Communication as a Special Case of Misunderstanding: Semantic Coordination in Dialogue

Patrick George Timothy Healey

Doctor of Philosophy
University of Edinburgh
1995

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Patrick Healey

Edinburgh, 1^{st} October 1995

For
Brian George Healey
1926–1984

Abstract

This thesis is an investigation of the theoretical and empirical consequences of idiolectical variation for the coherence of natural language dialogue. In essence, it is an elaboration of the intuition that we do not all mean the same thing by a given expression.

To the extent that semantic theories have been applied to modelling dialogue, it is argued that it has involved appeal, implicit and explicit, to a shared semantic code as the basic guarantee of mutual-intelligibility. The case is made that idiolectical variation should be understood as a semantic phenomenon, indeed may be fundamental to the normative character of semantic content, and as such it undermines models of communication that depend on the existence of such a code. A parallel problem is identified in empirical models of dialogue where there has been a tendency to concentrate on social and pragmatic factors in interactional coherence whilst remaining silent on the processes that contribute to the achievement of semantic coordination. It is argued that, again, there is often an assumption that the recognition of some code-like level of literal meaning proceeds automatically but no explanation of how this level arises. Two notable exceptions to this are the Collaborative Model of Dialogue and the Input-Output Coordination Model; however, they offer conflicting accounts of the mechanisms involved in semantic coordination: the former appeals to the pairwise establishment of mutual belief whereas the latter offers a group-based account of semantic conventions. Three experiments are reported which investigate in more detail the emergence of semantic coordination within experimental communities and provide a test of these competing explanations. The results indicate that neither of the existing accounts is fully adequate and an alternative, repair driven, model is proposed. This model allows for an arbitrary degree of idiolectical variation to obtain between individuals while still accounting for the convergence in interpretation necessary for the coordination of actions through dialogue. To provide a semantic model for this explanation a formal framework is proposed, using Channel Theory, in which both conventional cognitive and semantic analyses can be reconciled.

Acknowledgements

Without whom not: Jimmy Cuthbert and Carl Vogel. For patience far exceeding Job's: Nick Chater, Simon Garrod, Paul Schweizer and Melanie Nelson. Special thanks to Bean, infelicity smeller pursuivant, all my experimental subjects for their generous compliance and everyone at the Centre for Cognitive Science 1990-94, a place where psychologists can indulge in whatever degree of miscommunication they wish. Thanks also to SERC/EPSRC for financial support. May the sun shine brightly on you all.

Contents

Chapter 1

Communication and Natural Language

1.1 Introduction

Dialogue is the primary site of language-use in ontogeny and phylogeny. In evolutionary time, language is thought to have emerged with the Hominids, approximately 2 million years ago (Foley, 1987). By contrast, the earliest evidence of any written language is dated to approximately 10 thousand years ago (Schmidt-Besserat, 1991) and Cherry (1966) estimates that only 5% of languages have a written form. In development, children achieve communicative competence in their native language prior to engaging with the problems of becoming literate. The skills required to communicate successfully in a natural language seem to be prior to those required for negotiating the other forms of language to which we are exposed. Despite its centrality in the development of language, dialogue *per se* has received little attention from psycholinguists or semanticists (Clark, 1985). There is a tendency to assimilate dialogue to monologue under the heading 'discourse', obscuring the problems specific to the maintenance of the inter-individual

coherence of linguistic interaction. Intuitively, an important problem specific to dialogue is that individuals must attempt to coordinate meaning. Where interlocutors vary in their interpretation of words and expressions they must find some way of dealing with the threat this poses to the success of their interaction. The organising concern of this thesis is with the empirical and theoretical problems idiolectical variation poses for the mutual-intelligibility of dialogue. Empirically, the question is how mutual-intelligibility is achieved and maintained in dialogue despite idiolectical variation. Theoretically, the question is how theories of meaning, particularly those in the formal semantic tradition, can accommodate idiolectical variation.

Like much of cognitive science, what follows is an exercise in ecumenicalism, attempting to recruit findings from experimental psychology, formal semantics and philosophy of language to the exploration of this theme. Some obvious dangers attend an interdenominational approach, not only because it may offer a weaker analysis than a more focused investigation would allow, but also because it may violate some of the commitments that delineate each discipline it draws on. Sectarian antipathies notwithstanding, interdisciplinary research also promises substantial advantages, not least because it holds out the possibility of bringing a range of tools and methodologies to bear on a common problem. Whatever claim this work has on developing a coherent thesis, it is defended on the grounds that it starts from a specific problem and explores its implications in several areas in cognitive science. A more practical corrollary of the interdisciplinary approach adopted here is that no exhaustive, coherent literature review is possible. Rather than appealing for absolution, the strategy followed below is to try and deal with the relevant elements of each literature as they arise rather than corralling them all into one chapter.

One important tradition whose toes do not fall under the steps made by the following pages is that referred to by Markova (1990) as "dialogism" (see also

Leudar, 1991). In this literature, the notion of dialogue is elaborated beyond processes of face-to-face interaction to an epistemological stance on the nature of cultural and historical development as, for example, in the Hegellian notion of dialectic. Where psychological issues are raised, the emphasis is on psychology and cognition as socially constituted, rather than individualistic, phenomena. The preoccupations of cognitive science do not readily accommodate this tradition, not least because it draws extensively on phenomenology and 'modern' continental philosophy where, by contrast, the majority of work in cognitive science follows in the Anglo-American analytic tradition. It would be wrong to overplay these differences; however, they are sufficient to prevent any attempt by the ensuing discussion to span the gap between them. There are points of contact though, particularly with respect to the problems of perspectival relativity and the normative nature of meaning and, where relevant, these issues are raised.

In terms of the overall plan of this thesis, the remainder of this chapter concentrates on outlining the, often tacit, commitment of semantic theories to code based models of mutual-intelligibility. Chapter 2 explores in some detail the problems idiolectical variation poses for code-based accounts and its relation to several themes in the philosophy of language. Chapter 3 turns to the examination of a number of empirical models of dialogue and the extent to which they can address the difficulties raised for code theories. It is argued that only two proposals offer analyses which impact on questions of semantic coordination. Chapter 4 reports three experiments designed to separate the competing claims of these two models. The results of these experiments suggest neither account is entirely adequate and an alternative explanation for the mechanisms by which coordination is achieved is offered. Chapter 5 returns to the formal concerns, developing a proposal for a semantic framework which can address idiolectical variation and can provide a basic model for the phenomena observed in the experimental work. Finally, Chapter 6 attempts to draw together the arguments of the preceding chapters and discusses

their implications for the study of meaning.

1.1.1 Terminology

A practical difficulty that besets discussion of theories of meaning and comprehension is the confusion engendered by terminological clashes between different areas. Although it is impossible to detect and pre-empt all the potential conflicts it is useful to state at the outset how some terms are intended throughout the text.

For the purposes of characterising meaning in natural language a tripartite distinction between syntax, semantics and pragmatics has traditionally been imposed. Roughly, syntax characterises the structural or grammatical component of language, pragmatics characterises the factors relating to particular occasions of use and semantics characterises the aspects of meaning, sometimes referred to as literal meaning, that are understood as somehow given by a particular language independently of its deployment in particular situations. It is difficult to give an independent justification for the distinction between the contribution to meaning made by the semantic resources of a language and the contribution made by pragmatic factors. Usually, the distinction is drawn purely on theory-internal grounds. For example, Gazdar (1979) defines pragmatics as meaning minus truth conditions. There is no interest here in becoming embroiled in the defence of a particular application of this terminology. The ensuing discussion is understood, by the author at least, to be primarily concerned with semantic issues but often strays in the direction of pragmatics.

The term "concept" is used as an explicitly mentalistic term, appying to cognitive or psychological state(s) rather than the abstract, 'platonic', philosophical notion (Fodor, 1981, also notes this potential confusion). Some authors equate concepts with the notion of intension (see e.g., the discussion in Putnam, 1975). Here, intension is restricted to its technical sense of a function that determines the extension of a term, and concept is used in its psychological sense.

The term "dialogue" is used to refer to multi-party discourse in a natural language and is not restricted to situations in which there are only two interlocutors. "Discourse" itself is treated as a superordinate category of extended linguistic performances by one or more individuals; i.e. it subsumes both monologue and dialogue.

1.2 Theorising Meaning

Meaning is a nebulous notion, understood as anything from the emotional impact of a painting to the reference of a particular word. Despite the variety of contexts in which the term is used, informal discourse about meaning in language displays a consistent bias toward treating meanings as objects that are passed between individuals. Reddy (1979) estimates that, in English, 70% of expressions relating to meaning are structured by the "conduit metaphor" (see also Lakoff & Johnson, 1980). The conduit metaphor embodies the idea that communication is a process of transfer of thoughts and feelings using language as a container. For example, the following expressions are typical of talk about meaning (the numbering from Reddy's original article is retained):

1 Try to *get* your *thoughts across* better.

5 You have to *put* each *concept into words* very carefully.

11 The sentence was *filled with emotion.*

13 Your *words are hollow* you don't mean them.

These examples illustrate what Reddy terms the "major framework" of the metaphor. In each case there is some concept or thought which is placed into a word or utterance for transmission and then retrieved through extraction by the listener. In the major framework, an utterance serves as a pipe or conduit directly

linking interlocutors. However, the metaphor also displays a generalisation which suspends the requirement that thoughts are always contained in something. In this case ideas and feelings are 'released' into some general, public, space between individuals where they may exist in an independent or disembodied form until they are picked up again. For example,

27 *Put* those thoughts *down on paper* before you lose them.

30 That concept has been *floating around* for decades.

32 You'll *find* better ideas than those in the library.

Of course, 'common sense' talk about meaning is untroubled by questions of whether it offers a coherent analysis in the theoretician's eye. Such questions are academic, in the pejorative sense. Informal notions do, however, often provide the pre-theoretical basis on which more systematic speculations are based. As a pre-theoretical account of meaning, the conduit metaphor employs a reification of thoughts or concepts which are passed between individuals using language. However, it combines two potentially distinct views on what these objects are. The minor framework of the metaphor highlights the view that meanings are public objects: we see tables and we see meanings. The major framework highlights the intuition that meanings are psychological entities, that we put *our* thoughts into words. Consider Reddy's examples;

8 Your concepts come across beautifully

26 You only have a short time, so try to stuff your essay with all your best ideas

31 I can't seem to get these ideas into words

Concepts and thoughts are treated as both public and private objects, thus obscuring a distinction that forms a major fault line in semantic theory, where

there has been a tendency to treat semantic entities as either psychological states or public objects but not both.

1.2.1 Semantical Realism

Semantical realists view the analysis of meanings in a way resonant with the minor framework of the conduit metaphor. The clearest statement of this approach is found in the work of Frege (1892) who distinguished three components of meaning, Sinn, Vorstellung and Bedeutung, usually translated as sense, idea and reference. For Frege, senses are abstract objects, like Platonic forms, that individuals grasp (sic) with varying degrees of success. Like concepts and thoughts in the minor framework, senses are autonomous objects that exist independently of the human mind. Ideas, by contrast, are subjective psychological states, private to, and dependent on, the individuals who entertain them. As Schweizer (1991) emphasises, Frege understood the grasping of a sense not as a purely psychological state but as something analogous to the realist view of perception where to see an object is to stand in a relation to something external. On this view, a 'grasped' sense is not reducible to its associated psychological states in the same way that an object of perception is not reducible to perceptual sates. My perception of a tree is not numerically or qualitatively the same as your perception but, under the appropriate circumstances, we maintain that it is of the same tree. To preserve the possiblity that the truth or falsity of an expression is a matter of fact, independent of individual beliefs, Frege invoked senses as the semantical objects being perceived or grasped.

Frege's views have formed the cornerstone of subsequent work in formal semantics and many authors explicitly maintain the commitment to semantical realism (e.g., Lewis, 1972b; Dowty, Wall, & Peters, 1981). Montague, whose work was seminal in this area, advocated a view of formal semantics as a branch of mathmatics, not psychology (Partee, 1979). For authors in this tradition, theories of

meaning for natural language must account, at a minimum, for the way in which some expressions are true and some are false. In so far as meaning mediates, in a systematic way, between the form of expressions and the form of the world, formal semantics aims to characterise this relation.

Arguably, the most successful approach in this area has involved the specification of a model theory constructed in a way that provides an abstract structure reflecting the properties of the world relevant to the truth of particular expressions (e.g., Dowty et al., 1981; Kamp & Reyle, 1993; Lewis, 1972b). To provide a mathmatical foundation, the models are constructed using set theory. The standard approach is to determine a set of individuals, U_M, which represents the universe of the model and use subsets of the universe to analyse the simple and compound terms in the language being modelled. Typically, names are assigned to individuals in $i \in U_M$, n-place predicates are assigned subsets of n-place tuples of individuals and so on. Where possible worlds techniques are employed, for example, Lewis (1972b), Dowty et al. (1981), different assignments of individuals in U_M are made for predicates and names in different worlds. The functions that make these assignments are *intensions* which take as their arguments indices determining, for example, a set of possible worlds, a set of contextual coordinates to a speaker, time of utterance and audience, and an index to an assignment of individuals to variables. The range of an intension is its extension in the model. The intension of a sentence is a function from indices to propositions, analysed as sets of possible worlds. The intension of a common noun is a function from indices to sets of individuals in the model. For more complex linguistic categories, the associated intensions are more elaborate. To characterise verb phrases, appeal is made to a function from name intensions to sentence intensions. For adverbs, the function is from verb phrase intensions to verb phrase intensions. The defence of, and interest in, model-theoretic or referential semantics lies in the details of how the intensions create alternative structures for each expression in way that deals with issues such

as quantifier scope ambiguity and pronoun resolution.

1.2.2 Semantical Cognitivism

The metaphysical commitments of semantical realism have proved unattractive for many authors. The appeal to abstract, real entities that connect language directly with the world contributes little to understanding how people actually use and comprehend natural language. Semantical cognitivists maintain that it is the cognitive or mental states of individuals that mediate between language and the world and that a theory of meaning must be characterised in terms of these states. Gardenfors (1993) insists that "meanings are in the head" (p.288). The cognitivist claim, that knowing the meaning of a word consists in associating some thought or concept with it, has a long history. Sperber and Wilson (1986) and Putnam (1988) trace it as far back as Aristotle's *De interpretatione*. People associate some form of mental representation, in Aristotle's terms "affectations of the soul [...] that are in themselves likenesses of actual things", with words that determine the meaning or sense of those words for them. Semantical cognitivism thus gives expression to the conduit metaphor's treatment of thoughts and concepts as psychological entities that individuals 'put into words'.

Early attempts to provide a systematic, cognitive, theory of semantics typically attempted to determine a set of primitive features or properties that exhaust the meaning of a word (Katz, 1972; Katz & Fodor, 1963). Thus "boy" is analysed by appeal to the properties, Male, Human and Non-adult, "girl" by appeal to, Female, Human, Non-adult. The expression "Pat is a boy", is analysed as expressing the conjunction of propositions: Male(Pat) & Non-adult(Pat) & Human(Pat). Clark and Clark (1977) discuss the details of this and related proposals in some detail.

Ultimately, the attempt to isolate primitive properties in this way was abandoned for a number of widely discussed reasons (see, e.g., Putnam, 1970, 1975; Johnson-Laird, 1983). An important problem is that the cognitive structures that

individuals associate with the meaning of words act like vague theories that help to pick out typical instances rather than definitions consisting of various primitive properties. The definitional view suggests, for instance, that something either is a bird or it isn't: i.e. it either has the properties expressed by the propositions which analyse the meaning of "bird" or it doesn't. However, it has been repeatedly demonstrated that people regard some birds as better examples of birdhood than others. Penguins are generally considered less typical examples of birds than Robins (Rosch, 1973). There are also much cited examples of words, such as "game", that do not seem to possess any definitional properties (Wittgenstein, 1958). The task of accounting for these, and similar, observations seems to require appeal to a richer representation than that admitted by definitional approaches.

A range of proposals have been made that try to provide a more adequate account of the cognitive structures that underpin natural language semantics. Work in the area of cognitive semantics has concentrated on developing more elaborate structures for concepts and the nature of their connections, perceptual and experiential, with the world (e.g., Neisser, 1987; Lakoff, 1987; Langacker, 1986). In addition to the truth-functional aspects of meaning emphasised by semantical realists, cognitive semantics is also concerned with finer-grained semantic contrasts, such as "half-full" versus "half-empty", that do not directly impact on the truth of an expression. This has involved appeal to quite highly structured cognitive states such as image schemas and radial categories. Other proposals, e.g., Johnson-Laird (1983), Miller and Johnson-Laird (1976), have concentrated on a procedural analysis of meaning in which some simple, propositional, meaning of a expression is recovered in the first stage of language-processing and then elaborated into a specific interpretation through the construction of a mental model. There have also been attempts to recruit the apparatus of model theoretic semantics to the task of characterising the psychological states associated with production and comprehension. For example, Stalnaker (1987) suggests an interpretation of possible

worlds in terms of epistemologically possible worlds. However, there are a number of substantive problems with interpreting formal semantic structures in this way (see, e.g., Putnam, 1975; Partee, 1979; Gardenfors, 1991).

The details of the alternative models proposed in both the realist and cognitivist approaches are not directly of interest here. For the purposes of this thesis, the importance of semantic theory resides in the contribution it can make to understanding how mutual-intelligibility is achieved in natural language dialogue.

1.3 Theorising Communication

Semantical realism endorses the notion of thoughts as public objects and proceeds to characterise them in terms of abstract senses and propositions associated with the various expressions in a particular language. Semantical cognitivism endorses the notion of thoughts as private, psychological objects and proceeds to characterise them in terms of cognitive concepts and propositions entertained by individuals who speak a particular language. While both approaches develop quite elaborate proposals concerning the structure of thoughts or concepts, neither is concerned with developing a detailed explanation of how these structures relate to the communicative aspects of language. However, the intuition that the notion of meaning is intimately related to communication is evident in both approaches.

As noted above, Frege argued from the fact that individuals could, through communication, 'grasp' the same meaning to the conclusion that senses cannot be reduced to private psychological states (cf. Taylor, 1992). Semantical realists often explicitly acknowledge the importance of mutual-intelligibility as an *explanandum* for theories of meaning (e.g., Lewis, 1969; Davidson, 1977; Kamp & Reyle, 1993).[1]

[1] For example, "central among the problems [for a theory of meaning] is the task of explaining language and communication ..." (Davidson, 1977, p.215) "Languages are for communication. To know a language is to know how to communicate with it ..." (Kamp and Ryle, 1993, p.7).

However, the preoccupation of semantical realism with the specification of abstract structures for a language leaves only a vestigial account of how communication is actually achieved. Probably the most that can be said is that, in the ideal case, individuals who speak the same language make inferences in accordance with the structures specified by the model for that language.[2]

A concern for the possibility of successful communication also informs cognitivist approaches. A typical statement is found in Locke (1690), who explicitly referred to language as the "great conduit":

> "To make words serviceable to the end of communication, it is necessary, as has been said, that they excite in the hearer exactly the same idea they stand for in the mind of the speaker. Without this men fill one another's heads with noise and sounds but convey not thereby their thoughts, and lay not before one another their ideas, which is the end of discourse and language" (Locke, 1690; III.ix.6)

For Locke, successful communication, where it occurs, is due to a matching of ideas, and success is guaranteed only where different individuals respect the same pairing of ideas and words. However, like the accounts of semantical realists, little more than this is said about how the 'matching of ideas' is actually achieved. Oddly, this is still the case in contemporary cognitivist accounts where there is rarely more than a cursory nod in the direction of mutual-intelligibility. More will be said about this below.

On the view that semantics characterises those aspects of meaning given by a language or, perhaps, given in virtue of membership of a particular linguistic community, the relative silence on issues relating to communication may seem perfectly defensible. Communication, it might be argued, is precisely a matter

[2]Dowty et al. (1981) suggest that accounting for speakers' intuitions about entailments is an important goal for model theory.

for pragmatics, not semantics. An important theme of this thesis is to argue that this way of partitioning the explanation of meaning is unsustainable. A preliminary step in developing this claim is to establish that the application of semantic theories to the task of modelling communication imports the assumption that there is some basic, code-like level of meaning which underwrites the mutual-intelligibility of natural language dialogue.

1.3.1 The Naïve Code Model

In ordinary discourse, the notions of meaning and communication are deeply intertwined. The reification of thoughts and concepts in the conduit metaphor is bound up with an implicit theory of what mutual-intelligibility consists in. As Reddy (1979) emphasises, the conduit metaphor assimilates concepts or thoughts to words in a way that naturally gives rise to a view of languages as codes that pair symbols, or groups of symbols, with meanings. This fosters a view of successful communication, in the sense of mutual-intelligibility, where the parties to a dialogue respect the same pairing. While both strands of semantic theory depart substantially from ordinary discourse in their analyses of the nature of semantic structures, both appear to subscribe to this 'naïve code' model of mutual-intelligibility. In support of this claim it is useful to survey the ways in which semantic theory has been applied in modelling communication.

