calculation of optimal choice, though MDC is very close to such an optimization. They sometimes go wrong in that they lead us to choose a rhetorical connection that isn't the one the speaker has in mind as in:

- (5)     a.   John went to jail.
  - b.   He embezzled the company funds.
  - b'.  because he embezzled the company funds.
- 
- (6)     a.   John went to jail.
  - b.   He embezzled the company funds.
  - c.   But that's not why he went to jail.
  - d.   Rather, he was convicted of tax fraud.

As we see from (5), it is appropriate to draw the inference that (5b) explains (5a). This is not the correct discourse relation for (6b) as we find out from (6c). But the fact that this relation is inferred and then revised explains the appropriateness of *but* in (6),

So far we've argued that our simple rules have a firmer basis than more complex but equivalent rules. But we're a long way from justifying the particular default rules we've adopted. The easier case concerns our maximization rule, MDC, that governs the choice of attachment site and overall rhetorical structure. MDC maximizes the strength of the rhetorical connection, minimizes the number of overall underspecified elements and nodes after update. All of these tasks are needed to get the right rhetorical structures, as argued, e.g., in Asher and Lascarides (2003). But they appear to have rather separate justifications. Minimizing underspecifications and maximizing the strength of the rhetorical connections between constituents globally is an optimal strategy for interpretation if our utility function in fact maximizes for informativeness and coherence. That this should be so follows from a charitable attitude to the speaker's communicative intentions, but it does have its potential costs: for instance, it increases the chance for error in getting the intended message. Nevertheless such maximization and

minimization strategies are an established part of the Neo-Gricean approach to pragmatics. To wit, consider Horn’s *Q and I Implicatures*:<sup>5</sup>

- Q: Say as much as you can (given I)
- I: Say no more than you must (given Q)

The maximization of rhetorical relations and resolution of underspecified elements in MDC is a more precise version of a maximization principle than the Q principle. The other component of MDC, minimizing the graph (provided this introduces no inconsistencies or pragmatic conflicts into the message—again see Asher and Lascarides 2003 for details) is a precise and more limited version of an I type principle: the minimization of the graph size means that we’re trying to include the same semantic content within as small an informational structure as possible.

MDC and the other maxi-min principles like the Q and I implicatures seem decision theoretically sound principles for the speaker to use, provided our basic view of informativeness and cost are right. And if the interpreter assumes that the speaker is a decision-theoretic maximizer, then that’s all we need to ground a principle like MDC. No application of game theory appears to be needed (*contra* van Rooy, to appear). But if such maxi-min principles are more or less standard fare in pragmatics, our default rules for calculating rhetorical connections are not. There isn’t any obvious direct, decision- or game-theoretic justification of our particular defaults that we can see. We’ve given general arguments that establish the relative superiority of a system of simple, reasonably accurate defaults. But we can imagine lots of equivalent equally simple systems of defaults. The puzzle is how we coordinate on a particular set of defaults in a game with indeterminately many default theories (each understood as a strategy).

---

<sup>5</sup>Q implicatures are used to describe scalar implicatures of various kinds, as in (7ab); the Q principle together with (7b) taken as an answer to (7a) implies that not all the students came to the party. I implicatures, on the other hand, find applications in implicatures involving logical connectives or bridging (7cd); we note the implicatures after the  $\longrightarrow$ .

- (7)
- a. How many students came to the party?
  - b. B: Some did.
  - c. If you mow the lawn, I’ll give you \$5  $\longrightarrow$  If you don’t mow the lawn, I won’t give you \$5.
  - d. John unpacked the picnic. The beer (part of the picnic) was ice cold.

The basic problem with game theoretic accounts of coordination is that standard game theory simply tells us that in a coordination game, any coordination is a good one: all the points on the diagonal of a game in which the common alternative strategies define the rows and columns are Nash equilibria, and standard game theory tells us nothing about which of these points are to be selected. While standard game theoretic techniques tell us why it's important to coordinate, it does not tell us which coordination point to select. The experimental evidence, however, and work since the pioneering books of Shelling (1960) and Lewis (1969) have shown that coordination is far from random in many coordination games. Agents coordinate on what Shelling calls "focal points," equilibria that are particularly salient. The data are striking and show at least that standard game theoretic techniques need supplementation if we are to explain how rational agents behave in such situations. This has a special importance when it comes to the linguistic conventions that we adopt. Standard game theory has nothing to say about why we adopt the conventions we do.