As noted above, the conventional interpretation of formal semantics as characterising abstract, mathematical objects is not conducive to consideration of how agents might actually communicate. However, some recent approaches, such as Discourse Representation Theory, or DRT, (Kamp & Reyle, 1993), have attended more closely to the relationship between individual competence and formal structure in determining interpretation. DRT places certain restrictions on the models it allows and further restrictions are employed in the generation of possible discourse representation structures (DRSs). Nonetheless, where different speakers are

considered, although they may well vary in the assignment of, say, elements in the model to anaphora in a particular DRS, if they speak the same language they are assumed to employ the same model for a particular vocabulary. Of course, DRT is principally concerned with the inter-sentential coherence, not inter-individual coherence, of discourse. Some models with a more direct interest in multi-agent communication employ model theory to provide a basic semantic structure for the agents in their models (e.g., Airenti, Bara Bruno, & Colombetti, 1993; Cohen & Levesque, 1990; Galliers, 1989). Implicitly, agents are assumed to to determine identical interpretation functions for mapping expressions onto their extensions, and the semantic ontology, given by U_M, is effectively transparent to the agents in these models. As a result, the application of formal semantics to situations where more than one agent or individual is considered imports a shared semantic code respected by all the agents.

Work in Artifical Intelligence which examines issues such as multi-agent planning and coordination also assumes the existence of some shared semantic code in accounting for the basic interpretation of expressions by each agent (e.g., Grosz & Sidner, 1990; Houghton & Isard, 1987; Pollack, 1990; Perrault, 1990). Again, these models are not directly concerned with questions of semantic coordination: rather, they presuppose some basic level of coordination in order to concentrate on the investigation of issues such as how speech acts are recovered or how joint plans are constructed. The use of a unique semantic representation is also a common feature of research on natural language processing (Barr & Davidson, 1981), and forms the basis of many theories of discourse analysis (Levinson, 1983; Prince, 1988). In these cases, the investigators make no explicit claims about the inter-individual coherence of discourse but the implication is that every individual who speaks a particular language can be adequately characterised by the same apparatus.

The role of a shared code in determining mutual-intelligibility is particularly clear in information theoretic models of communication (Shannon & Weaver, 1964;

Cherry, 1966). Orignally, information theory was developed as means of characterising the quantitative aspects of communication, i.e. the average quantity of information carried in a channel between a source and a receiver. The qualititative nature of the information transmitted in a particular case can be determined only by reference to the processes of encoding and decoding that take place at the source and receiver (cf. Dretske, 1981; Bar-Hillel & Carnap, 1953). In the terminology of information theory, the message is a set of alternative states at the source which are encoded into a signal via a pairing of various physical characteristics of the signal with different states. The message itself cannot travel but is converted into a signal that is decoded by the receiver(s) to obtain the message. In order for communication to be successful, the process of decoding by the receiver must respect the same pairing of signal characteristics with states. Applied to natural language, this becomes the claim that the mental states or concepts intended by the speaker are encoded into words and then decoded by the hearer (Sperber & Wilson, 1986; Reddy, 1979). Successful transmission of a message therefore depends on the existence of identical copies of the code in both speaker and hearer.

Taylor (1992) discusses a number of code models that have been influential in Linguistic theory, perhaps the best-known of which is the Saussurean notion of *langue*. For Saussure, successful communication occurs and it is this fact that stands in need of explanation. He therefore proposes that languages consist of a set of signs that pair acoustic images with concepts. As before, mutual-intelligibility is possible because individuals who speak the same language respect the same pairing of sounds and concepts, internalised through their exposure to a particular speech community. In many respects Saussure's approach parallels that of formal semantics, the principle interest being in investigating the structure of the code, not the factors which determine how individuals, correctly or incorrectly, internalise it. It is worth noting that Saussure's requirement for a pairing of acoustic

images and concepts is especially strong, since concepts are defined by reference to their role in the whole system constituted by the *langue*. Taylor points out that this has the consequence that individuals cannot differ on just one word, since, if they do, they differ on the entire language.

In most of the cases considered above the code is understood as a property of a language, a property individuals must respect in order to count as speakers of the language. The cognitivist, by contrast, is directly concerned with the precise nature of the cognitive structures involved and the processes by which they become internalised. One (in)famous strategy in dealing with this question has been to adopt a nativist position with respect to concepts. The most radical formulation of this view is associated with Fodor (1975) who appeals to an innate language of thought, or *mentalese*, into which expressions in natural language are translated. The phylogenetic naturalism in Fodor's account is adduced in order to guarantee that the same primitive concepts are shared by the entire species, again providing a shared code that underwrites successful communication. Jackendoff (1992) also argues for an innate 'alphabet' of concepts or primitives shared by all individuals that underpin meaning, although, in conrast to Fodor, he adopts a more Saussurean line on how the meaning of these concepts is determined by the system of distinctions in which they are embedded. Interestingly, Hurford (1989) has developed a game theoretic model which illustrates how the Saussurean code might become fixed by processes of natural selection.

For cognitivists in the empiricist tradition, this kind of analysis is unattractive and other explanations of how a code can become internalised are sought. Locke devoted considerable attention to the imperfections in each individual's understanding which could lead to to "doubtfulness and uncertainty of signification". He cited a number of factors, such as the problems with the retention of complex ideas, understanding of vague ideas that have "no certain connexion in nature" and, somewhat presciently for the discussion of chapter 2, problems with deter-

mining the 'correctness' of ideas whose definition is not widely known or which do not capture the essence of the thing signified. In response to these worries, Locke focussed on the development of prescriptive principles according to which the ideal code-model of communication would more usually be met.

In contemporary cognitive semantics, the issue of mutual-intelligibility rarely receives explicit mention. However, even here code-theoretic assumptions surface in the appeals to factors which can underwrite the sharing of some basic conceptual structures. Significant attention is paid to experiential and perceptual factors that are common for all individuals. For example, Lakoff (1987) repeatedly points to basic-level concepts and the pre-conceptual structuring of experience as a means of defending against a perceived relativistic threat to communication:

> "The existence of directly meaningful concepts –basic-level concepts and image schemas– provides certain fixed points in the objective evaluation of situations. The image schematic structuring of bodily experience is, we hypothesize, the same for all human beings. Moreover, the principles determining basic-level structure are also universally valid, though the particular concepts arrived at may differ somewhat. Thus, certain things will remain constant in assessing situations." (Lakoff 1987, p.302)

Lakoff also discusses "grammaticised" concepts that are conventionalised in a particular community:

> "Concepts that are used in this way are fixed in the mind, or 'entrenched', as opposed to being novel, that is, newly made up. Conventional concepts, shared by members of a culture, are also fixed in the mind of each speaker." (1987, p.321)

For similar reasons, Keil (1981) appeals to universal constraints on learning

as a way of providing individuals with concepts similar enough to allow them to communicate.

The motivation for appealing to these various mechanisms seems to be to provide some basic shared code that offers a minimum level of mutual-intelligibility and in virtue of which more elaborate communicative transactions can be achieved. Thus Johnson-Laird (1983) assumes that people can recover enough meaning from an utterance to apprehend its literal or propositional meaning and then proceed from that to a more detailed model corresponding to a particular interpretation. It is worth noting that a degree of hedging is evident concerning the exact degree of similarity necessary between different individuals' concepts: they are required to be "similar enough" or only "somewhat different". However, nothing is said concerning how this is established or the circumstances under which concepts fail to be similar enough to support communication.

1.4 Summary

The foregoing survey was designed to establish that, inasmuch as semantic theory is bought to bear on questions of communication, there is a pervasive assumption that mutual-intelligibility depends, at root, on the existence of a shared semantic code. Effectively, people who speak the same language are treated as semantically transparent to each other. Some qualifications are in order. This is not to claim that this is a necessary feature of semantic theory: rather, it is a largely 'accidental' consequence of the focus on other issues. Neither is it to suggest that different theories are committed to the same sense of "shared". In semantical realism and approaches such as Saussure's the code emerges as an assumption about the nature of a language considered as a whole, independent of the individuals who speak the language. In cognitivist and information theoretic approaches the code emerges as something internal to individuals who speak the language. Furthermore, even

within cognitivist semantics, there is a range of assumptions about exactly what is shared. In Fodor's case it is a fully regimented formal language including primitive symbols, syntax and proof theory while in Lakoff's it is no more than a set of basic concepts that may be elaborated in different ways in different conceptual systems.

Despite these differences, it has been claimed that all these approaches are committed, in one way or another, to a naïve code model of mutual-intelligibility. In the following chapters it will be argued, on conceptual and empirical grounds, that this assumption is ultimately unsustainable and must be revised in order to provide an adequate account of successful communication in dialogue.

Chapter 2

Content and Commensurability

This chapter elaborates the claim that idiolectical variation undermines the naïve code model of communication. In pursuit of this, section 2.1 provides a general survey of the range of inter-individual differences with respect to normal[1] interpretation and understanding. This leads to discussion of two problematic aspects of semantic theories: firstly, the idealisation of individuals in a linguistic community as semantically transparent to one another –the problem of commensurability; secondly, the attempt to naturalise semantic content in terms of cognitive states –the problem of content.

2.1 Idiolectical Variation

The suggestion that no two people will understand exactly the same thing by a given utterance is tantamount to a cliché. For Fodor and Lepore (1992) it is a "...patent truth that no two speakers of the same language ever speak exactly the same dialect of that language" (p.10). In general, we do not expect members of the same linguistic community to share exactly the same vocabularies nor

[1]Normal here is simply intended to exclude pathological cases such as dyslexia and anomia.

recognise exactly the same meanings for the words they do share. This is, perhaps, unsurprising given the inevitable differences in each individual's exposure to, and use of, language during their lives. Furthermore, there is a general concensus that the principal source of idiolectical variation in understanding is variation in the cognitive structures (e.g., representations/images/concepts) that we each associate with particular words or utterances (amongst others: Chomsky, 1986; Frege, 1892; Lakoff, 1987; Johnson-Laird, 1983; Quine, 1960; Schutz, 1973).

The gross consensus is, of course, consistent with a range of proposals concerning the appropriate analysis of idiolectical variation. The underlying architecture of conceptual structures, their relative stability and their course of development are all contentious issues (see, e.g., Lakoff, 1987; Neisser, 1987). There are also interesting questions concerning their relevance to models of syntax. The concern here is to establish that idiolectical variation should be understood, at least in part, as a semantic phenomenon. *Prima facie*, several aspects of idiolectical variation appear to raise semantic issues.

2.1.1 Misconceptions

Probably the most common and least controversial examples derive from situations where an individual displays some 'deviant' understanding of a term. Burge (1979) discusses the prevalence of idiosyncratic differences between an individual's understanding of a term and its use in a community. Intuitively, there are many cases where people suppose, for example, that a contract is only binding if it is written and signed, or that "brisket" is a cut of beef without knowing exactly which part of an animal it is cut from. Such misconceptions infect much of our normal discourse but they do not ordinarily impede communication. In addition to these 'secular' examples, the everyday use of terms recruited from specialist areas also provides a range of familiar cases. For example, a number of terms from theoretical psychology have passed into common usage but often without

the theory that grounded their original introduction. A familiar case is the common confusion of schizophrenia with multiple personality disorder. Terms such as "neurotic" are often applied with enthusiasm by individuals who have only a vague understanding of the clinical definition of the condition.

Although individual misconceptions of relatively well defined terms provide the least contentious examples of semantic variation there are two respects in which this focus can be misleading. Firstly, they foster the idea that to know the meaning of a word is, in some sense, to internalise its definition. Departures from an established definition or norm yield the possibility of comparisons of, for example, expert and novice understanding. Clearly, such comparisons are only possible where there is some standard conception available to act as a reference point. However, emphasis on these cases can foster the idea that the *correct* concept to associate with a word is its intension as determined by, for example, scientific theory. Nonetheless, although ordinary usage presupposes that words such as natural kind terms pick out a set of things that share some essential essence, it is widely accepted that the concepts individuals actually associate with natural kind terms rarely, if ever, determine that essence (for example: Lakoff, 1987; Johnson-Laird, 1983; Neisser, 1987; Putnam, 1970, 1975). Rather, as noted in section 1.2.2, the concepts associated with terms like "lemon" and "tiger" act as vague theories or sets of defeasible conditions that pick out typical instances of the entity in question. Whatever their actual structure, these stereotypical concepts do not determine the extension of natural kind terms in the way that scientific definitions aim to. Nonetheless, normal practice often grants that an individual knows the meaning of a term even where he or she does not understand, or even necessarily know of the existence of, the relevant science.

The second side effect of the focus on individual deviation from a norm is to obscure cases of divergent interpretation where there is no clear standard of what counts as a correct conceptualisation. For example, there is no accepted

definition of a "friend" and there are probably as many concepts of what constitutes a friend as there are friendships. Domains such as interpersonal relations do not have definitive analyses which render comparisons of 'friendship expertise' tractable. Consequently, they tend to be underrepresented in studies of conceptual and semantic structures. If anything, idiosyncratic variations are probably even more common in these less well defined domains.

2.1.2 Cognitive Depth

Putting aside questions concerning accuracy or correctness, another semantic aspect of idiolectical variation is that it may have cognitively 'deep' effects. That is, individual differences in the use of language about a domain cannot be dismissed as superficial differences in, say, the labels applied to the underlying entities. This point is readily illustrated by work on the effects of conceptual structure on reasoning. Gentner and Gentner (1983) compared the influence of two predominant metaphors that people spontaneously appeal to when prediciting the behaviour of a circuit: as a moving crowd or as water flow. They found that these alternative conceptualisations have a differential influence on problem solving, displaying complementary strengths and weaknesses. For problems predicting the behaviour of circuits that vary in their configuration of resistors, subjects who use the moving crowd metaphor are more successful. By contrast, in problems which manipulate the configuration of batteries, subjects who use the water-flow model generate more successful predictions. Importantly, this indicates that adopting a particular metaphor is not simply a matter of exegetical or communicative convenience, it may reflect the structure of the underlying conceptualisation as evidenced by the way it constrains reasoning and prediction. Gentner and Gentner (1983) provide suggestive evidence that this is equally true of problem solving in scientific contexts where metaphors may guide researchers toward certain concepts and away from others. Analogy and metaphor are extremely pervasive aspects of natural

language (e.g., Lakoff, 1987; Lakoff & Johnson, 1980), and may have correspondingly pervasive effects on conceptualisation.[2]

2.1.3 Range of Variation

In addition to 'depth' it is important to consider the 'breadth' of conceptual variation that needs to be addressed. Cross-cultural research on mental models illustrates just how profound the range of variation can be. Hutchins (1983) compared the mental models of Micronesian and western navigators. Micronesian navigators successfully negotiate voyages of up to 450 miles, out of sight of land, with great accuracy. This is achieved without recourse to charts, compasses or any mechanical navigation aids. In fact, the conceptual models of experienced navigators appear to be so radically different from their western counterparts that these devices are not obviously of use to them. Rather than utilising the familiar absolute 'bird's eye' view of a voyage in two dimensional space, Micronesian navigators adopt an egocentric viewpoint against which the bearings of various reference points change. During a journey, reference islands move along the horizon, the goal moves towards them and the starting point recedes. The horizon itself is conceived as a line parallel to the canoe instead of a circle. Early anthropologists discovered that, without extensive tutoring, techniques such as using several sets of bearings to pinpoint an absolute location in two dimensional space were almost completely incomprehensible to experienced Micronesian navigators. Although the western and Micronesian methods of navigation evolved in response to the same problem they appear to achieve their results through radically divergent conceptualisations and processes of computation.

[2]This is not to suggest, like the strong reading of the Sapir–Whorf hypothesis, that language *determines* thought. The point is rather that what people say about a domain may reflect more than a stylistic preference or trope: it may reflect structural differences in their underlying conceptualisation.

2.2 Idiolectical Variation and Ontology

While it is uncontroversial to suggest that idiolectical variation is widespread, the extent to which it bears on semantic models is far less clear. While the examples discussed above indicate that idiolectical variation raises general semantic issues, a principal concern of this thesis is to press a specific claim about the semantics of idiolects, specifically, that idiolectical variation needs to be analysed, at least in part, as variation at the level of the semantic *ontology*.

2.2.1 The Realist Inheritance

As discussed in Chapter 1, semantical realists assume a set of ontological primitives, the 'universe', over which interpretation functions are defined. These primitives are designed to capture the properties of the world relevant to the truth of different expressions. Determination of exactly which primitives there are and how they are individuated is taken to be principally a question for the natural sciences, or perhaps careful conceptual analysis. The gamble is that the characterisation of semantic properties, such as reference and entailment, can proceed independently of any programme to naturalise the underlying ontology[3] (e.g., Dowty et al., 1981). While this seems to be a viable strategy under a platonist or realist interpretation, as noted above, attempts to adapt the machinery of semantic realism to models of dialogue also import the commitment to a single ontology. It is this commitment which comes under pressure from conceptual variation between different individuals. While realist semantics appeals to a direct mapping between language and the

[3]In fact, the ontology of semantic theories also usually includes, at a minimum, set theory and the logical connectives. The characterisation of these primitives is not normally considered a matter for the natural sciences, not least because they also draw on the same, or similar, machinery. The current discussion concentrates on the non-logical primitives, leaving aside the equally tendentious issue of exactly how the logical and mathematical apparatus should be understood.

world, cognitivist semantics interposes cognitive structures between language and the world and the interpretation function or reference relation must be realised by individuals in the relevant linguistic community. The realist commitment to a unique ontology becomes a psychological claim that there is some universal *conceptual* ontology. *Prima facie*, the examples discussed above suggest that this assumption cannot be maintained.

Pressure on the assumption of a universal semantic ontology poses, in turn, a direct threat to the applicability of semantic theory to accounting for mutual-intelligibility and communication. Divergent conceptualisations undermine the shared code account of communication because they imply divergent ontological commitments. If individuals 'carve up' the world in different ways they will determine different interpretation functions defined for different ontologies. However, naïve code models (see section 1.3.1) assume that the *same* concepts are associated with the same words:

> "Communication is achieved by encoding a message, which cannot travel, into a signal, which can, and by decoding this signal at the receiving end. Noise along the channel (electrical disturbances in our example) can destroy or distort the signal. Otherwise as long as the devices are in order and the codes are *identical* at both ends, successful communication is guaranteed." (Sperber & Wilson, 1986, p.4, emphasis added)

In a naïve code model, a language is actually defined as a particular pairing of signals with messages. If the requirement for identical codes which express those messages is not met, then successful communication is, at best, a happy contingency. In the case of human communication, the codes adopted by speaker and hearer must match in order for successful communication to occur. We obtain the bizarre consequence that, strictly speaking, interlocuters who do not have

identical codes, i.e. who do not determine identical interpretation functions, fail to speak the same language. Unmodified, the appeal to a common code ceases to play any explanatory role in the success or failure of communication.

There are some obvious objections to this analysis. We might be legitimately suspicious whether the difficulties for *naïve* code models really constitute difficulties for code models in general. The most obvious response in this spirit is to contest whether conceptual variation of the kind discussed above really is relevant to the semantics of communication.

2.2.2 Idiolectical Variation as Noise

Methodologically, it might be justifiable to restrict the scope of semantic explanation by stipulating that normal communication consists in precisely those interactions for which a code model is appropriate. While, strictly speaking, there may be no *unique* interpretation function, it might be legitimate to idealise to one in order to maintain theoretical tractability. Like the performance–competence distinction familiar in generative linguistics, semanticists might discount some variation by appeal to performance factors. Certainly, it seems valid to claim that a semantic theory should not be required to deal directly with, for example, memory limitations or pathological cases such as slips of the tongue. However, it seems that this approach is not really plausible for variation of the kind recorded by Gentner and Gentner (1983) and Hutchins (1983). The basic difficulty is that we observe conceptual variation even where perfomance factors, in all relevant respects, are constant.

The parallel with syntactic theory does, however, suggest a more robust line of response. Different subgroups of speakers who share, to some degree, particular conceptualisations might be regarded as speaking the same 'semantic dialect'. The prevalence of idiosyncratic variation suggests that this idealisation may ultimately be unsustainable, however, even the established syntactic notions of dialect

and language are often considered to be convenient fictions unsupported by any systematic linguistic distinction (Chomsky, 1986). The fact that, in the long run, such idealisations are difficult to defend empirically does not necessarily undermine their programmatic value. An obvious advantage of this line of response is that it offers a straightforward analysis of cases like those reported by Hutchins and Gentner and Gentner (*ibid*). The differences between subgroups or cultures can be accommodated as dialectical variation. More importantly, the parallel with syntax raises the possibility of appealing to a set of innate semantic universals which constitute a common semantic structure underpinning dialectical (and, to some extent, idiolectical) variation –in Chomskian terms, an I-language (cf. Jackendoff, 1992). If, following the Chomskian lead, we take the goal of psychological semantics to be elucidation of innate semantic universals then the threat posed by idiolectical variation is blunted. Some variation could be factored into the parameters which fix different dialects and some could be factored into variation in the integrity of each individual's 'semantic organ'. The resulting semantic theory could ignore the noise created by individual variability and concentrate on mapping the underlying semantic universals. In addition to defusing the challenge from conceptual asymmetries this approach also suggests a possible reconciliation with semantical realism since it offers a determinate ontology conditioned, through natural selection, by the structure of the world.