That conventions are arbitrary might be right for the basic lexical meanings. There may be no story, for instance, about why 'red' should pick out the property of being red. The wide variation in such basic semantic conventions across the world's languages suggests that our basic semantic conventions are arbitrary. However, things are otherwise with pragmatic defaults. For one thing, the rhetorical relations of Narration, Elaboration, Explanation, Background, Result, Commentary, Parallel and Contrast are relations relevant to discourse interpretation in many languages, as studies in RST and SDRT have shown. No doubt we could refine these discourse relations into a more finely individuated set, but they do appear to be universal features. Further, the defaults that enable us to infer these relations also appear to work in many languages besides English, though there is some variation because of the interpretive work that other elements in a language perform. For instance, in English the defaults for Explanation work when the past tenses of the main verbs in both constituents are the same—whatever the past tense, whereas this is not true for French. In French Explanation cannot link two clauses in the *passé simple*, though it can if the tense in both main verbs is the *passé composé*. This uniformity deserves an explanation. So what we want to do is to consider why a particular default or set of defaults for a particular relation, say Explanation, might become a convention—something that speakers of a language will coordinate upon. To do this we need to take a detour into a theory that enables us to pick between equilibria. This theory is known as Variable Frame theory.

## 5 Bacharach's Variable Frame Theory

Bacharach's (1993) Variable Frame Theory (VFT) offers us a way to re-describe a game by forming equivalence classes of actions in the original game. Such a redescription often favors one equilibrium, a focal point, over others. Formally, players in VFT have a set  $A$  of objects and a set  $C$  of families "about"  $A$ , where each  $c \in C$  is a set of descriptions of the objects in  $A$  such that  $c$  induces a partition of the objects in  $A$ ,  $P_c$ . A frame is a subset of  $C$ , and  $F = \mathcal{P}(C)$  is the set of all frames. Each frame  $f \in F$  thus induces a partition  $P_f$  of  $A$  that identifies each  $a \in A$  with a cell in the partition resulting from the pointwise intersection of all (available) families. The *availability* of a frame,  $\nu(f)$ , is a probability measure over frames; we'll take it to be the distribution of frames among the population of players. In VFT nature assigns each player a frame. This is analogous to an assignment of a type in a game of incomplete information, in keeping with Harsanyi's (1967-8) treatment of such games. Players' options  $o$  in frame  $f$ , are given as the cells of  $P_f$  where  $E(o)$  is the extension of objects in this cell. If there is only one object, choosing option  $o$  implies choosing that object. If there is more than one ( $|E(o)| > 1$ ), it implies randomizing over those objects. Let  $opt_f$  be the set of such options.

If  $f$  is a frame in  $F$ , let  $F(f)$  be the set of all subframes of  $f$  where a subframe of  $f$  is a frame that takes only a subset of the descriptions of  $f$  (for example color, shape is a subframe of a frame that takes color, shape, size into account). Let  $opt_{F(f)}$  be the options defined by the subframes of  $f$ , so  $opt_{F(f)} = \cup_{f' \in F(f)} opt_{f'}$ , and let  $Opt = \cup_{f \in F} opt_{F(f)}$ . Intuitively  $Opt$  is the set of all frames available for consideration.

The payoff function is a function  $U : Opt \times Opt \rightarrow \mathcal{R}$ , defined by

$$U(o, o') = \sum_{a \in E(o)} \sum_{a' \in E(o')} u(a, a') |E(o)|^{-1} |E(o')|^{-1}$$

where  $u : A \times A \rightarrow \mathcal{R}$  is the payoff structure given by the game. A decision rule is a function  $\phi : F \rightarrow opt_{F(f)}$

Define a belief function  $\pi(f'|f)$  as the probability that the other player has frame  $f'$  given that the player has frame  $f$ . No player can put positive weight on any frame not a subframe of her own. If a player were aware of such a possibility, it would thereby be a part of her own frame. Thus, the

expected utility of each option  $o \in \text{opt}_{F(f)}$  is given by

$$\mathcal{EU}(o|\phi, f) = \sum_{f' \in F(f)} U(o, \phi(f')) \pi(f'|f).$$

Here's a simple example of a game with two players, 1 and 2, where VFT produces interesting results. The players are presented with four shapes, a square and three triangles. There are 4 actions: choose shape 1 (the square) or shape 2, 3, or 4 (the triangles). Players get a payoff of 1 each if they agree on the shape, 0 otherwise. This yields the following game:

	1	2	3	4
1	1,1	0,0	0,0	0,0
2	0,0	1,1	0,0	0,0
3	0,0	0,0	1,1	0,0
4	0,0	0,0	0,0	1,1

The equilibria are coordinations along the diagonal (and mixtures thereof) where no pure strategy is preferable to any other (in this discussion we are only interested in pure equilibria). However, if we use VFT, with a frame that takes shape into account the players now have an option set of randomizing over all shapes  $s$  (the trivial partition), randomizing over the triangles,  $t$ , and picking the square,  $q$ . This yields the game

	s	t	q
s	0.25,0.25	0.25,0.25	0.25,0.25
t	0.25,0.25	0.33,0.33	0,0
q	0.25,0.25	0,0	1,1

This game has the pareto-optimal coordination point of both players selecting the square.

## 6 Dynamic Refinement of VFT

Consider the following example from Janssen (2001). There are five objects differentiated only by their color. Two are red, one is blue, one is green, and one is yellow. Players get a utility of one if they coordinate on an object, nothing otherwise. As before, in the standard game theoretic account, there are five equally good pure strategy equilibria, with nothing to distinguish them. However, if we use a frame that distinguishes colors we find the following game:

	c	r	b	y	g
c	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$
r	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{2}, \frac{1}{2}$	0,0	0,0	0,0
b	$\frac{1}{5}, \frac{1}{5}$	0,0	1,1	0,0	0,0
y	$\frac{1}{5}, \frac{1}{5}$	0,0	0,0	1,1	0,0
g	$\frac{1}{5}, \frac{1}{5}$	0,0	0,0	0,0	1,1

Notice in this case there are three equilibria ((g,g), (b,b), and (y,y)) that are indistinguishable. Janssen’s solution is to offer a refinement that lumps together all “payoff symmetric” solutions. Since the three solutions are indistinguishable, they are treated as equally likely to be used. This gives the result that (r,r), which offers a 50-50 shot at coordination is preferable. Note that this is equivalent to using the subframe descriptions red and non-red. Bacharach and Bernasconi (1997) expected this result, but it is not in line with the empirical results they found. While a strategy coordinating on red makes sense in a one-shot game, we think it misses the fundamental role focal points play, which is to aid the discovery of conventions that are useful over the long term, using information available from experience.

Our argument that focal points are essentially a dynamic tool comes mainly from the realization that the discovery of useful descriptions is not performed on the spot, but is rather the result of background world knowledge. Humans have a considerable amount of experience in interacting with others. Focal points provide a means of translating this experience into a useful tool for facilitating their interactions. Interactions of the sort described in the games played here add to this background knowledge, increasing their future



usefulness. An attempt to model this phenomenon in a static setting is terribly misguided. Any participants in such a game would bring a great deal of experience and knowledge about what other players are likely to notice and what they are likely to do.

The problem, then, with the sort of theory described by Bacharach and Berasconi, and by Janssen is their expectation that play will conform to the best play for the short term (picking red). However, in this case short-term and long-term viability of the solution diverge. As the repetitions rise above 5 in this example, the expected utility of picking green, blue or yellow would be better than coordinating on red from the beginning. This is because staying with the red picking equilibrium would never be better than coordinating half the time. While it would take longer to find a coordination on another color, once it was found, players would be able to coordinate all the time by keeping this color. Furthermore, with positive results confirming that the other player is picking red, one would not be able to escape from this coordination. Thus, in the long-term, a red picking strategy would be worse.

In empirical studies Bacharach and Berasconi observed other behavior that's surprising to a proponent of VFT. In a game asking people to coordinate on shape from a choice of two circles, a square, a diamond, and a triangle. Bacharach and Berasconi found a majority of participants picking the triangle, though they had predicted the circle. We feel this is because, even in a one-shot game, people are looking for long term conventions better served by a non-circle shape, althopugh far more empirical study is needed to substantiate this intuition.

Finally, although experimental efforts attempt to correct for this, with human subjects the number of possible frames is far too large. Even when deliberately attempting to limit frames to particular attributes in presenting the game to players, Bacharach and Berasconi found a part played by "oddities", nuisance descriptions that entered the self-reported frames of the participants. The use of focal points in normal interaction would, of course, leave this possibility wide open. There would be too many possible frames to consider for focal points to be a useful tool if they were simply a static concept, to be discarded when play finished rather than remembered for use in aiding future cooperation.