2.3 Commensurability

Although the parallel with the Chomskian approach to syntactic theory brings questions of innateness to the fore, nothing hinges directly on the, potentially controversial, appeal to genetic endowment. The central issue for current purposes is whether it is possible, in principle, to devise a set of universal semantic primitives. Idiolectical and dialectical differences must yield to an analysis which

renders them *commensurable*. If we can identify a semantic 'lowest common denominator', idiolectical differences can be reduced to the same basic concepts and thereby rendered mutually translatable. Given the right kind of inferential apparatus for encoding and decoding utterances, we obtain an account of the mutual-intelligibility of natural language.[4] The pivotal assumption for this general line of response is that idiolects and, *mutatis mutandis*, dialects do, in fact, respect some common semantic measure which guarantees their commensurability. The assumption of commensurability has, often tacitly, informed research programmes in many areas. It is instructive to consider its influence in three areas: machine translation, developmental psychology and philosophy of science.

2.3.1 Cross-Linguistic Commensurability

The grossest requirement for a set of universal semantic primitives is that they should capture all the potential distinctions available in all natural languages. Cross-linguistic comparisons between, say, French and English, often reveal distinctions in one language which are not found in another. For example, there is no French word that directly translates the English "chair", neither "chaise" nor "fauteuil" capturing the same set of properties. Conversely, the French word "porte" does not discriminate between the English words "door" and "gate" (Kay, Gawron, & Norvig, 1994). Another example from French (discussed by Kuhn, 1983) is "pompe" which translates to the English "pomp" in ceremonial contexts and to "pump" in hydraulic contexts. It is apparent that concepts such as CHAIR,[5] that might normally be invoked by a semantic analysis will need to be abandoned in favour of more finely individuated primitives. Nonetheless, given an appropriately

[4]Sperber and Wilson (1986), e.g., pp.26–27, pursue just such a heterogeneous approach arguing for both an innate code and inferential processes.

[5]Notational convention: where the concept is intended it will be rendered in upper-case e.g., CHAIR, the corressponding lexical items will be rendered in lower-case with double quotes to distinguish mention from use, e.g., "chair".

revised ontology, these examples might be amenable to an analysis which renders them commensurable.

The prospects for this strategy of re-analysis look worse when we consider more complex examples. Even restricting attention to French-English mismatches, less tractable cases are apparent. The French word "esprit" can, depending on the context, be translated as "spirit", "aptitude", "mind", "judgement" or "wit" (Kuhn, 1983). Furthermore, French and English share a common root; if we switch attention to translation mismatches between languages such as Japanese and English the difficulties multiply (see, e.g., Lakoff, 1987; Quine, 1969). Kay et al. (1994) provide detailed discussion of the contrast between the Japanese verb "nomu" and its normal English translation, "drink". In fact, "nomu" has a slightly wider range of application and can be used for medicines or small objects. This suggests a translation as "swallow". Assuming we require two different universal concepts, DRINK and SWALLOW, a problem arises because SWALLOW is also the right concept to associate with the English word "swallow". But this leaves no way of accounting for the fact that "nomu" should sometimes translate as "drink". While it is literally correct to decribe drinking as swallowing, in some contexts it sounds very marked. Translation of a Japanese phrase as "she is swallowing water" implies a context like drowning rather than drinking. Although the specific difficulty may be resolved by devising an extra translation rule, such a move seems *ad hoc* not least because the contextual factors which determine a "drowning" interpretation are quite diffuse.

Translation mismatches have had important repercussions for attempts to automate the process of translation. An important strand of research in machine translation has focussed on the attempt to specify a language-neutral semantic 'interlingua' which, like a universal conceptual ontology, captures all the basic semantic distinctions in the languages, or language fragments, it translates (Kay et al., 1994; Barr & Davidson, 1981). Once the interlingua has been determined,

translation consists in decomposing the source sentence into its primitives and then producing a meaning-preserving reconstruction in the target language. However as examples like "nomu" illustrate, the accuracy of a translation seems to depend not only on a particular source sentence considered alone but also on its relation to the broader context, including the narrative context and the conceptual resources and distinctions available in the language as a whole. Terms in one natural language often seem to involve concepts almost 'orthogonal' to those associated with another. Although any specific example may be resolved by adding an extra translation rule, the overall frequency of such mismatches raises the possibility that something approaching world knowledge is necessary for their resolution. Kay et al. (1994) conclude that the interlingual approach's lack of success is due to reliance on a literal, what has here been termed code, model of what constitutes translation. Instead of regarding translation as strictly meaning-preserving they argue for an approach which models translation as a process of active negotiation (cf. Bell, 1991).

2.3.2 Commensurability and Development

The literature on child development has seen a parallel dispute surrounding the validity of what amounts to a 'code theoretic' picture of development. As Gopnik (1983) notes, although the details of each proposal vary, there is a strong, roughly Piagetian, tradition in theories of conceptual development which maintains that a child's concepts of things like "animal" and "weight" can be analysed in terms of a fixed set of semantic/conceptual primitives. The course of development is envisaged as a process in which universal conceptual primitives are combined into successively more sophisticated structures as learning proceeds.[6] In turn, the child uses these structures in order to interpret the language it is exposed to,

[6]See Jackendoff, 1992, pp.57–59, for a particularly explicit statement of this view.

thus providing a model of how the child's intellectual and communicative abilities develop.

As in the case of cross-linguistic comparisons, this assumption appears to be empirically inadequate. Carey (1988) reports a detailed study of the interrelationships between the preschool child's concepts of "animal", "death" and "life". These concepts diverge from those in adult language in a number of ways. The concept of animal centres on a 'vitalist biology' which takes internally generated activity to be the key property of living things. There is no recognition, for example, that each species must solve universal problems such as obtaining food or reproducing and no understanding of internal bodily systems such as eating, breathing and circulation. Death is conceived of as a reversible form of separation where the dead effectively live on but in an altered state. For the preschooler, death is part of an undifferentiated concept that includes UNREAL, INANIMATE and NONEXISTENT. Carey claims that the meaning of the preschool child's biological concepts is partially determined by their interrelations within a whole nework of related concepts which inform their reasoning about the world *en masse*. Support for these claims comes from the influence these concepts have on structuring the child's inferences (cf. Gentner & Gentner, 1983); for example, a child who asks why it is possible to see objects like statues and tables if they are dead. Conversely, children frequently maintain that active, useful things are alive. They will classify, amongst other things, the sun, bicycles, the moon, fire and the buttons on their trousers as alive.

Karmiloff-Smith (1988) utilises an experimental paradigm to investigate how children's concepts of "weight" evolve during development. Six to seven year olds will balance blocks on a beam according their geometric centre. If the weight distribution of the blocks is shifted away from the centre, by attaching a weight to one side of the block, this severely interrupts their performance. Both younger and older childer, by contrast, are able to balance both 'off-centre' and normal blocks.

The detailed differences between the age groups show some marked parallels with the Kuhnian (see below) picture of scientific theorising. The younger children take each block seperately and use simple proprioceptive feedback in order to get the block to balance. They show no evidence of making any generalisations about balancing points for sets of blocks. Even if they have successfully balanced one block and an identical one is available they do not appear to have any expectation that it will behave similarly. In contrast to this, the six to seven year olds adopt a 'geometric centre' theory for predicting how each block will balance. In fact, they apply this theory so rigidly that when they fail to balance an off-centre block they respond as if the problem was with *their* method, for example, trying to place the block at the same point but much more slowly. For the older children, the failure to balance the block is treated as data about the block, not a failure of their method. As Karmiloff-Smith (1988) argues, like scientists at successive stages of a research programme, children appear to adjust the boundary between data and theory in a manner that frequently involves sacrificing data in order to preserve theory.

These examples call into question what amounts to an assumption of commensurability in developmental theorising. Models which appeal to conceptual primitives acting as basic building blocks do not fit comfortably with the studies outlined above. Rather, many developmental psychologists argue that children should be regarded as actively theorising about their environment in a manner that undermines the validity of the appeal to universal conceptual primitives (e.g, Gopnik, 1983; Carey, 1988; Karmiloff-Smith, 1988). As Karmiloff-Smith (1988) puts it:

> "They [children] constantly develop theories and create domains, carving and re-carving nature at new joints" (p.192).

Echoing the problems created by translation mismatches, elucidation of the

nature of a child's concepts appears to depend in opaque ways on the context, or theory, in which they occur.

2.3.3 Commensurability and Scientific Theory

The emphasis on children-as-theorists was explicitly inspired by the work of Kuhn (1970) and Feyerabend (1962) on diachronic change in scientific theories. This work highlights the problems created by the way 'semantic interdependencies' distribute throughout a theory. For example, Kuhn (1970, 1983) documents the relationship between successive theories in disciplines like chemistry. Eighteenth century chemical theory made extensive use of the notion of 'phlogiston' to account for various experimental data. As Kuhn (1983) points out, nothing in contemporary chemical theory can directly translate this term. Although in some cases it is possible to identify a referent in terms of modern theory, for example as oxygen or an oxygen-rich atmosphere, in other uses such as "phlogiston is emitted during combustion", there is nothing recognised by contemporary chemical theory which could act as a referent. In fact nothing, currently acknowledged, combines all the appropriate properties. Consequently, any translation of an eighteenth century text needs to adduce a range of different substitutions, and in some cases blanks. Without an extra gloss provided by the translator it is not even clear that a single notion or entity is actually intended (Kitcher, 1988). Contemporary theory also lacks the auxiliary concepts corresponding to "principle" and "element". Phlogiston was interdefined with these terms which determined a web of relationships that do not map directly onto to anything in subsequent theories (Kuhn, 1983).

Wiser and Carey (1983) draw parallel conclusions from their study of the concepts of HEAT and TEMPERATURE. Prior to Black, these two concepts were fused in a single notion, one that has no counterpart in current theory. Wiser and Carey carefully rule out the possibility that this was due to a false belief that there really were two different concepts involved but they were perfectly correlated. The

older concept of HEAT combined both causal strength and qualitative intensity, properties separated by the later theory. Nor were heat and temperature measured separately by experimenters: instead they related a single variable, "degree of heat", to a range of phenomena. From a contemporary perspective there is no such thing as "degree of heat". Although any use of the concept in a particular context might be held to correspond to either heat or temperature there is no translation into contemporary theory which successfully preserves the sense of the original concept.

Kuhn (1970, 1983) claims that changes in theory frequently amount to changes of 'world view' or 'paradigm' and that when such a transition occurs, the language of a scientific community, before and after the change, may not be mutually inter-translatable. For example, Kuhn argues that Newtonian mechanics cannot be derived as a special case of Einsteinian unless the concepts corresponding to "space", "mass" and "time" are actually *reinterpreted*. Conventional, homophonic, translation fails to respect the basic sense of each term. For instance, Newtonian mass is always conserved whereas Einsteinian mass is inter-convertible with energy. Only where relative velocities are low can they be measured in the same way (Kuhn, 1970, p.102). The attempt to reduce Newtonian Mechanics to Einsteinian also runs into difficulties because of the interdependence of each term within the theory. In Newtonian mechanics, the terms "force" and "mass" have to be acquired together with Newton's Second Law. They can't be learned independently because the meaning of these concepts is *defined* in terms of their mutual interdependence. In Einstein's theories, Newton's Second Law does not apply, leaving no counterpart to Newtonian "force" and "mass" which might admit a direct translation. Although the same terms occur in both theories their various interdependencies are transformed between the two in a "displacement of the conceptual network"(Kuhn, 1970, p.102).

2.4 Incommensurability

The parallel with syntactic theory raised in the previous section suggested an argument for discounting idiolectical differences as irrelevant to the interests of semantic theory. This move gains plausibility from a certain intuitive picture of conceptual differences as differences in beliefs about what are, in some sense, the *same* underlying entities. As Putnam (1981) puts it: "as the difference between the primitive concepts employed, by a theory, which are constant, and the conceptualisations in which they are employed which may vary" (p.116). Individual variation, synchronic and diachronic, is construed as a matter of differences in the way primitive concepts are combined, not as differences in the basic vocabulary.

Prima facie, each of the examples raised above undermine this picture. Substantial cross-cultural variation of the kind recorded by Hutchins (1983) (see also Lakoff, 1987) indicates that the goal of developing a universal ontology, if possible at all, will demand a radical revision of what are normally considered to be the primitives in semantic analysis. In order to accommodate translation mismatches, patterns of conceptual development and conflicting scientific theories it seems reasonable to suppose that we would require a vocabulary couched at a very low level, possibly in terms of perceptual or sensory primitives, in order to ensure adequate coverage. While these examples may shift the burden of plausibility against such a programme, thus far, they do not definitively rule it out.

However, there is a much more serious problem presaged by each of the examples in section 2.3. They call into question whether it is actually possible, even in principle, to isolate a set of semantic primitives common to all languages (or developmental stages, or theories) independently of the various conceptual and theoretical contexts in which they figure. Without such a division, there is no reason to suppose that idiolects and dialects are, in fact, commensurable.

Probably the best-known proposal concerning the incommensurability of different frameworks is associated with Kuhn (1970) who, argued that the changes

in 'world view' that result from radical revisions of scientific theory render many of the terms of each community incommensurable. Drawing on his historical analysis of patterns of theory change, Kuhn argues that the web of interrelationships between terms such as "force" and "mass" guarantee that substantial theoretical shifts, even where the same words (or perhaps strings) are retained, also entail ontological shifts. The examples above suggest that to the extent this holds for scientific theory it also holds for developmental change, cross-linguistic disparities and for conceptual differences in general. In each case, the diffuse influences of the wider conceptual or linguistic background undermine the attempt to isolate the theory-neutral content of a concept from its context.

Kuhn's argument is induced from the patterns of actual, historical, scientific practice. Quine's paper, "Two Dogmas of Empiricism", (Quine, 1953) provides a more systematic analysis of what is essentially the same difficulty.[7] Quine's attack begins with the first dogma, the analytic-synthetic distinction. Taking a number of proposals in turn, Quine shows that definitions of analyticity in virtue of, e.g., semantic rules, definition, or meaning, are inherently circular. Ultimately, they involve appealing to intentional notions which themselves presuppose some adequate definition of analyticity. An apparently attractive proposal for breaking this circle is offered by the second dogma: roughly, the suggestion that meanings can somehow be reduced to their primitive empirical content. The verificationist program represents the most notable example of this, aiming to systematically reduce the meaning of any 'high-level' statement to statements about the set of primitive sense data which could confirm or infirm it. If achieved, this reduction could provide a reconstruction of the notion of synonymy in terms of statements

[7] One obvious discontinuity between Kuhn's and Quine's positions is the latters focuses on the meaning of statements rather than terms, a consequence of his view that reference is secondary to truth. However, as Quine notes, the considerations he raises apply equally to terms and can be derived by verifying whether the substiution of terms in a statement maintains synonymy. See Quine, 1953, pp.37–38 and footnote 15.

having the same empirical content or verification conditions and a reconstruction of the notion of analyticity as statements which are always confirmed, no matter what. However, as Quine emphasised, the difficulty lies in the intimate interconnections between theory and data. In practice, scientific statements are neither exclusively empirical nor exclusively theoretical: "statements about the external world face the tribunal of sense experience not individually but only as a corporate body."(1953, p.41). There are no 'empirical' statements which cannot be salvaged in the face of recalcitrant data by appeal to auxiliary hypotheses, e.g., methodological factors or even hallucination. Conversely, there are no theoretical claims, including logical laws, which are immune from revision. The response to any particular piece of evidence depends, amongst other things, on global factors such as theoretical conservativism, simplicity and heuristic value.

Although widely accepted, the Quinean analysis is not uncontroversial. Fodor and Lepore (1992) in particular have suggested that, in fact, it cannot be coherently formulated as a semantic thesis. The crux of their argument, and the point from which all their objections stem, is that Quine holds both that a) the meaning of a statement can be held constant in the face of confounding data by revising the meaning of other statements and that b) because theories meet the "tribunal of sense experience [...] as a corporate body" changes in the interpretation of *any* single theoretical statement entail revisions in the meaning of *all* the theoretical statements. As a result, they urge, Quine is caught in a dilemma: semantic holism with respect to an entire theory is inconsistent with the suggestion that one statement may keep its meaning while others within the same theory change. It is not at all clear however that Quine is actually committed to either position. Firstly, he nowhere claims that the meaning of an ostensibly falsified prediction is held *constant* while others are revised. Rather, a threatened statement may retain its truth value by revision of the truth value of one or more auxiliary

statements.[8] Without establishing the additional step that truth value exhausts meaning, there is no contradiction. Revision of the truth value of auxiliary statements may well alter the meaning of a statement, for example, by changing its reference, but nonetheless leave its truth value intact. In point of fact, Quine has repeatedly emphasised that all the truth values of a theory can be held constant under systematic variations in their ontology (Quine, 1960, 1992).

Taking up the second horn of the dilemma, it also seems clear that Quine is not actually committed to the suggestion that revising the truth or, indeed, meaning of one statement in a theory necessarily implies revision of *all* other statements of a theory. In fact, this is an option which Quine explicitly rejects (Quine, 1953, 1992). For the current point, all that is required to get the difficulties going is that it be unpredictable which auxiliary statements are revised in the face of problematic data. As long as theory revision depends on nebulous considerations such as simplicity and conservativism then the attempt to determine a set of fixed points around which theories evolve is forlorn. There is nothing to guarantee that it won't be the truth value of ontological claims that is sacrificed. Returning to the other examples, the diffuse effect of embedding context on the content of particular terms (or sentences) is apparent in the case of translation mismatches and conceptual development in children. Just as theory revision may involve disparate, unpredicatable elements of the entire theoretical nexus, so the accuracy of a translation may depend on diffuse elements of the linguistic context. Similarly, a child confronted by "recalcitrant experience" may choose to revise any of the auxiliary hypotheses available to it, including those relating to its ontological commitments.

The problems identified by Quine also feed into the prospects for naturalising

[8]For example: "A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements" (1953, p.42) "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system" (1953, p.43).

a candidate set of semantic universals. Establishing a plausible revision of the conventional semantic ontology is necessary but not sufficient for a code-theoretic account of communication. As discussed in section 1.2.2, there must also be some mechanism which ensures that different individuals come to represent or realise the same basic conceptual alphabet.

2.4.1 Naturalising the Code

Models that appeal to learning-based accounts face the immediate difficulty that the requirement for a universal conceptual vocabulary that accommodates the distinctions encoded by *all* natural languages has the consequence that the requisite conceptual primitives are not discriminated directly by the resources of any *particular* natural language. As a result, the conceptual primitives required are necessarily more basic than those that a child is exposed to when learning any given natural language. This difficulty is compounded by the fact that nothing guarantees that a child will actually receive appropriate exposure during development (cf. Chomsky, 1986). Additionally, Quine (1960, 1969) has argued that even if we restrict attention to the ontology of a single natural language, the overt behavioural data from which a child might try to derive its conceptual primitives radically underdetermines the choice of a specific ontology or set of concepts. There are an indeterminate number of equally viable ontologies consistent with observed linguistic behaviour (see also Davidson, 1984; Putnam, 1975).

The most popular solution to the difficulties with learning-based models has been to propose that a species-specific conceptual vocabulary might be fixed to some degree by evolution (amongst others, Fodor, 1975; Sperber & Wilson, 1986; Jackendoff, 1992). This proposal has been widely criticised (e.g., Putnam, 1975, 1988), not least because there is no model of natural selection which could realistically drive such a vocabulary to fixation. Many concepts currently in use had no significance for the survival of our ancestors. For example, Putnam (1970)

discusses Katz's theory of semantic markers which attempts to analyse the meaning of a word in terms of innate conceptual universals such as UNMARRIED, ANIMATE, SEAL. As Putnam notes, it is difficult to see what theory of human evolution is going to provide an account of the fixation of SEAL, let alone CLOTHING or FURNITURE. In general, the conceptual resources of natural languages invoke a wide range of entities that simply did not exist in evolutionary time. Furthermore, it is not obvious how we can construct an argument that evolution would act to fix a *species-specific* code without appealing to group selection as the principal selective force. However, group selection provides by far the weakest form of selection pressure, one which is readily overturned by other pressures (Healey, 1991). In the absence of a detailed model that addresses these worries, this aspect of the nativist explanation is difficult to sustain.