In order to account for these aspects of focal points, we propose a dynamic extension to VFT. Instead of a one-shot game where players must calculate

the type of the other player (which corresponds to the frames used), we offer a repeated game which gives players a chance to learn the type by observing the play of the other players. We, nevertheless, agree that only subframes are available for consideration. This allows for the possibility that one player may not notice the feature on which the other is conditioning, though repeated play will allow for noticing it later. Even if the frames are distinct (say one player notices shape but not color, the other notices color but not shape) as long as the options selected overlap, allowing for some coordination, players will be better off than if they were choosing at random, which is not the case with VFT.

In Dynamic Variable Frame Theory (DVFT), games will be played as stage games, where each stage is a VFT game. We give each game iteration two stages: an analysis stage and a playing stage. In the first analysis stage,  $A_1$ , players consider frames that appear according to availability. Then they anticipate the other player's frames (always a subset of their own) with some (possibly empty) noticer bias (the other player is more likely to notice what I have noticed). Also players will form an expectation for the results of the playing phase based on beliefs about the frames being used (unlike Harsanyi's approach, neither DVFT nor VFT assumes common knowledge of the player types). They use these frames to construct a VFT type game for the first playing stage,  $P_1$ . Then in  $A_n$  for  $n > 1$  they adjust the probability of the other player using this strategy. If results of  $P_{n-1}$  meet their expectations, it will confirm the initial hypothesis that the other player is using a similar strategy. If not, the probability of this frame will drop. Probabilities will be updated in a typical Bayesian hypothesis testing way. Also in  $A_n$ , new frames may arise according to their availability, though none will ever entirely go away. Note that new strategies will be automatically updated according to the other player's past play. If what is known is inconsistent with this new frame, the new frame will start out judged less likely than otherwise. Strategies shouldn't go away entirely just because of inconsistent play, especially if no equilibrium has been reached: the other player might switch to the frame option that just failed. The existence of the trivial strategy that randomizes over all possibilities, as well as the availability implying that this strategy might be newly discovered by the other player, will keep it a viable option. Also, any coordination should boost the probability of all consistent frames by the likelihood of meeting there. So if players coordinate on the blue patch, they should play blue again rather than assuming this coordination is due to randomizing over all

strategies.<sup>6</sup> The information from  $A_n$  is then used to play  $P_n$ .

Formally, we define the terminal node,  $z_n$  as the result of play in each stage  $n$ . Since each player knows the terminal node reached at in each round,  $h_t$  is the sequence of terminal nodes  $(z_1, z_2, \dots, z_{t-1})$  reached in the first  $t - 1$  rounds of play, and  $h_\infty$  is a complete history i.e. history at time  $\infty$ , and if fixed,  $h_t$  are the first  $t$  results in this history, while  $\mu_t^i(h_t)$  is a belief rule, a function from histories to player  $i$ 's beliefs at time  $t$ , given history  $h_t$ . If we let  $\hat{H}(h_\infty)$  be the information sets reached with positive probability along history  $h_\infty$ , and the empirical distribution of play at information set  $h^j$  be  $d(h^j|h_t)$ .

Now we turn to the updating the availability function in each round. After the initial move by nature where  $f$  is the frame assigned by nature and  $f'$  is a subframe, as in VFT, players have the belief function  $\pi(f'|f)$ . After observing the results of each round of play, players must take the following into account:

1. The consistency of the results  $(h_t)$  with  $\pi(f'|f)$
2. The consistency of  $h_t$  with any newly discovered frames
3. The possibility that player <sup>$j$</sup> , for  $j \neq i$ , just became aware of a frame of which player <sup>$i$</sup>  was already aware

This information is then used to update the respective belief functions for the next round of play. This is taken into account by using the belief rule

$$\mu_t^i(h_t) = P(h_t|\mu_{t-1}^i(h_{t-1})) + \frac{(t-1)!}{t!} \nu(f)(1 - \nu(f))^{t-1}$$

and  $\mu_0^i(h_0)$  the initial belief function is given by  $\pi(f'|f)$ . In words this says that you update by taking probability of the history conditional upon the result of the belief rule from the previous round, while adding the probability of the other player encountering the frame in round  $t$ . Note that this gives us a vector of probabilities, one probability for each subframe of the updated frame.