The natural way to resist these difficulties is to propose that evolution fixes a set of more primitive concepts which could then be built up, during development, into successively more complex structures. This avoids the unhappy conclusion that concepts such as CARBURETTOR and COMPACT DISC are innate. However, the Quinean argument indicates that this is, in principle, impossible. Indeed, Fodor (1981, 1983) accepts that empiricist and verificationist attempts at reduction to "sensation concepts" failed precisely because no such reduction is possible. For Fodor, a natural language is effectively as expressive as it needs to be and "you can't hardly reduce it at all" (1981, p.213). As for languages, so for idiolects.

2.5 Ontological Pluralism

The first part of this chapter has concentrated on the problems idiolectical differences create for code-theoretic models of communication. Idiolectical variation is both pervasive and marked, an observation that undermines naïve code models. The attempt to salvage code-based theories by appeal to a set of primitive

semantic universals runs up against arguments to the effect that the ontology embodied by a particular idiolect (or theory, or language) cannot be isolated from the theoretical commitments which serve to individuate it. The attempt to 'step outside' idiolectical variation to determine an absolute, universal, semantic ontology of concepts or sense data cannot succeed without compromising or reinterpreting the terms and statements of the originals. Even those authors who are sceptical both about the coherence of claims about incommensurablity and its diagnosis as a semantic condition offer no alternative proposals which might obviate the problem. Quine (1992) concludes that the reification of objects, abstract and 'real', is a *retrospective* theoretical move motivated by the desire to integrate our system of beliefs with the world. The ontologies associated with theories and, *mutatis mutandis*, idiolects emerge as "ideal nodes at the foci of intersecting observation sentences" (p.24). As systems of belief vary, so, frequently, will our ontological commitments. Models that propose to account for communication must accommodate some degree of ontological pluralism between interlocutors.

Perhaps the principal motivation for suggesting that, at some level, there is a shared conceptual code is the intuition that there *must* be one in order for communication to be possible at all. The proposals for innate mentalese, conceptual universals in development and a language-neutral interlingua all trade on the code-theoretic assumption that what makes translation between conceptual schemes possible is a common set of conceptual elements. This intuition also informs criticisms of the arguments for incommensurability. For example, Kuhn's analysis of theory change has been challenged on the grounds that it is self-refuting. In the course of his argument Kuhn apparently offers inter-translations of theories that are *ex hypothesis* not mutually translatable (see Kuhn, 1983; Putnam, 1975). The same worry is apparent in critiques of the literature on incommensurability between adult and child language (Carey, 1988). Adults and preschool children are perfectly capable of communicating and this may seem mysterious if they

do not employ commensurable languages. Furthermore, developmental psychologists, including those who endorse the notion of incommensurability, frequently *do* translate children's conceptualisations into adult language. Of course, the accuracy of these example inter-translations between adult and child language and different scientific paradigms is largely an empirical issue. To insist that these are clear cases of inter-translation, and that successful inter-translation presupposes a common semantic ontology, courts an obvious circularity.

More importantly, the claim that two idiolects are incommensurable does not entail that there can be no informational commerce between them. In fact, Kuhn (1983), elaborating the metaphor, points out that incommensurable magnitudes can be compared to an arbitrary degree of accuracy. Chapter 5 of this thesis is precisely an attempt to provide a semantic framework for dialogue which does not depend on assumptions of commensurability. Rather than rejecting the notion of incommensurability as incoherent on the grounds that successful inter-translations occur, we might equally conclude that it is the concept of accurate translations as 'ontologically invariant', or, perhaps, reference-preserving, which must be abandoned. The issue is not whether it is useful or productive to have a strong notion of translation but rather whether the actual practices to which it has been applied, for example, translation between languages, idolects and theories, ever actually meet the strict criterion of preserving reference. Kuhn (1983) and Carey (1988) both question whether the strong notion of translation is ever appropriate. What we ordinarily refer to as translations are, on this account, really interpretations, or perhaps instances of language acquisition (see also Lakoff, 1987). Carey argues that in the case of child development the psychologist is effectively learning the child's language and attempting to teach it to us, not passively reporting its meaning. In the applied case of machine translation, semantic mismatches between natural languages have proved so problematic that the attempt to identify a universal interlingua has been all but abandoned. Kay et al. (1994) argue that the

notion of a pure interlingua fundamentally misconceives the nature of translation. Rather than regard translation as a function between languages that preserves meaning in any absolute sense they advocate a view of translation as a negotiated compromise.

Having elaborated an essentially negative thesis about the consequences of idiolectical variation, the next section develops the claim that, in fact, idiolectical differences may be instrumental in giving semantic and, more generally, intentional states their distinctive character.

2.6 Idiolectical Variation and Content

Recent debate in the philosophy of mind has generated some surprising conclusions about which objects or states can be sensibly identified as the 'bearers' of meaning. The most obvious suggestion for a psychological theory is to analyse the meaning of a word in terms of the associated psychological state(s) of the speaker or hearer, and perhaps allow this to vary from individual to individual. However, a number of arguments have been advanced to show that such a move fails to satisfy some basic, 'pre-theoretic', intuitions about meaning. In particular, it seems that ordinary usage presupposes that the meaning or semantic content of words is a *distributed* property of individuals, their linguistic community and the physical environment. This implies (at least) two challenges to a *cognitive* account of meaning. The weaker challenge is that the meaning of a word cannot be determined solely by reference to the cognitive states of a particular individual or community; instead it is minimally a composite situation in which some state of affairs in the head stands in an appropriate relation to some state of affairs in the environment. On this account we might still maintain a theory of narrow content, i.e., those characteristics shared by two individuals who "mean the same thing" or "think the same thought" after the "broad" influences of the socio-physical

environment have been discounted (e.g., Fodor, 1980). The stronger challenge is that whatever the mental structures associated with language turn out to be, they bear *no* constitutive relevance to the meanings of words (see Pettit & McDowell, 1986). The problems raised by these arguments are obviously of great significance for any programme which aims to develop a 'psychological semantics'.

2.6.1 Content and the Environment

The best known arguments for the distribution of semantic content originate with Hilary Putnam (e.g., Putnam, 1975, 1988). Putnam's principal concern is with the 'world-involving' aspects of the meaning of natural kind terms, i.e., nouns associated with entities that have explanatory importance within some, possibly informal, theory. As discussed in section 2.1.1, the standards of normal usage require only a relatively underspecified knowledge of a natural kind term for competent usage. As a result we might accommodate this within a cognitivist semantics by determining a representation that captures a set of incomplete (i.e., not necessary and sufficient) conditions that each person who competently uses the word has internalised. However, Putnam shows that not only do the cognitive states associated with knowing a word underdetermine its 'actual' extension as determined by the relevant science, the same intension, conceived as a cognitive state, may determine indefinitely many different extensions. Putnam (1975, 1988) illustrates this possibility with the aid of a thought experiment.

Two situations are considered, one on earth and one on a possible twin earth. Both situations are assumed to be identical in all respects, including linguistic conventions, except that water on earth has the chemical structure H_2O whereas on twin earth it has the chemical structure XYZ. The argument turns on the intuition that the word "water" on earth has a different extension from the word "water" on twin earth. Specifically, someone using the word "water" on earth is referring to H_2O whereas the counterfactual twin using the word "water" on twin

earth is referring to XYZ. Crucially, this intuition holds even though, by hypothesis, there is no difference between the cognitive states, narrowly understood, the twins associate with the term "water". The only variation between the two situations is the ultimate composition of the liquid they both call "water". Neither twin needs to be able to discriminate H_2O from XYZ through taste, touch or any other superficial characteristics. Indeed, it does not even matter whether anyone in the relevant communities, including scientists, can discriminate H_2O from XYZ, only that they might in principle do so (see section 2.6.3).

The general conclusion Putnam draws is that what a natural kind term means cannot be determined solely by reference to the internal cognitive state(s) of speaker-hearers. The reference of any natural kind term is determined partly by the ultimate nature of the stuff referred to itself. As a consequence we cannot adopt a cognitive interpretation of semantic theory which satisfies both the claim that meanings are mental states and that meanings determine extensions.

2.6.2 Content and The Community

While Putnam emphasises the role of the physical environment, Burge (1979, 1986) develops arguments which focus on the contribution of the social environment in determining the content of mental states. Burge (1979) sets up a thought experiment involving two possible situations. In both situations we consider an individual who uses the word "arthritis" perfectly competently under the same circumstances to the same effect. For example, to agree that "arthritis is painful", that "it affects the old" that "stiffening of the joints is one of its symptoms" and so on. In both cases it is assumed that the individual's physical, functional and phenomenological history is identical. He is exposed to the same perceptual experiences, has the same physiological history and exhibits the same behaviour. In short, the situations are identical in all non-intentionally described respects. In the first situation the individual makes a visit to the doctor during which he suggests

that his arthritis has spread to his thigh. The doctor explains that this cannot be the case since arthritis is specifically an inflammation of the joints. The patient has expressed a false belief. Burge then contrasts this with a second counterfactual situation where, as before, the patient visits their doctor and declares the same worry. However, in this situation the term "arthritis" *can* be legitimately applied to the patient's condition. The relevant bodies of expert and lay opinion all hold that the term "arthritis" applies to any rheumatic ailment. Under these circumstances the patient has expressed a belief with a different content; a true belief. In both cases this is the first time the patient has ever expressed his belief and, *ex hypothesis*, his disposition to express it has resulted from precisely the same (non-intentionally defined) patterns of experience. The only difference between the two situations is the conventional application of "arthritis" in the patient's linguistic community, a difference of which the patient is ignorant.

The contrast between the situations brings out the intuition that the content of the patients' "arthritis-beliefs" changes from the first situation (essentially ours) to the second. In the second case none of his beliefs, e.g., that his father had it, that it afflicts the old, are about arthritis as we understand it. They are beliefs about an extensionally and definitionally different condition. This is demonstrated by the change in the truth of the belief that "my arthritis has lodged in my thigh" between the two situations. However, nothing about the patients' internal states, narrowly understood, varies between the two situations. The difference in the content of the patients' beliefs arises purely as a result of the change in the conventional interpretation by the relevant community. While the patients' misconception brings out the difference between the two situations it is their other, related, beliefs that give the thought experiment its force. Importantly, ordinary practice indicates that we commonly attribute to individuals beliefs whose contents they may incompletely understand. The kind of deviation from common usage of "arthritis" envisaged in the thought experiment does not preclude us from attributing beliefs whose

content is specified by reference to the meaning of "arthritis". Under such circumstances we normally regard an individual as still holding beliefs about arthritis (or contracts, or neurosis), just that some of them are false.

Burge's thought experiment turns on the observation that mental states are commonly individuated by reference to the content ascribed to them in the subordinate 'that-clause' of propositional attitude descriptions. For example, the difference between the psychological states associated with the attributions, "believing that arthritis is painful" and "believing that rheumatism is painful", depends on the distinction between the meaning of arthritis and rheumatism respectively. Burge draws out the intuition that this difference of meaning is exhausted only by appeal to the linguistic practices of the relevant community: discrimination of a psychological state as a state of a particular kind is a distributed property of an individual and their linguistic context. A given individual may be accredited with mental states having a certain content, including attributions by and of themselves, even where these 'narrow' mental states are, by communal standards, incorrect. The socio-linguistic context plays a constitutive role in individuating mental states, challenging the viability of individualistic cognitive models.

This formulation is ambiguous in an important respect. While it is clear that the narrow cognitive states of some arbitrary individual in a community may not determine the meaning of a word, it might still be argued that there is always some expert who commands a definitive understanding of the concept in question. On this view, the distribution of content is a simple consequence of what Putnam (1975) termed the distribution of linguistic labour.

2.6.3 The Division of Linguistic Labour

In essence, Putnam's (1975) proposal was that although most individuals have no knowledge of the theories which might determine what count as instances of a term like "gold" or "tiger" there is a distribution of knowledge within a com-

munity on which they can rely to provide the 'official' meaning. For example, although I may not know what the ultimate constitution of gold is I can nonetheless use the word in a particular instance, deferring to experts for the definitive judgement on the integrity of a sample and, thereby, the appropriateness of the term. The temptation to read the (subsequent) arguments for broad content solely as elaborating aspects of the distribution of linguistic expertise in a community is compounded by the fact that both Putnam's and Burge's arguments suggest appeals to expert opinion for adjudication on the meaning of the contested word. In Burge's argument reference is made to the existence of expert medical opinion, in Putnam's, physical scientists. According to this interpretation, Putnam's twin earth argument is a special case of Burge's, concentrating on the particularly clear case of natural kind terms and the importance of scientific opinion in determining their reference and, thereby, meaning (cf. Pettit & McDowell, 1986).

This reading of the arguments for broad content has obvious attractions for cognitive semanticists. Meanings are still analysed in terms of cognitive states, albeit those of particular subsets of the community, possibly supplemented by some account of the source of each sub-community's expertise. This is certainly the interpretation entertained by Johnson-Laird (1983, pp.191–195) and is explicitly adopted by Gardenfors (1993). Gärdenfors' model is particularly important here as its principal concerns are similar to those of this thesis. Gärdenfors carefully builds up a series of formal structures which capture various possibilities for the distribution of "linguistic power" within a society. In outline, L is the set of atoms of a language (sentences or predicates), M is a set of meanings (propositions or concepts) for the language and U is the set of speakers/users. For each individual $i \in U$, there is an individual semantics, m_i, mapping from L into M. There is also a distinguished mapping m_s, the *social semantics*, from L into M.[9] A semantic

[9]Gärdenfors' reliance on set theory and model theory has the consequence that the set of meanings, M, is fixed for a particular language. For example, in the case where M is a Boolean

situation S is a set of individual mappings, one for each $i \in U$. By defining conditions on these structures, namely, compsitionality, contingency-preservation and designators for social meaning, Gärdenfors derives an interesting representation of alternative possible social power structures that determine m_s. For example, the set $D \subset U$ of individuals is decisive for L in S if it holds that $m_s(a) = m_i(a)$, whenever $m_i(a) = m_j(a)$ for all $i, j \in D$. Mathematically, D forms a *filter* which, depending on various other parameters, such as whether U is finite, can model various degrees of democratic and oligarchical control on m_s. Importantly, while no specific individual necessarily determines the entire mapping m_s, for all $a \in L$ there is some m_i such that $m_s(a) = m_i(a)$. Less esoterically, for any given expression and its associated social meaning, some individual will determine that meaning.

This analysis does not, however, succeed in meeting the problems raised by Putnam's twin earth argument. As noted in section 2.6, there may not be *any* individuals in a community whose cognitive states accurately determine the actual extension of a given natural kind term. Science may be insufficiently advanced to make the necessary discrimination between H_2O and XYZ. Nonetheless, given that there is some, in principle discoverable, difference between the essential nature of water on earth and that on twin earth then "water" refers to, and thereby means, different things in each case. Johnson-Laird and Gärdenfors (*ibid.*) are both aware of this difficulty and try to meet it by challenging the validity of essentialism. Johnson-Laird raises the (somewhat bizzare) possibility that it may transpire that nothing actually does have an essential nature: perhaps we are all subject to a profound cartesian delusion. Assuming this idea could be coher-

algebra it is taken to be the powerset of the finite set of fixed objects $O = o_1, o_2, o_3 \ldots$. Agents vary in the assignment of particular meanings to particular expressions but not on the set of meanings. Clearly, this runs contrary to the arguments of section 2.4 as it constitutes an assumption of a unique, fixed, ontology of concepts. However, for current purposes, the main interest is in the assumptions relating to the distribution of content.

ently formulated, it would follow that essentialism about meanings was wrong, presumably leaving room only for cognitive accounts of meaning. Gärdenfors attacks essentialism on the grounds that many putative natural kind terms such as "phlogiston" and "caloric" have proved to be without any essential nature. Pressing a weaker claim than Johnson-Laird, Gärdenfors argues that we have no reliable method of distinguishing between 'genuine' natural kind terms, historical aberrations and non-natural kind terms. Without independent justification for the existence of categories which actually do have an essential nature in common, Putnam, according to Gärdenfors, is assuming the very thing he wishes to show.

However, while there may well be legitimate concerns about the viability of essentialist metaphysics, the point on which the difficulties concerning content turn is that, rightly or wrongly, ordinary practice *presupposes* essentialism about the meaning of at least some words, those that Putnam calls natural kind terms. In the absence of some argument which accounts, in non-essentialist terms, for the intuiton that the meaning of "water" is different on earth and on twin earth the difficulty stands. Furthermore, it is not clear that the problems are restricted to cases that trade on essentialist intuitions. Burge's thought experiment suggests an additional problem. Intuitively, some terms, which are neither natural kind terms nor the province of some group of experts, may still be used with the intention that their actual content is fixed by the community in which they have their currency. One candidate, raised in section 2.1.1, is "friend". Certainly, there are neither experts to whom we, as a community, defer for definitions of what a friend is, nor some defining essential nature we suppose that all friends share. Individuals may, paralleling Burge's scenario, harbour various misconceptions about what constitutes a friend and, as noted above, it is entirely possible that no two people will have exactly the same concept of a friend. Nonetheless, we still use "friend" normatively, intending it to mean what it normally means in the community, i.e., we would, under appropriate circumstances, accept correction. Examples such

as these are problematic for cognitivist accounts. It would seem more promising to propose that the meaning of "friend" is somehow constituted by the relevant practices and utterances of the linguistic community as a whole.

2.7 Discussion

This chapter began with an attempt to establish that idiolectical variation should be understood, at least in part, as a semantic phenomenon and that as such it poses important problems for any attempt to provide a semantic model for dialogue. In Chapter 1 a number of alternative semantic models were discussed that, it was claimed, assume that mutual-intelligibility depends on the existence of some shared semantic code. Given the arguments in the first half of this chapter, it seems that this assumption must be abandoned in favour of accounts which do not imply that individuals are semantically transparent to each other. Specifically, an adequate semantic model of dialogue must be able to accommodate a degree of ontological pluralism between interlocutors. A promising framework for dealing with this concern is offered by cognitive semantics where a great deal of attention has been paid to factors that may differentially influence each individual's interpretation of particular expressions. Furthermore, these models are noticeably non-commital about exactly how similar different individuals' conceptual structures should be in order for communication to be possible (see section 1.3.1), holding out the prospect that, with some modification, these accounts could be brought to bear on situations in which interlocutors with divergent ontological commitments are engaged in dialogue.

The arguments considered in the second half of this chapter, however, raise important difficulties for the cognitive approach. Theories which attempt to naturalise meaning by reduction to cognitive states, whether those of an individual or groups of individuals, appear to violate important intuitions about the nor-

mative, distributed, nature of meaning. Burge, in particular, argues that this fact, combined with the normal practice in both 'folk' and cognitive psychology of individuating mental states with respect to the content of attitude attributions places the intentionality of mental states beyond the reach of any individualistic explanation.[10] The non-individualistic nature of mental states is, on this account, a corollary of the distributed socio-physical influences on semantic content. For the purposes of this chapter, the important point is that the semantics of natural language expressions cannot be naturalised by reduction to mental states, individualistically understood. Attempts to do so may actually fail to provide a theory of meaning at all (cf. Putnam, 1988). Discussion of the consequences this has for the intentionality of cognitive states is deferred until chapter 6.

The contribution of idiolectical variation to the arguments for distributed content is salient. A critical point of departure for Putnam's and Burge's views is the existence of conceptual differences between individuals. In Burge's case the role of idiolectical variation is quite clear. In Putnam's case it is less obvious since in the twin earths scenario we are invited to consider a situation in which individuals with identical narrow concepts of "water" may nonetheless mean different things by their utterances about it. However, it is only in order to demonstrate the inadequacy of this account in meeting certain intuitions about meaning that he sets up the hypothetical identity in the twin's cognitive states.[11] In fact, Putnam's views, particularly those concerning stereotypes, depend on the assumption that idiolectical variation is the norm (see section 2.1.1). Putnam (1988) emphasises that it is precisely because we normally discount differences in belief, or even definition,

[10]This threat is sufficiently serious for some authors to adopt an eliminativistic stance toward the intentional, regarding it as a consequence of inadequate 'folk' theories that should form no part of a scientifically respectable study of cognition (e.g., Churchland, 1981).

[11]Burge (1979) interprets Putnam's emphasis on individualistic mental states to be a point of departure from his own account. However, Putnam (1988) emphasises that this was for purely exegetical reasons and should not be read as an endorsement of individualisitic conceptions of the mental.

when assessing whether a term means the same thing on different occasions of use that the cognitive states of speakers are, to a certain extent, irrelevant to the fixation of semantic content. Normal interpretation proceeds via the application of a principle of charity which attempts, all things being equal, to hold meanings constant in the face of synchronic and diachronic variation.