Now we can define an equilibrium concept for such games as a mixed profile  $\sigma$  such that there exist beliefs  $\mu^i$  and for each  $s^i$  in the support of  $\sigma^i$ ,

---

<sup>6</sup>Actual (human) player are unlikely to assign numerical probabilities or calculate a Bayesian update; this is a formal convenience.

$u^i(s^i|\mu^i) \geq u^i(\hat{s}^i|\mu^i)$  for all  $\hat{s}^i \in S^i$ , and  $\mu^i(P^j(\sigma^{-i}|H)) = 1$ . The first criteria requires that each player's strategy be optimal, given her beliefs. The second condition requires that the beliefs be well behaved probabilistically at every information set. This is extended to give a unitary self-confirming equilibrium as a mixed profile  $\sigma$  such that there exist beliefs  $\mu^i$  and for each  $s^i$  in the support of  $\sigma^i$ ,  $u^i(s^i|\mu^i) \geq u^i(\hat{s}^i|\mu^i)$  for all  $\hat{s}^i \in S^i$ , and  $\mu^i(P^j(\sigma^{-i}|\bar{H}(\sigma))) = 1$ . The first clause is the same, while the second requires that the players have correct beliefs for information sets actually reached along  $\sigma$ , and learn nothing about the others.

The idea, for our purposes is that self-confirming equilibria will exhibit infrequent switches (as experimentation will never pay) so it will be asymptotically myopic: players will only learn about situations where the world knowledge is used by everyone and so learn that it works and so the defaults will be inescapable. See Fudenberg and Levine (1999) p.184 proposition 6.1 and 6.3. Now we can look at equilibrium and stability concepts for dynamic extended form games. There is a convergence of players to the largest option that overlaps the other player's frame even if it's not shared by both players. To see this consider a game where player 1 is sensitive to shape but not to color and player 2 is sensitive to color but not to shape. They have to coordinate on a pick between two shaded triangles and three squares one of which is shaded. So the initial preference for each is for player 1 to randomize among the triangles, and player 2 to randomize among the unshaded shapes. This gives a payoff matrix below.

	c	d	w
s	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{5}, \frac{1}{5}$
t	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{9}, \frac{1}{9}$	$\frac{1}{3}, \frac{1}{3}$
q	$\frac{1}{5}, \frac{1}{5}$	$\frac{1}{3}, \frac{1}{3}$	0,0

Note, however, that this initial strategy never coordinates, doing far worse than chance. Even if we assume that players never become aware of the frame the other is using, they will soon realize that they are doing worse than the coordinating half the time that they expect. This will cause their beliefs in the other possibilities to increase, yielding at some point a coordination on either an unshaded square or a shaded triangle. This should increase the probabilities in whichever they happen to find first so that, for example, player 1 always plays square and player 2 always plays white. This result is equivalent to coordinating on the crudest partition; always playing square.

While player 2 may notice that she is not doing as well as she expects, she nevertheless cannot profitably switch. The other possibility is that player 2 always plays shaded and player 1 always plays triangle. In this case, player 1 notices that he isn't doing as well as he should; nevertheless, there's no more profitable strategy he can play. In this case the result is equivalent to coordinating on the crudest partition held by player 2—always playing shaded. So even if the frames never overlap at all, we are still doing better than random chance by using DVFT. This gives a formal confirmation of the results that Garrod and Anderson (1987) noticed with coordination on the Edinburgh maze games. There players started with different frames describing a situation but they were able to coordinate through repeated trials.

## 7 Extension of DVFT to Signalling Games

First we briefly describe signalling games. The basic signalling game has two players, a sender and a receiver. The sender has private information about her type,  $\theta$ , and chooses some action  $a_s \in A_s$ . Player 2 (whose type is assumed to be common knowledge) chooses some action  $a_r \in A_r$ . The spaces of mixed actions ( $\mathcal{A}_s$  and  $\mathcal{A}_r$  with elements  $\alpha_s$  and  $\alpha_r$ ) are set of mixed strategies that the sender and receiver can play. So player i's payoff is  $u_i(\alpha_s, \alpha_r, \theta)$ , and the receiver's beliefs about the sender's type are common knowledge. A strategy for the sender is a probability distribution  $\sigma_s(\cdot|\theta)$  over actions  $a_s$  for each type that the sender admits. A strategy for the receiver is a probability distribution  $\sigma_r(\cdot|a_s)$  over actions  $a_r$  for each action  $a_s$  she can receive.<sup>7</sup>