This chapter has been devoted to the theoretical importance of idiolectical variation to semantic theory. However, it also presents an important practical problem which individuals must somehow routinely solve in the course of interaction. The next chapter turns to consideration of the extent to which empirical theories of dialogue can account for the achievement of semantic coordination between interlocutors.

Chapter 3

Empirical Models of Dialogue

In Chapter 1, it was noted that while there are a number of semantic theories which aim to characterise *discourse*, few, if any, can be held to address themselves directly to dialogue *per se*. The tendency to assimilate dialogue to monologue, concentrating on the problems associated with the sequential, rather than inter-individual, coherence of discourse is paralleled in psycholinguistics (cf. Clark, 1985). With some notable exceptions, discussed below, there is a paucity of studies directly concerned with issues related to the inter-individual coordination of dialogue. The majority of research on the psychology of language has concentrated on processes of production and interpretation in individuals effectively isolated from ordinary conversational context. Examples of work in this tradition include studies on lexical semantics (e.g., Collins & Quillian, 1969; Morton, 1970; Rumelhart & McClelland, 1986), higher-order theories of scripts and schemas (e.g., Bower, Black, & Turner, 1979; Minsky, 1975; Schank & Abelson, 1977) and discourse coherence (e.g., Anderson, Garrod, & Sanford, 1983; van Dijk & Kintsch, 1983). Amongst the empirical research that does address issues of inter-individual coordination, a significant proportion has been concerned with the influence of factors such as gender, dominance and other personality traits on behaviours such as interruptions,

frequency of eye contact and persuasiveness (e.g., Linkey & Firestone, 1990; Roger & Nesshoever, 1987; Kleinke, 1986). While these issues clearly do bear on the effective conduct of face-to-face communication, they do not appear to bear directly on the essentially semantic issues under consideration here. The intuition is that idiolectical variation would create independent problems for mutual-intelligibility even where non-verbal signals are factored out (e.g., in telephone conversations) and personalities, assuming it were possible, are matched.

This aim of this chapter is to survey a range of empirical studies which explicitly raise the problem of inter-individual coherence in dialogue and promise some insight into the mechanisms and processes that contribute to the coordination of meaning in dialogue. For exegetical convenience, the studies that fall under this rubric can be divided into two broad approaches: those that utilise existing philosophical and linguistic analyses in formulating empirical questions, roughly, 'theory-driven' approaches, and those that concentrate on the empirical phenomena in their classification and theorising of dialogue, roughly, 'data-driven' approaches.[1]

3.1 Theory-Based Approaches

An ostensibly promising area for consideration is Speech Act theory and its variants (eg., Searle, 1969, 1976). It offers an analysis of the ways in which utterances can constitute actions, actions whose performance sometimes depends on appropriate relations between speaker and addressee. For example, a successful bet requires, amongst other things, that both parties mutually accept their undertaking. Despite its elegance, Austin's analysis is not always regarded as generating unequivocally empirical predictions. In fact, Levinson (1983) goes so far as to suggest that most analyses that elaborate Austin's insights have proved to be un-

[1]Naturally, this is a distinction in emphasis rather than in principle.

falsifiable and therefore vacuous. Levinson's basic complaint is that, in general, subsequent authors have attempted to develop 'well-formedness' conditions on sequences of speech acts or conversational moves. A precondition for achieving this is the specification of some determinate procedure for identifying what act a particular utterance performs or is a response to. However, there is a high degree of indirection in mapping from utterance form to speech act type. Utterances only rarely contain an explicit performative such as "I request you ..." or "I order you ..." and a particular utterance may count as an attempt to perform any of several possible direct illocutionary acts. For example, "I will return" may function equally well as a promise, a warning or a prediction (cf. Sadock, 1988). Indirect speech acts compound these difficulties by requiring the specification of the inferences that bridge between an utterance of "It's cold in here" and the implied request that the addressee closes the window. Additionally, responses may address the *perlocutionary* rather than illocutionary force of an utterance. As a result, the illocutionary type of an utterance in a given sequence and, thereby, its 'grammaticality', can only be judged by reference to the linguistic and extra-linguistic context. Further difficulties are created by the normative, Gricean, nature of utterance interpretation. An apparently 'ungrammatical' or 'ill-formed' sequence of utterances is likely to be deemed an intentional exploitation of conversational conventions in order to achieve effects such as irony and other tropes (e.g., see Clark, 1985). As a result, there appears to be no principled way of discriminating 'grammatical' from 'non-grammatical' sequences in order to assess the viability of the empirical claims.

Taking a different approach, Clark (1979) reports a series of experiments aimed at elucidating some of the cues individuals use in judging which to respond to of the direct and indirect illocutions associated with an utterance. For example, Clark identifies several factors that influence whether an addressee will respond to the direct illocutionary act of a request such as "Can you tell me the time?" by

asserting "Yes" they can *and* that "It's six" or whether they respond only to the indirect request with "It's six" alone. One cue to which people show sensitivity is the relative 'conventionality' with which the question is asked. The two questions below make what is, strictly speaking, the same basic request but differ in their conventionality. Intuitively, question 2 is a less common formulation than question 1.

1. "Can you please tell me what the interest rate is on your regular savings account?"

2. "Are you able to tell me what the interest is on your regular savings account?"

Clark predicts that the less conventional the presentation of the question, the more likely the direct request will be understood as seriously intended and, as a result, elicit an answer. Indeed, out of a total sample of 150 bank clerks, 92% of those asked question 1 responded to the indirect request alone whereas only 64% of those asked question 2 did. There is also evidence that addressees make inferences concerning a speaker's likely goals in assessing whether an illocutionary act is intended *pro forma* or more seriously. For example, 50 restaurants were phoned and asked "Do you accept credit cards?" or "Do you accept American Express cards?". In answer to the latter question 100% of those who did accept American Express cards replied "Yes", but offered no further information. By contrast, of those who answered yes to the former question 46% also offered a list of cards that they accepted. The restaurateur makes the natural inference that the caller is making their inquiry because they actually intend to pay, at the restaurant, with a specific credit card.

Although Clark's results reveal interesting patterns in the way people determine what is being asked of them in a conversation, it is less clear whether they speak directly to the specific claims of speech act theory. They do not address (and

are not designed to address) the difficulties raised by Levinson, since the classification of utterances as direct and indirect requests and their respective responses is made on intuitive rather than formal grounds by individuals fully apprised of the context in which they occur. A more secure interpretation would be to treat the data as shedding light on general patterns of conversational inference rather than providing specific support for speech act theory itself.

An empirical proposal which relates to the informational coherence of a dialogue is the given-new distinction (Clark & Haviland, 1977; Halliday, 1967; Prince, 1969). This distinction was first discussed in detail by Halliday (1967) as a means of characterising the information structure of an utterance as signalled by its pitch contour. Briefly, 'new' information is taken to be marked by the principal pitch focus in an utterance, which determines its rightmost element. By contrast, 'given' information is intonationally unmarked and corresponds to information that the speaker presents to their interlocutor(s) as already shared. This definition leaves the leftmost edge of the 'new' information unit undefined, making the precise division between 'given' and 'new' in an utterance difficult to determine. The scare quotes around given and new emphasise the fact that, as Halliday defined them, they refer to information presented *as* given or new, thereby encompassing cases where information is presented as new (or given) for essentially rhetorical purposes.[2] Another approach to informational asymmetries between conversational partners, which relates to the semantics of quantifiers, is found in Moxey and Sanford (1993) who demonstrate that quantifiers play an important 'rhetorical' role in manipulating focus in a discourse. Amongst their findings is evidence that individuals deploy quantifiers in a manner which is sensitive to their beliefs about their interlocutor's expectations.

[2] As Humphreys (1993) points out, this and several other aspects of the distinctions emphasised by Halliday, for example between given-new, theme-rheme and topic-comment have tended to be conflated by later discussions.

Each of the models discussed so far offers proposals concerning the way in which interlocutors present and deploy utterances in a manner that influences the inter-individual coherence of interaction. They also suggest ways in which these factors may influence or alter interpretation and, in this sense, impact on the meaning of what is said. However, the processing of elements of an utterance as given or new and the influence of quantifier choice in manipulating focus, both presuppose that some roughly 'literal' degree of interpretation has been achieved before they can produce their various effects. Similarly, determination of what speech act is being performed by an utterance speaks more to pragmatic than to semantic concerns and, as a result, does not directly impact on the issue of idiolectical variation. Again, the intuition is that even where these 'rhetorical' aspects of language-understanding are factored out, there will still be a residual, semantic, problem posed by interpretational asymmetries.

The research discussed so far is inspired, to a greater or lesser degree, by some of the theoretical approaches developed within linguistics and philosophy. There are some notable advantages to investigating the fit between pre-existing frameworks or taxonomies and empirical phenomena, not least because it often provides a relatively explicit specification of various dependencies within a model and, more diffusely, provides a basis for inter-disciplinary interaction. However, the theory-inspired nature of this research, especially in the case of Austinian approaches, has also been criticised as a weakness. Levinson (1983) argues that the attempt to provide a 'syntax' of successive turns in conversation is "fundamentally inappropriate to the subject matter" (p.289). For Levinson, the mapping of linguistic categories and methods of analysis onto conversational data is undermotivated, encouraging research which overlooks important aspects of conversational organisation. Schegloff (1992) adopts a more radical stance, insisting that Speech Act theory and, indeed, any theory in the cognitive/analytic tradition provides at best a superficial, and at worst a completely inadequate, analysis of conversational

data.[3] (see also Heritage, 1984). In the absence of specific explanations that offer more complete or more robust accounts of the experimental findings, it is difficult to justify dismissing the research reviewed above on these grounds, not least because, bearing in mind the arguments of section 2.3, Schegloff's position seems vulnerable to the criticism that no empirical observations are free from theoretical commitments. Nonetheless it is apparent that in psycholinguistics as a whole, too little attention has been paid to analysing the conduct of dialogue.

3.2 Data-Driven Approaches

Probably the most resolutely empirical approach to the study of dialogue is conversation analysis (henceforth, CA) which has its roots in the ethnomethodological tradition in sociology. Many of the distinctive commitments of CA derive directly from ethnomethodology, the tradition in which it arose, and it is consequently important to place CA within this context.

3.2.1 Ethnomethodology

The term ethnomethodology, coined by Garfinkel in the 1950's, was devised with the intention of providing a cognate to terms such as ethnobiology and ethnomedicine. It designates the study of the 'folk' methods by which individuals reason about, and make sense of, everyday problems (Heritage, 1984).

Garfinkel's position can be most clearly articulated as a reaction to Parsons' "voluntaristic theory of action" (see Heritage, 1984; Taylor, 1992). Parsons' work

[3] e.g., Schegloff (1992) " [...] speech act theory is [...] an analytic resource that in effect casts action as atomistic, individualistic, atemporal, asequential, and asocial" (pp.1338-1339). Schegloff, by contrast (*ibid*) advocates analysis of the "...procedural infrastructure of interaction" (p.1338) concluding that "...one should ask what grounds there are for continuing to take seriously theories whose analytic center of gravity is located elsewhere" (p.1339).

aims to develop a theory of social order, founded on scientific methods, that accounts for the resilience of socio-cultural institutions in the face of apparently divergent individual interests. A key element of Parsons' theory is its appeal to internalised 'social norms' as the causal determinants of individual action. An individual, through learning and experience with particular cultural institutions, internalises certain rules that govern their subsequent behaviour. Taylor (1992) illustrates this with the example norm: "wear a tie on formal occasions" which, in order to regulate the behaviour of males in a given society, would need to be internalised, independently, by each individual. There is a sense in which this example can be misleading. Parsons' concept of social norms was strongly influenced by the work of Durkheim and Freud (Heritage, 1984). Consequently, the type of norms which Parsons held to be determinants of behaviour are not conceived of as something of which an individual would ordinarily have cognisance. Like the Freudian notion of a complex, they are considered to be deeply buried psychologically, with the actor whose behaviour they determine having little or no insight into their operation. In the Parsonian framework, norms are externally defined theoretical entities elucidated by the methods and procedures of social science and are only accidentally apprehended (if at all) by the individuals subject to them.

Garfinkel objected to the conception of norms as hidden, causal determinants of behaviour. Drawing on the phenomenology of Schutz (1973), and detailed empirical studies of his own (e.g., Garfinkel, 1967), Garfinkel argued that, far from being opaque to the individuals influenced by them, social norms are directly utilised in accounting for and making sense of actions and activities. For Garfinkel, the social phenomena of interest to sociologists are precisely those which the participants themselves frame and interpret in intentional, meaningful, terms. Parsons's 'cultural dope' view of individual agency effectively relegates these aspects of individuals' reasoning about the social to a residual category of epiphenomena. Garfinkel, by contrast, insisted that it is precisely the processes of everyday ac-

counting for, and reasoning about, the social which stand in need of investigation. For example, Garfinkel (1967) highlighted how inadequate the Parsonian approach is for analysing the processes by which a jury agrees on a verdict. The majority of jurors' deliberations involve determining, in their own terms, what is fact or fancy, what actually happened and what merely appeared to happen, what is credible and what is contrived and so on. As Heritage (1984) points out, an account that aims to elucidate the norms which determine the jurors' deliberations while discounting the juror's own interpretations of the situation as irrelevant fails to address a significant, if not critical, aspect of the situation.

In emphasising the role of individuals' interpretation of their circumstances, ethnomethodologists have drawn attention to the way in which norms themselves can be used as resources for generating some particular understanding of a situation. Returning to Taylor's example, rather than viewing the norm, "wear a tie on formal occasions" as a descriptive or regulative rule, ethnomethodologists emphasise the way in which individuals may use such a norm in order to constitute some situation as a formal occasion. As Heritage puts it:

> "... the basic relationship between normative rules and socially organised events appears to be a strongly cognitive one in which 'rules' (concertedly applied) are *constitutive* of 'what the events are', or 'what is really going on here"' (1984, p.83).[4]

Thus the injunction to "wear a tie on formal occasions" might well be deployed as an indirect way of informing someone that he was at a formal occasion, even though he didn't consider it such, or perhaps that the tie he had on was not, under the circumstances, a tie but rather an indecorous eyesore (see also Wieder, 1974).

[4]It is worth noting that the cognitive turn in Heritage's and other interpretations of ethnomethodology is not uncontroversial (see e.g., Button & Sharrock, 1994).

In summary, ethnomethodology is the study of practical sociological reasoning emphasising the ordinary, situated, particulars of everyday talk and conduct. It aims to avoid 'premature' theorising by concentrating on the detailed empirical analysis of how individuals orient to, and make sense of, social activity. Against this background CA has developed as a strongly empirical approach to analysing the 'lay' practices and procedures through which the intelligibility of interaction is maintained.

3.2.2 Conversation Analysis

The basic objective of CA is to identify, and formally describe, the structural or procedural organisation of interaction and, thereby, the competences on which individuals rely when they engage in conversation (Drew, 1990; Heritage, 1984; Sacks, 1984). Reflecting the influence of ethnomethodology, the principal validation for any proposed analysis is evidence that conversational participants actually do orient to the proposed structure in making sense of their interaction. As a result 'conventional' categories of analysis, such as illocutionary acts and mixed versus single sex dialogue, are discarded unless conversational participants themselves can be shown to be sensitive to such distinctions. The extremely large literature that now exists within CA renders it impossible to provide a representative survey of findings in the space available.[5] The selection that follows is therefore determined principally by its relevance to the concerns here.

The goal of avoiding premature theorising has meant that the principle data for CA are detailed transcripts of naturally occurring conversation, usually telephone calls. The emphasis on 'natural' conversation extends to a rejection of experimental manipulations such as structured interviews and the production of invented examples in order to investigate intuitive judgements of acceptability. It

[5] Wooffitt (1990) mentions a recent bibliography of over twenty pages covering several hundred entries.

also focuses attention on 'mundane' conversations, avoiding academic discourse and other idiosyncratic contexts. An important methodological commitment of CA is that no detail of conversation, including pauses, coughs and other apparent 'disfluencies' can be dismissed *a priori*. As a result, the transcriptions that form the basic data for CA are extremely detailed. For reasons of clarity much of this detail is omitted from the examples that follow.

The basic unit of analysis in CA is the turn, identified on the basis of linguistic surface structure, prosodic and intonational cues (Sacks, Schegloff, & Jefferson, 1974). A speaker is assigned a turn construction unit (TCU) at the end of which there is a transition relevance point (TRP). In order to account for the smooth distribution of turns in conversation a set of rules are proposed which characterise transition between speakers (Sacks, Schegloff & Jefferson, 1978, modified by Levinson, 1983). If C is the current speaker and N is the next speaker at a TRP:

1. (a) If C selects N in current turn, then C must stop speaking, and N must speak next, transition occurring at the first TRP after N-selection.

   (b) If C does not select N, then any (other) party may self-select, first speaker gaining rights to the next turn.

   (c) If C has not selected N, and no other party self-selects under option (b), then C may (but need not) continue (i.e., claim rights to a further turn-constructional unit).

2. When rule 1(c) has been applied by C, then at the next TRP Rules 1 (a)-(c) apply, and recursively at the next TRP until change of speaker is effected.

These rules successfully accommodate the observation that only 5% of speech in conversation overlaps. They predict that only one person should speak at a time and where overlap does occur it should be restricted principally to competing

starts as in 1 below or misprojected TRPs as in 2 below. Where competing starts occur, one speaker will drop out rapidly and the one left in the clear will recycle that part of the turn obscured by the overlap.[6]

1. J: Twelve pounds i think wasn't it=
   D: =//Can you bel*ieve* it?
   L: Twelve pounds on the weight watchers scale

2. A: Uh *you* been down here before // havenche
   B: Yeah

The rules also offer a way of discriminating deliberate from accidental interruptions, as in example 3, where the overlapping speech does not occur at a TRP.

3. C: We:ll I wrote what I thought was a a-a rea:s'n//ble explanation
   F: I: think it was a *very* rude *le:*tter

They can also discriminate between silences which are treated as gaps and silences which are understood as significant or attributable because another speaker has been selected under rule 1(a), for example:

[6]Notation: "=" indicates no discernible gap between utterances, "//" indicates the point at which the next utterance overlaps the current utterance. ":" and "::" indicate lengthening of the preceding vowel sound. Italics indicate a word, or part of a word, uttered with extra emphasis. Numbers in brackets, such as (1.0), give the time elapsed in seconds.

4. A: Is there something bothering you or not?
   (1.0)
   A: Yes or no?
   (1.5)
   A: Eh?
   B: No.

The rules provide a skeleton framework around which the local, turn by turn, organisation of conversation emerges. Importantly, this organisation is anchored to the 'surface structure' of a turn, not functional or conceptual units. As a consequence the rules are held to operate irrespective of content or length of a turn and are independent of the number of possible interlocutors. In keeping with the precepts of ethnomethodology, the examples suggest that individuals are indeed sensitive to the organisation they characterise.

A series of more sophisticated proposals builds on the basic framework proposed for turn-taking. Sacks et al. (1974) address the relationship between pairs of utterances such as question-answer and offer-acceptance which are drawn together under the notion of *adjacency pairs*. These are pairs of utterances that are a) produced by different speakers, b) ordered as a *first part* and a *second part* and c) typed, so that a particular first part requires a particular type, or range of types, of second. Complementary first and second pair parts need not occur as immediately adjacent turns. For example, on recognising a first pair part the next speaker may respond with another first pair part, delaying completion of the 'prior' pair until later in the conversation. This behaviour gives rise to nested sequences of adjacency pairs such as those in example 5.

5.	A:	Can I borrow the car?	(Q1)
	B:	How long do you need it?	((Q2))
	A:	For a few hours	((A2))
	B:	Sure	(A1)

An adjacency pair which is nested in this manner is referred to as an insertion sequence (Schegloff, 1972). There is no hard limit to the number of levels of nesting which may occur with the consequence that an adjacency pair may be separated by a large number of intervening utterances (Levinson, 1983, offers some examples).

Consonant with Garfinkel's conception of norms, the relationship between first and second pair parts is understood in normative rather than regulative terms. The production of a first pair part makes the production of the second pair part conditionally relevant rather than 'grammatically' necessary. As a result, whatever follows a first pair part is interpreted as relevant to it, even though it might not strictly count as completion. Even the absence of a second, for example, silence in response to a question, may be interpreted as significant, perhaps leading to a restatement of the original question.[7] Adjacency pairs therefore do not constitute statistical generalisations or conditions on 'well-formed' discourse.

On its own, this formulation is quite weak, most first pair parts take a broad range of second pair parts. For example, a question may receive, amongst other things, a protestation of ignorance, a 're-route': "Better ask John", a refusal to answer or a challenge to its presuppositions. As a means of strengthening the generalisation, the notion of a preference organisation is invoked (Sacks & Schegloff, 1979). This allows an ordering of possible second pair parts according to whether they are *preferred* or *dispreferred*: e.g., the preferred second to a request is an acceptance, the Dispreferred a refusal. Disprefered seconds display a number of

[7]In fact, conditional relevance can also project back from the second pair part to the first pair part as, for example, where someone utters, "Oh you're welcome", after holding the door open for a stranger who has not thanked them.

systematic differences from preferred seconds. Their production is usually relatively delayed, prefaced by some marker, such as, "well", and is often accompanied by an account. Importantly, this is a structural rather than a psychological notion of preference:

> "Preference here does not refer to any personal psychological or motivational dispositions of individual speakers. It refers instead to the finding in CA research that these alternative actions are routinely performed in systematically distinctive ways" (Drew, 1990, p.14)

This brief selection of findings from CA serves to illustrate some important properties of the approach. Its goal is to generate substantive generalisations about the procedural organisation of ordinary conversation. The resulting structures, such as adjacency pairs, are defined independently of the content of the utterances to which they apply, relying, instead, on a taxonomy derived from the patterns of organisation to which they correspond. This has two consequences for the concerns here. Firstly, it is not in the spirit of CA, and may well be incoherent, to treat these proposals as claims about cognitive processes that underpin the interpretation of utterances, although, like the other frameworks discussed above, they obviously are relevant to interpretation in some sense. As the quotes from Schegloff and Drew suggest, there is no attempt, and apparently no desire, to provide an account that directly impacts on claims about cognitive structures. Secondly, CA offers generalisations that idealise over individual differences, including idiolectical variation, and, on face-value, this focus renders it neutral with respect to the principal concern of this thesis. The issue of how mutual-intelligibility is addressed in ethnomethodology will be returned to in Chapter 6.

3.2.3 The Collaborative Model

Clark and coworkers (Clark, 1993; Clark & Schaefer, 1989; Wilkes-Gibbs & Clark, 1992; Clark & Wilkes-Gibbs, 1990; Schober & Clark, 1989) have elaborated an empirical model of dialogue which draws on the basic framework developed in conversation analysis but is aimed at explicitly psychological and experimental concerns. The collaborative model adopts a central metaphor of multi-party discourse or conversation, as a concerted, collective activity. Rather than treat dialogue as the simple 'sum' of two autonomous activities, speaking and listening, they focus on the collaborative nature of conducting a discourse. This leads to a distinction between two basic types of action, joint or collective actions, performed by an ensemble of people, and individual actions. (e.g., Clark, 1993; Clark & Schaefer, 1989; Schober & Clark, 1989). The distinction is usually illustrated by reference to paradigmatic cases of collaborative activities such as dancing, playing a duet or a game of football (c.f. Searle, 1990). In general, we intuitively recognise a class of collaborative or joint acts which are performed by an ensemble of people: it takes two to Tango. However, Clark and coworkers also advance a stronger thesis, namely, that collaborative acts cannot be analysed by reduction to chains of individual, autonomous, acts. Instead, joint actions must be further subdivided into *autonomous* actions which are performed independently of other people and *participatory* actions which are perfomed in collaboration with others.

The claim is that dialogue is not simply a sequence of utterances where each can be understood as an autonomous act. Instead, it is a series of participatory acts that are constituted by reference to the joint activity of which they are part. Placing appropriate restrictions on context, the physical description of someone's finger depressing a particular key on a keyboard is identical whether the key is pressed as part of a duet or a solo. Clark's claim, however, is that under (at least) some descriptions important to the analysis of dialogue these two actions are not equivalent: they are different acts, involving different intentions (Clark,

pers. comm.). For example, "I intend to play C" versus "I intend to play C as part of my C plus your E". Clark and Schaefer (1989) propose that while many aspects of an utterance, for example, words and phonemes, can be successfully analysed as autonomous acts their function in a discourse demands analysis as participatory acts. On this model, analysis of any behaviour can proceed at a number of levels (Clark, 1993). Adapting a framework developed by Alvin Goldman, Clark proposes that participatory acts should be understood at four levels:

1. Vocalising and Attention.
2. Presentation and Identification.
3. Meaning and Understanding.
4. Proposal and Consideration.

Each level of action is cotemporal and may be performed by the same utterance token; however, for an utterance to count as an act at any given level presupposes that it is also an act at the lower, in terms of the numbering, level. A question may simultaneously perform all four of these acts. It acts as an appeal for the interlocutor's attention, it is an utterance, it has a particular meaning and it makes a request of the interlocutor. If the vocalisation does not gain the attention of the addressee (level 1) then it obviously fails to qualify as the presentation of an utterance (level 2), is not understood and makes no request of the interlocutor (levels 3 and 4). Alternatively, a vocalisation may be identified as an utterance (level 2) but the meaning may not be understood, perhaps it is delivered in an unfamilair language (level 3) and, again, thereby fails to qualify as a request (level 4).

The viability of the distinction between participatory and autonomous acts is not generally defended on conceptual grounds. Instead, the principal support for the model comes from a series of experimental studies on task-oriented dialogue

which aim to challenge a particular view of the means by which individuals accumulate common ground. This view can be summarised as three assumptions (Clark & Schaefer, 1989):

1. *Common ground:* The participants in a discourse presuppose a certain common ground.

2. *Accumulation:* In the course of a discourse, the participants try to add to their common ground.

3. *Unilateral Action:* The principal means by which the participants add to their common ground is by the speaker uttering the right sentence at the right time.

It is the third assumption, which treats utterances as autonomous acts, that is contested. Tacitly, it implies that an utterance, once made, is automatically added to the common ground. A listener essentially decodes an utterance and interprets it against the current common ground. By contrast, Clark and Schaefer claim that interlocutors can only add to the common ground through a collaborative process they term contributing. Structurally, a contribution consists of a presentation phase and an acceptance phase:

- *Presentation Phase:* A presents utterance u for B to consider. He does so on the assumption that, if B gives evidence e or stronger, he can believe that B understands what A means by u.

- *Acceptance Phase:* B accepts utterance u by giving evidence e' that he believes he understands what A means by u. He does so on the assumption that, once A registers evidence e', he will also believe that B understands.

A contribution is completed once A and B mutually believe that B understands what A meant by his presentation. Completion can only be determined

retrospectively once B has demonstrated acceptance and it does not draw further comment from A. Any directed signal, at any level, is treated as a presentation for acceptance by the interlocutor. Consequently, B's acceptance is simultaneously a presentation to A which, in turn, requires acceptance before it is added to the common ground. As described so far, the cycle of presentations and acceptances could go on indefinitely. However, the types of evidence for acceptance of a presentation are ordered according to their strength. From weakest to strongest these are:

1. *Continued attention:* B shows he is continuing to attend and therefore remains satisfied with A's presentation.

2. *Initiation of next relevant contribution:* B starts on the next contribution that would be relevant at a level as high as the current one.

3. *Acknowledgement:* B nods or says "uh hu," "yeah" or the like.

4. *Demonstration:* B demonstrates all or part of what he has understood A to mean.

5. *Display:* B displays verbatim all or part of A's presentation.

This ordering combines with a strength of evidence principle which governs the degree of evidence appropriate in accepting various presentations:

> "The participants expect that, if evidence e_0 is needed for accepting presentation u_0, and e_1 for accepting the presentation of e_0, then e_1 will be weaker than e_0." (Clark and Schaefer, 1989, p.268)

This principle ensures that the alternation of presentations and acceptances 'bottoms out' since the evidence required to establish mutual belief of an acceptance is always weaker than that required for the presentation it is designed to address.

This basic structure can build up into quite elaborate sequences of embedded presentations and acceptances. A given contribution may span several turns, for example when an initial presentation needs to be repaired (cf. insertion sequences). Conversely, a single turn may correspond to several contributions, with backchannel responses acting as signals of acceptance within a turn. Like adjacency pairs, a presentation is viewed as projecting for its acceptance with continued attention following an utterance, all things being equal, acting as a signal of acceptance rather than a noncommital pause. The notion of contribution thus imposes the constraint that there must be *positive* evidence that all parties have reached the mutual belief, at some level, that a presentation has been accepted before it is accumulated to the common ground between them at that level.

The supporting evidence for this model derives from a series of experiments (e.g., Clark & Wilkes-Gibbs, 1990; Wilkes-Gibbs & Clark, 1992; Schober & Clark, 1989) examining its predictions for the conduct of task-oriented dialogues. The typical format, based on a paradigm introduced by Krauss and Glucksberg, involves a director, who has a set of twelve tangram figures arranged in a target sequence in a numbered grid, and a matcher, who has the same grid and set of figures but arranged in a random order. The task is for the director to communicate to the matcher the desired order of the figures. Both individuals are separated by a screen and must therefore achieve this through the production and interpretation of referring expressions that efficiently discriminate between the different figures. Over the course of a number of trials, each with the same director, matcher and figures but different target sequences, a regular pattern emerges. For example, Clark and Wilkes-Gibbs (1990) report a regular decrease in both the number of words used in identification of each figure the number of turns taken. Over the course of six trials the average number of words per figure falls from 41 to 8 and the average number of turns from 3.7 to 1. Thus, a description which is initially complex, e.g., "looks like a person who's ice skating, except they're sticking two

arms out in front" becomes progressively contracted to "the ice skater". This decline is reflected by shifts in the type of noun phrases used to pick out each figure. For example, episodic noun phrases which consist of several separate clauses, e.g., "the goofy guy that's falling over, with his leg kicked up", and provisional noun phrases which are subject to unprompted elaboration, e.g., "the next one is also the one that doesn't look like anything. It's kind of like a tree?", both fall across trials. By contrast, elementary noun phrases consisting of a single, unrepaired, clause increase across trials. To account for this pattern (Clark & Wilkes-Gibbs, 1990) cite a principle of least *collaborative* effort:

> " speakers and addressees try to minimise *collaborative effort*, the work both speakers and addressees do from the initiation of the referential process to its completion." (Clark & Wilkes-Gibbs, 1990, p.486. Original emphasis)

This is contrasted with a principle of least *autonomous effort*, derived from the three assumptions mentioned above, which predicts that individuals should produce referring expressions that are sufficient to uniquely identify the referent in context. Clark and Wilkes-Gibbs argue that only the principle of least collaborative effort predicts the observed decline in complexity of referring expressions. On the autonomous account, once an appropriate expression has been found by an individual it should not receive further modification. The principle of least collaborative effort predicts that referring expressions will continue to contract to beyond the point where they are adequate in context. To some extent this explanation trades on an ambiguity in the notion of context. If the autonomous models are modified to include, for example, details of the dialogue history with a particular addressee as part of the relevant context, then the predictions converge since the least autonomous effort with a particular speaker is not equivalent to least autonomous effort *per se*. Of course, to a degree, this broadening of the notion

of context is just what the collaborative model aims to make explicit; however, it doesn't necessarily require appeal to the process whereby participants actively work toward the achievement of mutual-belief.

Stronger evidence for this aspect of the collaborative model is provided in a study by Wilkes-Gibbs and Clark (1992). Using the same basic paradigm, they examine the effects of level of participation on effectiveness at the tangram task. The important contrast is between two types of 'acknowledged overhearer', "omniscient bystanders" and "side participants" who both watch, and listen to, a pair performing the basic tangram task, again over six trials. Both types of overhearer are fully apprised of everything the director does, they know which figure is being referred to and they listen to all of the exchange between director and matcher, who are themselves aware of the presence of overhearers in both cases. In both conditions the overhearers are silent, making no direct contribution to the course of the task. The key difference between the two conditions is that the omniscient bystander observes events via a video and audio link whereas the side participant sits at the director's table, about 1m from the director's chair. In crude informational terms both types of overhearer are equivalent. However the side participant is also a ratified party to the conversation i.e., the director sees them as an acknowledged participant in the dialogue with the matcher. After the inital phase of six trials the bystander from the first phase becomes the matcher for a second phase of trials. Interestingly, the dyads composed of former omniscient bystanders and director are reliably less efficient at the task, on a number of measures, than pairs composed of former side participants and director. Directors are faster with a former side participant, 44 seconds per trial as opposed to 66 seconds per trial, and produce 33% fewer words. It seems that participation as a ratified overhearer is sufficient collaboration to improve performance, their silent participation in the first phase providing a degree of acceptance to the relevant presentations that the director treats as evidence of a degree of established mutual-belief.

Overall, the collaborative model draws several levels of participatory action, driven by cycles of presentation and acceptance, within its scope. The specific experiments discussed above are essentially neutral on the question of how idiolectical variation contributes to the observed coordination between speakers. Nothing in the data needs to be interpreted as collaboration on the meaning, in the sense of semantics, of the referring expressions used. However, the third level of participatory action, that of meaning and understanding, is understood as subject to the same mechanisms and, under the appropriate conditions, could be implicated in the same way. As a result, the process of collaboration to secure mutual-belief offers a possible mechanism for dealing with idiolectical differences and the threat they pose to mutual-intelligibility.

3.2.4 Input-Output Coordination

A second empirical model that offers a mechanism for dealing with idiolectical variation and is explicitly concerned with the issue of coordination of meaning, is input-output coordination (Garrod & Anderson, 1987; Garrod & Doherty, 1994). The development of this model has been driven by analysis of dialogues generated by the maze task (described further in section 4.2). In this task, two players are each seated at a VDU in separate rooms connected by a 2-way audio link. On their screens both players see the same maze configuration, consisting of boxes with links between them. Identical in all other respects, there are several additional features marked only on a particular player's maze. These are: a marker indicating their the player's own current position, a goal point which the players must move toward, and a set of switch points and gates each positioned differently on the respective display. The task is completed when both players reach, through a succession of alternating moves, their goal points. The collaborative nature of the task derives from the fact that when player A moves into a switch point marked on player B's maze all the open gates on player B's maze close and all the closed

gates open. Thus, if confronted by a closed gate player B must try to communicate to player A the location of a switch point, marked on B's maze but not A's, to which A can move thus opening the gate. Consequently, the resulting dialogues contain a number of exchanges of descriptions of positions as the players try to coordinate their understanding of both their current locations and various target locations to which they must move.

Analysis of the transcripts from these tasks reveals that descriptions tend to fall into one of four broad types: Figural which utilise some salient feature or aspect of the configuration, Path which pick out a route to be traversed between boxes, Line which order the maze into a set of rows or columns giving locations as n boxes along the row or column, and Matrix which appeal to a set of cartesian coordinates to identify a position (a more detailed discussion of each category is found in section 4.2). Garrod and Anderson provide detailed evidence that the different description types are not just arbitrary labels for various positions but rather depend on different mental models or conceptualisations of the maze used in the interpretation of descriptions (cf. section 2.1.2). For example, even amongst Line description types, choice of a particular expression describing the middle row seems to constrain the choice of description for top or bottom rows. Thus, where some ordinal numbering scheme has been introduced the bottom row is referred to as "row one", not "bottom" or "first". Similarly, the occurrence of prenominals such as "top" or "bottom" patterns with middle row descriptions of the form "third bottom row". The way choice of one description constrains the form of other descriptions indicates that a coherent overall interpretative scheme is being employed, one that varies between individuals. In contrast to Clark and Wilkes-Gibbs (1990), Wilkes-Gibbs and Clark (1992), this directly raises the issue of coordination of interpretation between members of a dyad.

Garrod and Anderson (1987) demonstrate that the distribution of different description types displays some reliable patterns across dyads. In particular, mem-

bers of the same dyad are much more likely to select the same description type than would be expected by chance, calculated on the basis of the overall distribution of description types in the corpus, indicating a tendency for pairs to coordinate their use of description scheme and the underlying conceptualisation of the maze it implies. Importantly, the observed coordination of description types within pairs is not achieved through a process of explicit negotiation as to which description type to use. Firstly, it is rare, occurring in only 27% of dialogues. Where it does occur it is almost always subsequent to the completion of several descriptions and then only where there have been substantial problems in coordinating on some scheme. Secondly, where pairs do try to arrive at a negotiated solution the resulting scheme only predicts the form of 59% of the subsequent description types (similar patterns are reported in Garrod & Doherty, 1994).

Garrod and Anderson emphasise that the level of coordination guaranteed by membership of a linguistic community, such as English speakers, is insufficient to explain the degree of coordination individuals achieve at the task. Commonly, they begin with a range of interpretations of words like "row", for example as horizontal, vertical and even diagonal lines but during the course of the task pairs tend to settle on a single interpretation. In order to account for this pattern of coordination without appeal to explicit negotiation Garrod and Anderson propose an interactional principle:

> "*output/input coordination*, ... may be simply stated as one of formulating your output (i.e., utterances) according to the same principles of interpretation (i.e., model and semantic rules) as those needed to interpret the most recent relevant input (i.e., utterance from the interlocutor)." (Garrod and Anderson, 1987, p.27)

Adherence to this principle ensures that pairs will tend to be locally consistent, as the data show, and achieve this without recourse to 'higher-order' beliefs about

how their addressee is interpreting descriptions, or explicit negotiation. Allowing for convergence on more specific interpretations of words such as "row" and "bottom" than are available prior to the task, discrepancies between two individuals' schemes are most likely to become apparent during the process of applying the interpretation arrived at for the previous input to the generation of a new output. Any discrepancy can provide the impetus for an individual to restate the original description according to their assessment of the currently accepted interpretation, entrenching it further as the common scheme.

Applied rigidly, input-output coordination does not allow for schemes to develop beyond a certain point: once coordination has been achieved modifications cannot be introduced without violating the principle. Garrod and Anderson suggest that one means by which individuals may overcome this is through a division of control. One speaker effectively takes control of the scheme used (the 'leader'), correcting their partner's descriptions ('follower'), and occasionally introducing new systems of, e.g., counting. This is supported by examples from the corpus which show pairs in which one speaker always conforms to input-output coordination while the other switches schemes and occasionally rephrases the descriptions offered by their partner.

A second experiment, reported by Garrod and Doherty (1994), has prompted further modification of this idea. In this study, degree of coordination is found to depend on membership of some 'virtual community'. Some details of the design of this experiment are given in section 4.2; however the basic contrast is between three groups who engage in nine maze task games. In the isolated pairs condition dyad composition is constant, always consisting of the same two individuals for all nine games. In the community condition players meet a different individual in each game but always drawn from the same pool of ten. Thus, on later trials, dyads develop a progressively larger common history of individuals they have both met already or, through indirect links, individuals one of them has already played who

have previously met individuals their partner has already played and so forth. The third, non-community, condition also involves dyads composed of different individuals on each trail, but in this case they are not drawn from a common pool. Although all three conditions show coordination of the kind predicted by the input-output coordination model there are also reliable differences between the groups. Although they initially show a lower degree of coordination than the isolated pairs, the community group rapidly converge on a Matrix scheme and by the third trial show almost perfect entrainment of description types. The non-community group, like the community group, initially show lower coordination than the isolated pairs but the degree of entrainment does not increase across trials, if anything, falling.

To accommodate these findings, Garrod and Doherty consider the various mechanisms that may operate to determine how conventions can emerge in the different groups. Considering the community group and isolated pairs first, the isolated pairs are modelled as conforming to the input-output coordination principle. Their choice of description type is influenced both by the maze configuration and the precedent set by their partner in their previous description. This mechanism provides a degree of local stability in choice of description types but cannot become a global convention, in the sense of Lewis (1969), as this depends on predicting the behaviour of several different individuals with respect to the task, i.e., it must become common knowledge amongst a group of players that they will usually choose a scheme of a particular type. In isolated pairs the local choice of scheme cannot be fixed in this way: any violation of the current scheme sets a new precedent. In the community group, however, individuals are exposed to a range of different individuals and therefore can begin to fix a more global convention. In this case a violation of the current scheme by one partner does not automatically undermine the generalisation that, across a range of partners, that scheme is the one usually chosen. In support of this Garrod and Doherty observe that in all

conditions, as players near their goals, the more likely they are to depart from the currently accepted scheme and produce a goal-related Path description, reflecting their sensitivity to the salience of the goal. Comparison of the frequency with which isolated pairs versus community group pairs are influenced by this cue reveals that isolated pairs are approximately twice as likely to choose a goal-related description. This supports the suggestion that isolated pairs are developing a different, local, form of coordination that is more readily disturbed by salience.

In addition to this basic contrast, they draw an analogy with exemplar-based models of concept learning to account for the stronger convergence observed in the late games for the community group. Roughly, the greater the range of exemplars an individual is exposed to in a concept learning task, the more stable the associated concept, in this case interpretive scheme, will be. Because of the greater number of people involved, individuals in the community group are exposed to a greater range of descriptions (2.8) than the isolated pairs (1.6) prior to convergence, suggesting that, on this account, individuals in the community group should derive more stable representations of the particular interpretation scheme.