---

<sup>7</sup>This gives the sender of type  $\theta$ , playing strategy  $\sigma_s(\cdot|\theta)$  who encounters a receiver playing strategy  $\sigma_r(\cdot|a_s)$  receives payoffs of

$$u_s(\sigma_s, \sigma_r, \theta) = \sum_{a_s} \sum_{a_r} \sigma_s(a_s|\theta) \sigma_r(a_r|a_s) u_s(a_s, a_r, \theta)$$

and receiver who plays strategy  $\sigma_r(\cdot|a_s)$  when encountering  $\sigma_s(\cdot|\theta)$  receives an (*ex ante*) payoff of

$$\sum_{\theta} p(\theta) \left( \sum_{a_s} \sum_{a_r} \sigma_s(a_s|\theta) \sigma_r(a_r|a_s) u_r(a_s, a_r, \theta) \right)$$

in order to take her beliefs about the senders type into account. In particular, the receiver should update her beliefs about the sender's type given the sender's signal and base her choice on the posterior distribution  $\mu(\cdot|a_s)$ , where  $\mu(x)$  is the probability that player i assigns to event x given that she has reached some information state. This yields a payoff

A Perfect Bayesian Equilibrium of a signalling game is a strategy profile  $\sigma^*$  and posterior beliefs  $\mu(\cdot|a_s)$  such that players optimize on the conditioning that will happen: the sender takes into account that the receiver will condition her beliefs about the sender's type on the signal, while the receiver will respond optimally to the sender's signal, given her beliefs; finally, the receiver will use Bayesian updating to determine her best response to the sender.

An example of a signalling game is the “Beer-Quiche” game. In this game, player 1 is supposed to be eating lunch in a bar, player 2 is a bully who may or may not fight player 1. We assume that player 1's types are weak,  $\theta_w$  and surly,  $\theta_s$ , where the probability of  $\theta_s = .9$  and the probability of  $\theta_w = .1$ . While both types would prefer to avoid a fight, the bully would want to fight a weak type but not a surly type. Player 1's actions are to order beer and order quiche. Surly types prefer beer while weak types prefer quiche. Player 2's actions are attack and ignore. Player 2 would prefer to attack weak types and ignore surly types. Player 2 will prefer to attack if she believes that the probability that player 1 is weak is greater than 0.5. Player 2 observes player 1's meal then updates her beliefs about player 1's type, so this is the signal.

Let us briefly look at the pure strategy equilibria of such a game. If player 2 believes that it is more likely that player 1 is weak after updating with the information that player 1 has quiche (or that player 1 has beer) we will get a “pooling equilibrium” where player 1 always orders beer (quiche). Otherwise, player 1 will order beer if surly and quiche if weak.

Note that the optimal equilibrium for the signalling game is one in which the players optimize on the conditioning that will happen: the sender takes into account that the receiver will condition her beliefs about the sender's type on the signal while the receiver will respond optimally to the sender's signal given her beliefs. Clause 3 in the definition of a perfect Bayesian equilibrium requires that the receiver use Bayesian updating to determine a best response.

Adapting DVFT to signalling games should be straightforward as types will be given by frames. The idea is that there will be another stage of the game where players send a signal that suggests how they are describing the decision

of

$$\sum_{\theta} \mu(\theta|a_s) u_r(a_s, \sigma_r(\cdot|a_s), \theta) = \sum_{\theta} \sum_{a_r} \mu(\theta|a_s) \sigma_r(a_r|a_s) u_r(a_s, a_r, \theta)$$

problem. The only difference between the regular dynamic VFT and the signalling game is the addition of a signalling stage. After the signal stage we will give some (arbitrary) probability that an unnoticed frame that that is signalled will be noticed by the hearer. Then the hearer updates her belief function in the same way as above. Then the game proceeds as in DVFT.

## 8 DVFT Applied to Pragmatic Defaults

The basic idea is that defaults in SDRT for inferring discourse relations capitalize on certain aspects of the description of a dialogue. Take for instance, the basic default for Explanation in SDRT. *Top* below stands for the relation that tells us that a particular label outscopes all the other labels related to another label, while *Aspect*( $\alpha, \beta$ ) says that the aspect of  $\alpha$  and  $\beta$  may be either perfective or imperfective in the sense of introducing an event or state discourse referent into the representation.<sup>8</sup> In effect *Aspect*( $\alpha, \beta$ ) summarizes a whole host of more specialized defaults, one for each specification of the aspects of  $\alpha$  and  $\beta$ . On the other hand, the *PAST*( $\alpha$ ) and *PAST+*( $\beta$ ) tell us that the main verb of  $\alpha$  must have a past tense and that  $\beta$  must have either a past or pluperfect tense.