Turning to the disparity in coordination between the community and non-community groups, in both conditions players are exposed to a wider range of exemplars and might therefore be expected to display similar convergence. To account for the difference between these groups Garrod and Doherty look at what happens where description types are in conflict. They suggest that conflicts are resolved through the interaction of two mechanisms. Firstly, the initial degree of coordination is achieved through input-output coordination but where this fails a dyad will adopt the scheme most commonly used by both players in the preceding games and this pattern is supported by the data from the community group. The difference arises because, in the community group, where a degree of 'common interaction history' emerges, adherence to input-output coordination amongst the different pairs predicts that, where a conflict arises, there is more likely to be an es-

tablished common scheme for the pair. By contrast, players in the non-community group do not have the same history of common individuals with whom they have interacted and are therefore less likely to achieve a high degree of coordination. Thus, input-output coordination can underwrite the stronger convergence on a common scheme in the community group than in the non-commuity group.

3.3 Discussion

Amongst the range of empirical models discussed in this chapter only the collaborative model and the input-output coordination model provide accounts which can address the issue of mutual-intelligibility in the face of idiolectical variation. In view of the widespread acknowledgement that idiolectical differences are the norm, this is surprising. The majority of work that has addressed questions of inter-individual coordination has concentrated on factors relating to social and pragmatic concerns which, though important, leave unaddressed the question of how semantic coordination is achieved. Conversation analysis does consider the need for an account of mutual-intelligibility but approaches it in a way difficult to reconcile with formal and experimental concerns. The concentration on the "procedural infrastructure of interaction" obscures the question of what individuals must know in order to engage in successful conversational transactions.

The collaborative model and the input-output coordination model both offer mechanisms in virtue of which individuals can overcome the interpretive asymmetries that may obtain between them. Furthermore, both are pitched at levels which directly address the semantic concerns such differences entail. However, there are also differences between the two accounts. The collaborative model takes coordination to be achieved through an explicit cycle of presentations and acceptances by the parties to a particular dialogue. The input-output coordination principle, by contrast, offers a group-based account that does not appeal to explicitly

negotiated mutual beliefs. This contrast is investigated further in Chapter 4.

Chapter 4

Experimental Studies of Coordination

Chapter 3 concentrated on idiolectical variation as a practical problem for the maintenance of mutual-intelligibility and surveyed exisiting empirical models for possible solutions. It was claimed that only two models, input-output coordination and the collaborative model, address this issue, providing mechanisms which facilitate semantic coordination between interlocutors. This chapter examines the empirical problem in more detail, concentrating on an experimental investigation of semantic coordination. The general rationale behind these studies is that the action of mechanisms which promote coordination can be most effectively revealed under circumstances in which a degree of interference between them is created. That is, if the processes which enhance coordination can be brought into conflict, then the factors which govern their operation should become more empirically tractable.

In pursuit of this, two conditions need to be fulfilled. Firstly, it is necessary to identify conditions under which the problems caused by idiolectical variation should be particularly apparent, promoting greater efforts at coordination. The

conversational domains most likely to provoke this will be those least familiar to the participants, i.e., domains for which there are few semantic precedents, beyond those constituted by membership of a given linguistic community, for dealing with the topic in question (see Lewis, 1969). As a rule of thumb, we might expect the threat idiolectical variation poses to mutual-intelligibility to be inversely proportional to the familiarity of the domain of discourse.

Guided by this heuristic, these experiments display a preference for abstract materials and relatively contrived tasks, a preference which raises specific questions of generalisability in addition to those normally provoked by experimental studies. The strategies that individuals employ in dealing with experimental tasks involving tangram figures or maze-like grids could turn out to have only an indirect bearing on everyday conversational transactions. There is no straightforward way to address this worry; however, a minimum hope is that these studies at least provide a plausible starting point for investigation. Conversely, the underdeveloped state of theory in this area suggests that it would be premature to rule out experimental studies of the kind pursued here since there is, *ipso facto*, no uncontentious way of motivating judgements of ecological validity.

The second condition that must be met is the creation of circumstances under which interference in the processes of coordination is predicted. If different groups of interlocutors achieve a degree of coordination for some domain then there is no *a priori* reason to expect that each group will do this in the same way. This follows from the suggestion that semantic conventions constitute *arbitrary* solutions to recurrent problems of coordination (Lewis, 1969). If this holds then interference should arise where individuals from different groups are faced with the same need to coordinate, for the same domain, but the previously achieved basis for that coordination has been removed. Any disturbance in performance that arises in these circumstances indexes the importance of semantic coordination and the manner in which it is disturbed provides clues to the mechanisms which operate

to achieve it.

Beyond the general rationale, the design of these experiments is also guided by the more specific aim of resolving the tension, raised in chapter 3, between the input-output coordination model (section 3.2.4) and the collaborative model (section 3.2.3). A central claim of the collaborative model is that utterances are subject to a process of presentation and acceptance before they are accumulated to the common ground between participants in a dialogue. This cycle is seen as crucial in underwriting the mutual beliefs held to be necessary for communication (Clark & Schaefer, 1989; Clark & Marshall, 1981; Wilkes-Gibbs & Clark, 1992) and is singled out as critical to establishing the use of particular referring expressions. As Garrod and Doherty (1994) note, *prima facie*, the collaborative model cannot account for the finding that convergence on particular reference schemes is observed within experimental 'communities' because this occurs before there has been a chance for each individual in the community to establish, separately, the mutual belief with each other individual that scheme X represents the conventional way to refer to particular positions in a maze.

Although, on the collaborative model, a scheme of reference could only become conventionalised through the active establishment of mutual belief between the appropriate parties, other explanations of Garrod and Doherty's results are possible. Wilkes-Gibbs and Clark (1992) note that, in addition to collaborative processes of adjustment, the referring expressions used in the tangram task also evolve as a function of each individual's experience with the task. Thus an experienced participant faced with a naïve partner will offer more readily identified descriptions on the first trial with the new partner than was achieved on the first trial with their previous partner. Wilkes-Gibbs and Clark (1992) attribute this difference to changes in expertise with the task, suggesting that more experienced participants are, in some sense, more skilled at generating effective referring expressions. A similar explanation could be adduced to account for Garrod and Doherty's re-

sults: the observed convergence on a particular scheme results from increasing expertise with the task, not from community based mechanisms of coordination. If individuals in the community are each becoming more expert at communicating information about spatial locations, the emergent scheme might represent an optimal solution to the task, arrived at independently by each participant.

It is worth noting that this explanation faces some difficulties in dealing with specific aspects of Garrod and Doherty's findings. In contrast to the community group, isolated pairs who repeatedly perform the maze task do not all converge on the same scheme. Accounting for this contrast requires appeal to the development of expertise in performing the task with a succession of different partners rather than development of expertise with the materials *per se*.[1] An expertise explanation is also complicated by the comparison of the community group and a control, 'non-community' group constructed so that individuals do accumulate experience with different dialogue partners but share only one previous partner in common with them. As discussed in section 3.2.4, the non-community group doesn't converge on a single scheme in the same way as the community group. However, there are reasons why this does not necessarily undermine the expertise explanation. Firstly, the degree of experience of different partners is not equivalent across the community and control groups: in fact only 'key' individuals participate in all five trials, their partners experiencing only one game prior to meeting them. By contrast, the structure of the community group dictates that each individual has, on average, been involved in the same number of games. Secondly, although the control group doesn't converge on the same scheme as the community group neither is it as heterogenous as the isolated pairs. The partial confound of degree of experience combined with the disparity between the isolated pairs and the control

[1]Some support for this explanation derives from the significantly higher degree of coordination that isolated pairs show in early trials. This disparity could be interpreted as indicating that individuals treat repeated trials with the same partner as a different task from repeated trials with different partners.

group leaves a line of defence open for an expertise based explanation.

These considerations combine to suggest a stronger test of a collaborative model explanation of Garrod and Doherty's data. In outline, the strategy adopted in the studies reported below is to encourage, over the course of several trials at a collaborative task, the emergence of a number of different 'semantic communities' or sub-groups. This is followed by a final trial in which the task is performed again but in this case half the pairs consist of individuals drawn from the same sub-groups, a homogenous condition, and half the pairs consist of individuals drawn from different sub-groups, a mixed condition. Assuming that variables relating to experience are appropriately controlled, the collaborative model predicts there should be no significant difference between the performance of the homogenous and mixed conditions.

All three experiments reported below are designed with the principal aim of providing a robust test of the prediction that there should be no difference between mixed and homogenous groups. However, they are also intended to satisfy two other interests. Firstly, the paucity of experimental studies dealing explicitly with issues of semantic coordination has the consequence that, with the exception of studies by Garrod and coworkers (e.g., Garrod & Anderson, 1987; Garrod & Doherty, 1994), there is little established methodology to draw on in addressing these questions. In addition to the uncertainty surrounding the selection of suitable materials, little is known about what measures are appropriate for the detection of possible effects and there are few precedents for the selection of appropriate tasks. Consequently, an avocational element of this investigation is the attempt to identify and develop suitable methodologies and tasks. Secondly, the experiments are intended to provide a corpus of dialogues suitable for informing the semantic modelling of chapter 5.

4.1 Experiment 1

Echoing work on the collaborative model, the first experiment was based on a tangram sorting task. However, the exact design developed by Krauss and Glucksberg (Clark & Wilkes-Gibbs, 1986) was not directly adopted on the grounds that the asymmetry between roles of director and matcher in producing descriptions might dilute the degree of negotiation between interlocutors.[2] Instead, a revised task was used involving two phases. In the first phase pairs of subjects negotiate various criteria by which they can exhaustively divide a group of tangram figures into two equal sets. The negotiation phase is then followed by a test, carried out independently by each member of a pair, of the accuracy and speed of their recall for the agreed classification. The rationale behind this is that because tangram figures are relatively abstract shapes the process of partitioning them into two categories should invoke a high degree of negotiation in order to reach agreement on both the expressions appropriate for referring to individual figures and the possible categories into which they might fall. Assuming this is the case, subjects performing the task with a succession of partners drawn from a particular sub-group are predicted to develop a degree of intra-group semantic coordination of the kind observed by Garrod and Doherty (*ibid*). Given a degree of emergent coordination, the performance of individuals who perform a final trial in dyads composed of individuals from different sub-groups can be compared with those who remain in the same sub-group.

[2]The possibility of using the map task (Anderson, Brown, Shillcock, & Yule, 1984) was rejected on the same grounds as a similar asymmetry obtains between the route-giver and route-follower.

4.1.1 Method

The experiment consisted of three trials, each divided into two phases. In the first phase subjects negotiated, in pairs, the classification of a set of twelve tangram figures. Each pair's negotiations were recorded and subsequently transcribed to allow for detailed content analysis. The second phase consisted of a speeded decision task, performed alone, in which the figures from the first phase were presented, one at a time in random order, to each subject and they were asked to indicate, via a keypress, to which of their agreed categories the figure belonged. The measure of response time was selected on the grounds that it should provide a sensitive measure of the confidence subjects had in the classification of each figure.

Materials

The materials consisted of 96 tangram figures selected from Elffers (1973) representing a range of degrees of abstraction; from those that clearly resemble a figure or animal to more abstract geometrical forms. These were divided up into eight sets of twelve figures, taking care to ensure that no set contained more than two figures readily recognisable as being of a particular type such as animals (see figure 4.1).

The materials were presented using the psychology testing software Superlab, version 1.4, running on Apple-Macintosh SE30's under system seven. Script files were written to control the timing and order of presentation of materials. In the first phase instructions were displayed until the space bar was pressed (coordinated by instruction from the experimenter) followed by presentation of the complete set of 12 figures for two minutes. The second phase also began with instructions presented until the space bar was pressed (again, on instruction) followed, consecutively, by each of the twelve figures. In each case, a fixation point was displayed for 350ms followed by presentation of the target figure, terminated by an appropriate keypress. The order of presentation was automatically ran-

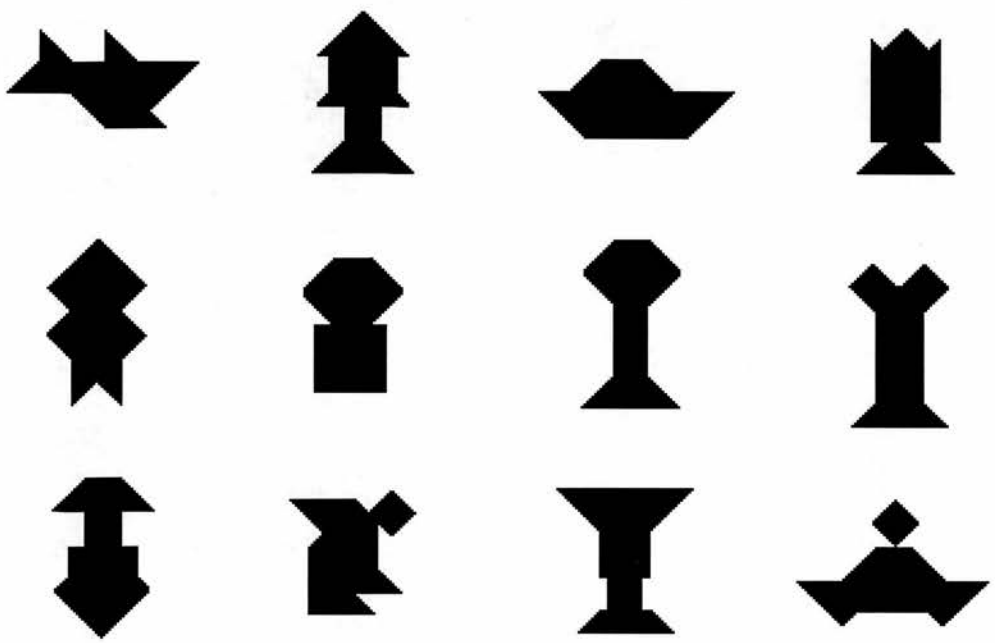

Figure 4.1: Example Set of Tangram Figures

domised and the response times logged to a results file. Response times were recorded via the Apple Desktop Bus keyboard using the Toolbox Time Manager with an accuracy of ±1ms.

Design

Subjects were randomly assigned into sub-groups of four, each constituting one 'community'. In the first two trials subjects performed the task with different individuals drawn from the same sub-group. On the third trial, two members of a sub-group formed a new pair and the remaining two were combined with

Trial:	1	2	3	Condition
Pair:	1+2	1+3	1+4	Homogenous
	3+4	2+4	3+6	Mixed
	5+6	5+7	5+8	Homogenous
	7+8	6+8	2+7	Mixed

Table 4.1: Dyad Composition Across Trials for 8 Subjects

subjects from a different sub-group. The resulting combinations generated for two sub-groups of four subjects are illustrated in table 4.1.

In order to control for effects of materials the eight groups of twelve tangram figures were counterbalanced across trials and conditions in a latin square assignment. Each subject, on each trial, encountered a new set of materials and a new dialogue partner. This resulted in a factorial design with a between-subjects independent variable of dyad composition (mixed versus homogenous) and dependent variables of response time and degree of agreement between dyad partners' classification of figures in phase 2.

Subjects

Thirty two subjects took part, recruited from staff, graduate and undergraduate students from various disciplines at the University of Edinburgh. They consisted of 8 females and 24 males ranging in age from 18 to 39 with an average age of 24 years. They were paid £3 each for taking part in the experiment.

Procedure

The experiment was carried out in four sessions with eight subjects per session making up two sub-groups of four. It was made clear to all subjects that their dialogues would be recorded and transcribed (but anonymously coded) and that they were free to withdraw from the study if this presented them with any problem. On each trial, subjects were divided into four dyads, seated opposite each other at

a desk with an SE30 in front of each subject, the screen of each computer visible to only one member of the pair. On all three trials they were instructed that the task would proceed in two phases, the first phase to be carried out in collaboration with their partner, the second to be carried out individually. In the first phase they would both be presented with a set of twelve tangram figures, arranged differently on each screen so that they couldn't be identified by position. The figures would be displayed for two minutes during which they are asked to decide, jointly, on some classification that sorted the figures into two groups of six. This was to be achieved by discussion rather than pointing or using gestures. They were informed that this would be followed by a test of the speed and accuracy with which they could categorise each figure according to their agreed classification. After the first phase they were asked to write down the names of the categories they had agreed, labelling them A and B. At the start of the second phase subjects were instructed that they would be presented, in random order, with each of the twelve figures they had previously classified and asked to indicate, as quickly and as accurately as possible, which of the two categories the figure had been assigned to (the d and K keys on the keyboard were relabelled A and B for this purpose). Care was taken to ensure that no indication was given to the subjects, either in the materials or the instructions, that they were divided into two sub-groups or communities and during debriefing no subject reported detecting this element of the design.

4.1.2 Results

Two tests were performed to provide general information about the improvement in performance as experience with the task increased. Firstly, the average response time for each subject on each trial was calculated and entered in an analysis of variance with a single within-subjects factor of experience with 3 levels corresponding to trials 1,2 and 3. There was a significant main effect of experience, $F_{(2,62)}=8.34$, $p=0.001$, with mean response time falling from 1086ms in the first trial to 801ms

and 720ms in the second and third trials respectively.

In order to index the changes in effectiveness of the agreed classification as experience with the task increased, each pair was scored for the proportion of the target items, out of twelve, that they assigned to the same classification. This was entered into an analysis of variance with a single between-subjects factor of experience with three levels corresponding to each trial. There was no main effect of experience: $F_{(2,45)}$=1.359, p=0.267. However, the means did suggest an increase in the expected direction as experience increased: trial 1: 0.81, 2: 0.88 and 3: 0.91 and a linear trend analysis confirmed this, $t_{(45)}$=1.747, p (one-tailed) =0.050.

The comparison of mixed and homogenous groups was made in two ways. Firstly, the average response times for each subject in trial three were entered into an analysis of variance with a single, between-subjects factor of group composition (homogenous vs. mixed). The mean response time for the homogenous pairs was 717ms compared with 723ms for the mixed pairs and these were not significantly different: $F_{(1,30)}$=0.005, p=0.944.

A second, more sensitive, comparison was made by calculating the difference between each subject's average response time in trials 2 and 3, providing an index of how each individual's response time was altered in trial 3. This was entered into an analysis of variance with a single within-subjects factor of group composition (homogenous vs. mixed). The anova revealed a significant main effect of group composition, $F_{(1,30)}$=4.23, p=0.048, with a mean decrease in response time for individuals in homogenous pairs of 3ms compared with a mean decrease of 159ms for individuals in mixed pairs.

Transcripts

Forty-eight two minute dialogues were collected in phase one and subsequently transcribed. Analysis of the transcripts revealed a very consistent strategy across trials and conditions. In contrast to the methods adopted by Clark and Wilkes-

Gibbs (1986) this task did not successfully elicit negotiation of referring expressions for individual figures. Instead, by far the most common strategy (94% of dyads) involved identifying some general property that six figures held in common, for example, "pointy tops" or "triangular base", and using this to discriminate between the two sets, labelling one as possessing the property and the remainder as "others". This was achieved by direct discussion of candidate discriminating properties and consequently required very little discussion of individual figures except where the number of figures possessing a particular discriminating property was not exactly six. In these cases, rather than revise their choice of property, the common pattern was to identify a single figure to either include or exclude from the set possessing the key property. For example, "triangle top and straight down not including the triangle out the way".

A clear majority of the properties dyads used to divide up the tangram figures were based on the two-dimensional geometric form of the figures. Furthermore, a number of properties recurred across trials allowing a classification into types. The most commonly cited property was possession of a pointed top or vertex, realised variously as, "triangular top", "single pointing arrow top", "pyramid top", "thin ones with roofs" and "pointy tops". This description type formed the basis of 30% of the categories. The second most commonly cited discriminating property involved appeal to the relative height of the figures, with descriptions such as; "tower-like", "long tall", "long vertical length on both sides","tall thin" and most frequently just "tall". This property was used to partition the figures in 19% of cases. The third most common type was possession of a triangular base which occured as "flat bottom with incline sides" and "triangular base". This property was used to determine 15% of the categories. Two other properties were utilised in more than one description: these were "flat base", 6%, and "flat top", 6%.

The remaining categories were all determined on the basis of properties that were not used more than once. These idiosyncratic methods of discriminating

Criterion	Trial 1	Trial 2	Trial 3
Unique	6	2	3
Relative Height	5	2	2
Pointed Tops	2	7	6
Triangular Base	1	4	2
Flat Base	2	0	1
Flat Top	0	1	2

Table 4.2: Frequency of Criteria Across Trials

amongst the figures accounted for the remaining 23% of the categories. Although necessarily heterogeneous, this group of properties did display some common features. They tended to be closely tied to the particular set of figures in question and to fall into one of two classes. The majority were more specific or elaborated versions of the general geometric properties described above: "four concavities point at top", "small square blocks attached", "right-angled isoceles triangle at least one", "no more than two points", "pointy top and flat or downward bottom" and "forty five degree wedge missing". The remainder appealed to a richer interpretation of half the figures as representations of objects, "familiar", "living things", "objects", contrasted with a set of abstract, uninterpreted figures.