- **Explanation:**

$$(?(\alpha, \beta, \lambda) \wedge PAST(\alpha) \wedge PAST+(\beta) \wedge Top(\sigma, \alpha) \wedge cause_D(\sigma, \beta, \alpha) \wedge Aspect(\alpha, \beta)) > Explanation(\alpha, \beta, \lambda).$$

Of special interest here is the predicate *cause<sub>D</sub>*, which sums up a variety of factors that lead to the possibility of a causal link. The idea is that a choice of main verb and arguments for that verb sometimes get associated with a causal connection: for instance, if someone falls and that person is pushed, there is an association between these that comes from world knowledge—these events can instantiate a typical causal pattern: x pushes y and as a result y falls. Note that we *wouldn't* want to say that this typical connection is implicated simply by the use of such verbs; but these causal patterns can help determine the type of speech act of the second utterance—i.e., that is functioning as an Explanation, when the constituents are linked and other features obtain. For instance, this information is quite sensitive to the form of the linguistic description. For instance, if the main verb falls within the scope of a negation, something that the ULFs of SDRT will reflect,

---

<sup>8</sup>We follow here Kamp and Reyle's (1993) treatment of verbal aspect.

then *cause<sub>D</sub>* won't be inferred. The default also shows the inference to Explanation is sensitive to the tense forms used. This sensitivity is needed in virtue of examples like (8a-b) though not (8c) (whose analysis involves a topic for alternations and is something we can't go into here).

- (8)      a.    John didn't fall. Mary pushed him.  
           b.    John had fallen. Mary pushed him.  
           c.    John fell or slipped. Mary pushed him

It is this association that primes the interpretation of the second clause as a speech act of Explanation. Asher and Lascarides (2003) argue that a careful examination of the syntax/semantics interface and the argument structure for lexical entries sometimes enable to derive these connections from MDC, if the lexical entries contain underspecified elements that the second clause can resolve. But there are certainly many cases where we infer causal links where the evidence for complex lexical entries is not at hand.

DVFT furnishes us an explanation for why such defaults should be adopted, even in the absence of evidence from the lexicon and the syntax/semantics interface. Let's suppose that the context satisfies the antecedent of the default for Explanation. There are many other rules that are alternative possibilities. For instance, we could have a strange Narration rule.

- **Weird Narration:**  
 $(?(\alpha, \beta, \lambda) \wedge Top(\sigma, \alpha) \wedge PAST(\alpha) \wedge PAST+(\beta) \wedge cause_D(\sigma, \beta, \alpha) \wedge Aspect(\alpha, \beta)) > Narration(\alpha, \beta, \lambda).$

A natural partition of our different defaults for a situation in which the antecedent of **Explanation** holds would be the following: there are those rules with that antecedent that have a discourse relation whose semantics coheres with the causal connection associated by speakers with the information summarized by *cause<sub>D</sub>*; there are those rules with the antecedent of the Explanation rule but with a discourse relation whose semantics doesn't have anything to do with the causal link (like **Weird Narration**) and then all the other rules without the causal information in the antecedent. DVFT should prompt us to select **Explanation** since it's the only rule in the cell of the partition with the "coherent" rules; the other partition contains many rules like **Weird Narration** and the other cell contains indefinitely many rules that use irrelevant or "nuisance" information in the antecedent (we suppose



in this case that there are no cue words or discourse particles present). There's something special about Explanation in that it tracks the associated, stereotypical causal information. The idea that falling is a potential result of pushing in effect primes for the interpretation of the second speech act as an Explanation.

Different discourse situations where we need to attach new information to a given context will group naturally into those that verify *Cause<sub>D</sub>* predicates and those that verify antecedents of the other defaults, like those for Narration, Elaboration, and Result. As agents continue to play discourse games, the probabilities of the rules on which coordination succeeds will continue to increase, while those on which coordination fails or which are not selected will get lower probabilities. DVFT allows us to keep the successful frames for future coordination. Replicating the analysis of DVFT for **Explanation** in these other contexts will yield justifications for these other defaults. A one shot coordination game should suffice to coordinate on any individual default, if both players share the same partition that we gave above. What if speaker and interpreter don't share the same partition or frame? One possibility, since we're in a signalling game, is that the signal conveys itself the relevant aspects of the speaker's frame. Then coordination can proceed as before. If that is not the case, then as we saw earlier, iterating the coordination game will produce a coordination on whatever is the largest cell that overlaps some element in each player's frame. This will amount to having some sort of an approximation to our full default, one say in which only some features relevant to *Cause<sub>D</sub>* are noticed. But in any case, we think that the aspects to which our defaults are sensitive are elements that speakers will readily notice and so we expect a much better than worst case convergence here to the full defaults.

It remains to be seen whether this strategy of justification fits all the other defaults that are used in SDRT's module of pragmatic reasoning. The axiom for Background, for instance, exploits aspectual information that doesn't appear to be directly connected to the semantics for the Background relation. And it's not clear how world knowledge functions in that case. Axioms for other relations like Parallel, Contrast or even Correction exploit syntactic and semantic structure or intonational information, and that too requires more study. Nevertheless, DVFT is a very sophisticated tool, and the prospects of using it to justify individual defaults in SDRT appears good. In future work, we hope to extend this approach to handle not only other linguistic defaults but defaults in other areas.

## 9 Selected References

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

Asher N. 1999: "Discourse structure and the logic of conversation", *Current Research in the Semantics-Pragmatics Interface* 1, pp. 1-28.

Asher N., Aurnague M., Bras M., Sablayrolles P and Vieu L. 1995: "Computing the Spatio-temporal Structure of Texts", The First International Workshop on Computational Semantics, Tilburg, the Netherlands.

Asher N. and Bras M. 1994: "Le raisonnement non monotone dans la construction de la structure temporelle de textes en franais," RFIA 1994.

Asher N., Hardt D. and Busquets J. 1997,. "Discourse Parallelism, Scope, and Ellipsis", in S. Peters and H. de Swart eds. *Proceedings of SALT 7*.

Asher N. and Lascarides A. 1998a: "How Questions in Dialogue", *Linguistics and Philosophy*, 21, pp. 237-309.

Asher N. and Lascarides A. 1998b: "The Semantics and Pragmatics of Presupposition", *Journal of Semantics* 15, 1998 pp.239-299.

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

Asher N., Sher I and Williams M. 2002: 'Game Theoretic Foundations for Gricean Constraints', *Proceedings of the 2001 Amsterdam Colloquium on Formal Semantics*, Amsterdam, The Netherlands.

Asher N. and Wang L. 2003: 'Ambiguity and Anaphora with Plurals in Discourse', *SALT 13*, Seattle, Washington.

Bacharach M. 1993: 'Variable Universe Games', in K. Binmore, A. Kirman and P. Tani (eds.), *Frontiers of Game Theory*. MIT Press, Cambridge, Mass.

Bacharach M. and Bernasconi M. (1997): 'The Variable Frame Theory of Focal Points: An Experimental Study', *Games and Economic Behavior*, Volume 19, Issue 1, pp 1-45.

Fudenberg D. and Levine D. 1999: *The Theory of Learning in Games*, MIT Press.

Garrod S. and A. Anderson 1987: 'Saying What you Mean: A Study in

- Conceptual and Semantic Coordination, *Cognition* 27, pp. 181-218.
- Harsanyi J. 1967-8: 'Games with Incomplete Information Played by Bayesian Players', *Management Science*, 14: 159-82, 320-334, 486-502.
- Janssen M. 2001: 'Rationalizing Focal Points', *Theory and Decision*, 50: 119-148.
- Knott A. 1995: *A Data-Driven Methodology for Motivating a Set of Coherence Relations*, Ph.D. Thesis, University of Edinburgh.
- Lewis, D. 1969: *Convention*. Harvard University Press.
- Schelling, T. 1960: *The Strategy of Conflict*. Harvard University Press.
- Skyrms, B. 1996: *Evolution of the Social Contract*, Cambridge University Press.
- Stone M. and Thomason R., 2002: 'Context in Abductive Interpretation', EDILOG 2002: 6th workshop on the semantics and pragmatics of dialogue. The University of Edinburgh, September, 2002.
- van Rooy, R. in press: 'Signalling Games select Horn strategies', *Linguistics and Philosophy*.
- Wu J. 2003: *An Analysis of Aspectual Markers in Mandarin*, Ph. D. Dissertation, University of Texas at Austin.