Table 4.2 illustrates the distribution of the various criteria for partitioning figures across trials. Inspection of the raw frequencies suggests a shift in the pattern of criteria adopted across trials. Unfortunately, there are insufficient data to determine the reliability of this pattern.[3]

Collapsing all the criteria that repeated across trials into one category of generalisable or 'abstract' criteria, it was possible to make a focussed comparison of unique versus 'abstract' on trials 1 and 3 using the Fisher exact probability test. This was not significant, p=0.159.

[3]Calculation of omnibus Chi-square is inappropriate since no cell has expected frequencies which rise above 5. Fisher's exact probability test could not be applied as it is computed only for 2-by-2 comparisons (Rosenthal & Rosnow, 1991).

Criterion	Homogenous	Mixed
Unique	1	2
Relative Height	0	2
Pointed Tops	4	2
Triangular Base	1	1
Flat Base	1	0
Flat Top	1	1

Table 4.3: Frequency of Criteria in Homogenous and Mixed Dyads

Table 4.3 reports the relative distributions of each criterion type in the homogenous and mixed dyads in trial three. There is a weak indication that the homogenous dyads were more uniform in their choice of criteria than the mixed dyads. However, as above, there are too few data points for meaningful comparisons to be made.

4.1.3 Discussion

In respect of methods and choice of dependent variables, this experiment was a qualified success. The results indicate that as experience with the task increased, subjects' performance also improved, becoming both faster and more accurate in the assignment of figures to their agreed categories. The data suggest that the task was comprehensible to subjects and that the chosen measure of response time was, at least to some degree, an effective index of subjects' expertise and confidence in their classification of the figures. Against this background, the comparisons between the homogenous and mixed dyads in trial 3 appear warranted and might be sensibly brought to bear on the experimental hypothesis. The between-subjects comparison of response times for the two groups shows no difference in task performance as the collaborative model would predict. However, the more sensitive, within-subjects comparison of changes in each individuals response times shows a suprising decrement in perfomance for individuals in the homogenous dyads

compared to the mixed dyads in the third trial. On the face of it, this result is predicted neither by the collaborative model nor the rationale for the experiment. Nonetheless, it does appear to undermine the attempt to account for emergent semantic coordination as a product of individual task experience.

Although encouraging, there are several reasons why this conclusion, and the evidence that supports it, are not decisive. The clearest weakness is that although the results suggest a difference between the mixed and homogenous groups, they provide no convergent evidence that this is actually due to the emergence of local semantic 'dialects' within the different sub-groups. As noted, the task did not elicit negotiation of referring expressions for each figure which could then be compared across trials and conditions. Instead, subjects concentrated on isolating some discriminating property that could be used to isolate a subset of figures. This led to few or, in most cases, no exchanges of descriptions concerning individual figures. Furthermore, the clear majority of these properties were based on simple, two-dimensional, geometric considerations, only the unique, idiosyncratic, criteria showed any of the richer interpretations evident in Clark and Wilkes-Gibbs (1986). As a result, the available data were not amenable to the type of analysis used in previous work, for example, contraction/expansion of referring expressions and distribution of description types, and there were too few data points, in the analysis that was possible, to permit meaningful statistical comparisons.

The concerns raised by the lack of evidence for emergent conventions are compounded by the fact that the design of this experiment admitted only two trials prior to the experimental manipulation. Practical constraints restricted the study to sub-groups of four, which can only support three trials before individuals must meet twice, allowing only a short period for convergence to occur. While Garrod and Doherty (1994) report evidence of convergence within three trials of the maze task, this is the minimum reasonable number of trials, and it is particularly questionable whether it is valid to generalise this expectation to a different task.

This consideration weakens confidence that the effects observed in this experiment are due to community-based emergence of semantic conventions. Without some independent measure of convergence it is difficult to draw any firm conclusions. Furthermore, the significant difference that was found to obtain between response times in trials 2 and 3 is subject to a possible confound. Although the materials were counterbalanced across all three trials the *shifts* experienced by individuals between trials 2 and 3 were not (i.e., A $\rightarrow$ B may not be equivalent to B $\rightarrow$ A). This opens up the possibility that there may have been a systematic difference in the difficulty of the shifts experienced by individuals in the mixed and homogenous groups.

Although susceptible to criticism, the experiment does indicate that the collaborative model might not provide an adequate account of the emergence of conventionalised reference schemes within sub-groups. The absence of a convincing demonstration that local, intra-group, convergence occurred in this study means that no strong conclusion can be drawn. Nonetheless, the distribution of criteria types across the mixed and homogenous groups is suggestive and, combined with the within-subjects comparison in response times, provides sufficient motivation for pursuing the investigation further.

4.2 Experiment 2

Experiment 2 aimed to address the design faults in experiment 1 while preserving the same basic rationale. As before, the aim was to produce conditions under which sub-groups would converge on particular local conventions and then determine whether, and in what ways, crossing-over between subgroups affects performance.

4.2.1 Methods

The shortcomings identified with the tangram task used in the first experiment prompted adoption of the maze task (Garrod & Anderson, 1987; Garrod & Doherty, 1994), described in section 3.2.4, as the basic experimental task. This task has the advantage that the number of descriptions each subject produces can be manipulated and these expressions can be classified according to an established set of categories, overcoming some of the problems that arose with the more open-ended tangram task. As well as creating more chances for the exchange and modification of descriptions, the maze task offers established metrics for comparing the incidence of, and convergence on, description types across individuals, dyads and groups.

Materials

The original, electronic, version of the maze task takes approximately 20 minutes to run and requires two Apple Macintosh computers (SE-30 or better) per dyad. The aim of using large sub-groups in this study generated impractical demands on hardware and in order to overcome this problem a paper version of the maze task was employed. Screen dumps were made of thirty of the displays used in the original task and edited, using the graphics application Superpaint, to produce 168 basic maze configurations based on a 6-by-4 grid. These were divided into twelve sets of 14 and each set of 14 was made up into a pair of booklets, one with target locations, indicated by a circle, marked only on the odd numbered pages and one with locations marked only on the even numbered pages. Thus, for each pair of booklets, any given page number had identical maze configurations; however, on alternate pages only one booklet had a location marked, the other identical in all respects apart from a circle indicating the target location. An example pair of pages is illustrated in figure 4.2.

Using these booklets a modified version of the maze task was adopted in which

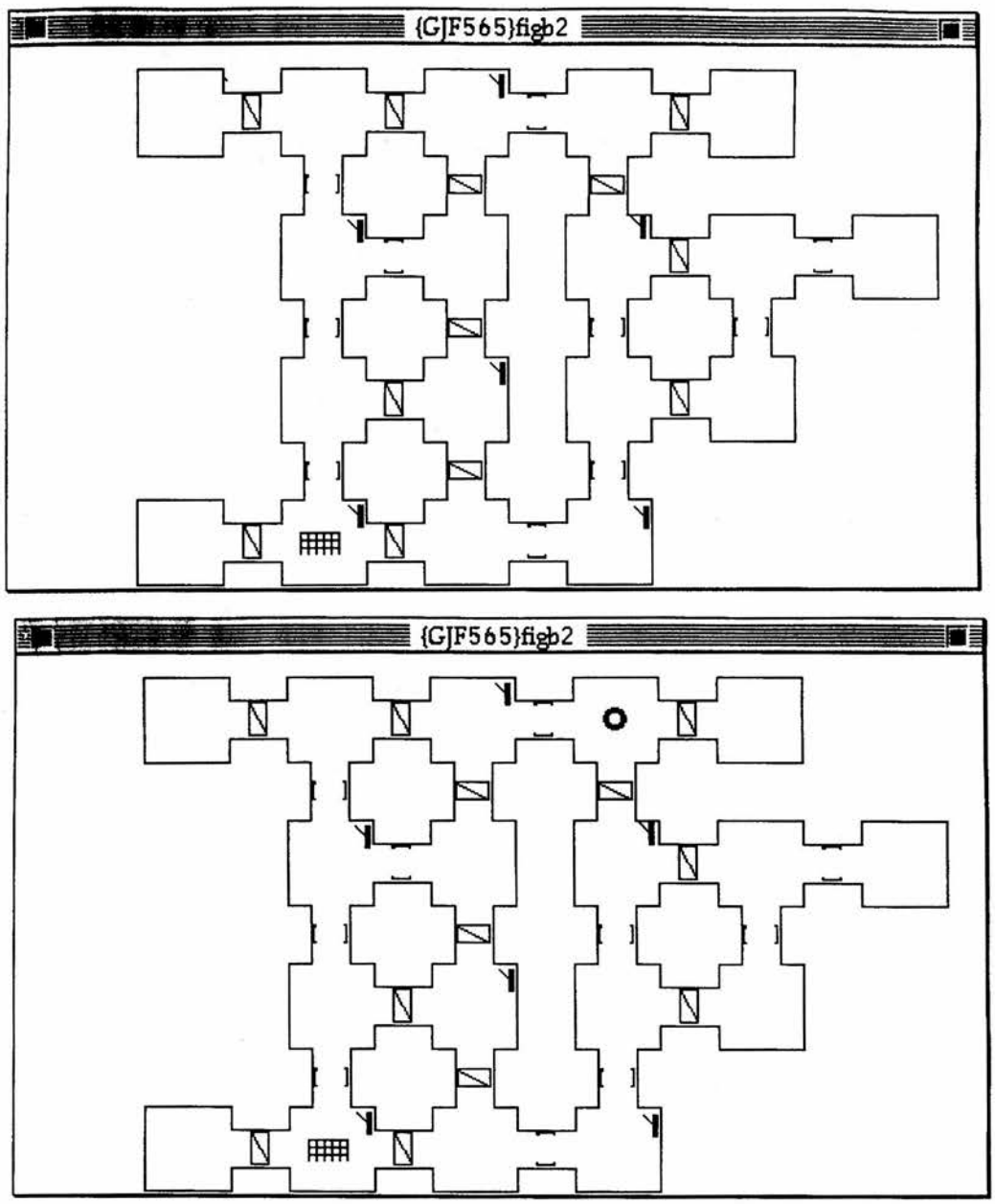

Figure 4.2: Example Pair of Maze Configurations

members of each dyad, alternately, describe the location marked on their maze and their dialogue partner indicates, using a pen, where they think it is. This preserved the requirement for the production and comprehension of spatial descriptions while obviating some of the practical difficulties associated with the original.

Design

The design of this experiment employed the same basic strategy as experiment 1 but was modified, employing larger sub-groups of subjects and a larger number of trials before the experimental manipulation, in order to promote a high degree of intra-group coordination. It was also intended to provide tighter control of two aspects of dyad composition not addressed in earlier studies. The practical problems with running the electronic version of the maze task had the consequence that in the original studies (Garrod & Anderson, 1987; Garrod & Doherty, 1994) subjects could not all carry out the task within a short period and were required to return on several different occasions. This often resulted in trials involving dyads composed of individuals with differing degrees of experience with the task and who had been subject to different intervals between trials. Although these differences were unlikely to have introduced systematic bias in the earlier work, they had the potential to act as nuisance variables in the current study. This was overcome by ensuring that both number of trials experienced and interval between trials were equivalent for all subjects at each stage of the experiment.

As for experiment one, materials were assigned in a latin square design in order to counterbalance possible biasing effects. This is particularly pertinent for the maze task as Anderson and Garrod (1987) report that certain types of maze configuration tend to elicit disporportionate numbers of descriptions of a given type. Some configurations are particularly likely to invoke "figural" descriptions while others appear to suggest "line" or "matrix" description types.

Satisfying these constraints resulted in a basic design consisting of multiples

of twenty-four subjects, divided into three sub-groups of eight, who participate in six trials each consisting of a two minute task-oriented dialogue. For the first five trials each dyad is composed of subjects drawn from within a single sub-group of eight. In the sixth trial half of the dyads are composed of subjects drawn, as before, from within one subgroup of eight while the remaining six dyads are composed of subjects drawn from different sub-groups. Labelling the three sub-groups of eight, A, B and C, the mixed dyads consisted of two pairs for each of the combinations; AB, BC and AC, counterbalancing the combinations. On the sixth trial half the subjects perform the task with individuals drawn from the same sub-group and half perform the task with individuals drawn from a different subgroup. The combinations generated for a set of twenty-four subjects are illustrated in table 4.4. As for experiment one, each subject, on each trial, is exposed to a different set of materials and a different dialogue partner.

Trial:	1	2	3	4	5	6	Condition
Pair:	1+2	1+3	1+4	1+5	1+6	1+7	Homogenous
	3+4	2+4	2+7	2+6	2+5	2+8	Homogenous
	5+6	5+7	3+6	3+7	3+8	3+13	Mixed
	7+8	6+8	5+8	4+8	4+7	4+22	Mixed
	9+10	9+11	9+12	9+13	9+14	14+15	Homogenous
	11+12	10+12	10+15	10+14	10+13	10+11	Homogenous
	13+14	13+15	11+14	11+15	11+16	9+21	Mixed
	15+16	14+16	13+16	12+16	12+15	6+12	Mixed
	17+18	17+19	17+20	17+21	17+22	17+23	Homogenous
	19+20	18+20	18+23	18+22	18+21	18+24	Homogenous
	21+22	21+23	19+22	19+23	19+24	5+19	Mixed
	23+24	22+24	21+24	20+24	20+23	16+20	Mixed

Table 4.4: Dyad Composition Across Trials for 24 Subjects

The resulting design was a simple factorial with dyad composition (mixed vs. homogenous) as a between-subjects independent variable and trial number as a within-subjects independent variable. The switch to an off-line version of the

maze task required the abandonment of a response time as a dependent variable. Instead, to provide basic measures of effectiveness/efficiency at the task, the scores each pair achieved for the total number of items attempted and for the proportion of items that involved erroneous identification of the target location were chosen as dependent variables.[4]

Subjects

The experiment was run in two parts with a total of 48 subjects participating. They were recruited from students studying A-level psychology at two colleges of further education in Edinburgh. They consisted of 9 males and 39 females ranging in age from 16 to 54 years with an average age of 33.

Procedure

The same procedure was followed in both parts of the experiment. Subjects were allocated randomly to sub-groups and each was assigned a number that was used throughout the experiment when pairing individuals into dyads. On each trial dyads were seated opposite each other at a desk with a partition between them in order to obscure their view of each other's booklets whilst permitting eye contact. At the start of each session it was explained that subjects would be asked to work in pairs, each member of the pair having a booklet of mazes on which one of them would have a target location, marked by a circle, and their partner would have the same maze but without the circle. The task being to communicate the location of the circle, without pointing or showing, to their partner who should then mark it on their copy. They were informed that they would perform this task six times,

[4]The decision to score by pair rather than by individual was dictated by the fact that, if scored by individual, these measures would violate the assumption of independence for analysis of variance. If one subject managed n items within the time allowed their partner could not, in virtue of the task structure, manage more than $n \pm 1$. This had the unhappy side-effect of requiring 48 subjects in order to generate twelve data points in each condition.

each time with a different partner. It was explained that on each occasion they would have two minutes to perform the task, start and finish to be signalled by the experimenter, and they were asked to complete as many pages of the booklet as possible within this period while preserving, as far as possible, accuracy. It was made clear that all the dialogues would be recorded, although anonymously coded, and that they were free to withdraw if this presented them with any problem. As before, no indication was given, either in the instructions or in the materials, that they were divided into subgroups and no one reported detecting this aspect of the design.

4.2.2 Results

Only one dyad managed to complete all 14 items within the time allotted, achieving this on trial 6. Overall, the average number of items completed within two minutes was low but increased across trials: 2.4, 5.0, 5.5, 6.8, 7.1, 8.1. This trend is illustrated in figure 4.3 with trials 1 & 2, 3 & 4 and 5 & 6 averaged to provide three levels of task experience: low, medium and high. The reliability of this pattern of increase was confirmed by an analysis of variance perfomed on the number of items completed by each pair with trial number as a between-subjects factor: omnibus $F_{(4,138)}$=17.71, p=0.000, linear trend; $t_{(138)}$=9.03, p (one-tailed)=0.000.

Each pair was also scored for the proportion of items that resulted in incorrect identifications of the target location. This suggested a reverse pattern, illustrated in figure 4.4 (with trials again collapsed to give three levels of experience, low, medium and high), with the average proportion of errors tending to fall across the first five trials: 0.32, 0.24, 0.29, 0.17, 0.20. An analysis of variance with trial number as a between-subjects factor confirmed the pattern of decrease: omnibus $F_{(4,138)}$=1.057, p=0.386, linear trend; $t_{(138)}$=1.930, p (one-tailed)=0.027.

The effects of the experimental manipulation of group composition on the dependent measures of items attempted and proportion of errors was tested using

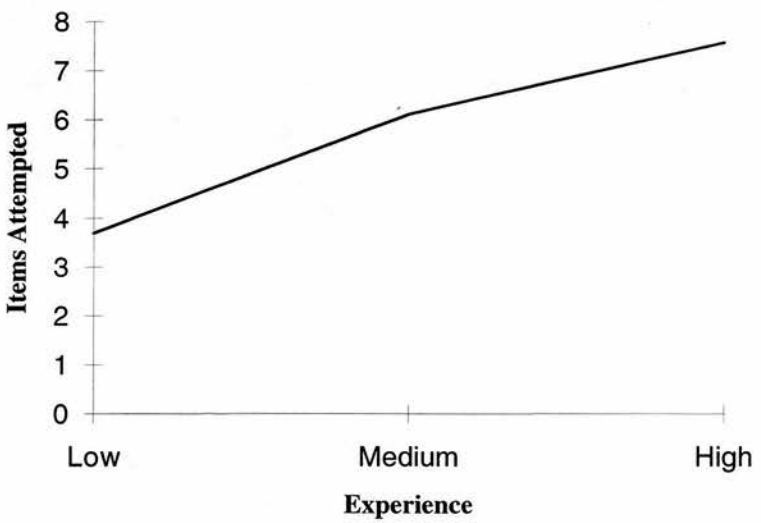

Figure 4.3: Number of Items Attempted with Experience

two analyses of variance with group composition (homogenous vs. mixed) as a single, between-subjects, factor. No reliable difference was found either for number of items attempted; $F_{(1,22)}$=0.04, p= 0.843, or for the proportion of errors; $F_{(1,22)}$=0.006, p=0.936.

Transcriptions

A total of 144 two minute dialogues were transcribed and coded for the occurrence of the description types, Figural, Path, Line and Matrix. Repetitions of all or part of an original description were not counted as completed descriptions, nor were procedural clarifications or clarifications, including reformulations, of part of a description. This classification was guided by the criteria offered in Garrod and Anderson (1987), drawing on the same conceptual and lexical discriminations. The criteria, with examples drawn from the current corpus, are illustrated below:

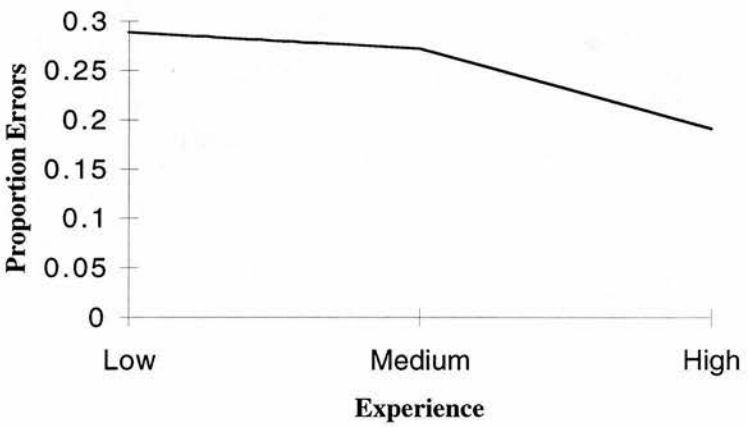

Figure 4.4: Proportion of Errors with Experience

Figural: draws on some element of the configuration or location of particular features to identify the target. For example:

"on the bottom row there's one missing it's diagonally to the: diagonally up to the left"

Path: identifies a route to be traversed through the maze to the target location. Sensitive to the layout of boxes and connections and frequently involves an interdependence in the enumeration of the vectors so that a location whose cartesian coordinates are 2,2 might be described as "two up, one along". Often contain adverbs such as "across " or "along".

"right from the right go to your left one and it's umm down one"

Line: classifies the maze into a set of elements corresponding to rows, columns or diagonals. The target line is described first, followed by the target box as a position along it. Frequently an ordering is imposed on the set of lines giving rise to descriptions that refer to lines as e.g., "second" or "last".

"right second column from the left and it's the second one down"

Matrix: effectively imposes a cartesian coordinate system on the maze with locations identified via the specification of two vectors either as rows and columns or in terms of two numbers, one for each axis.

"two, four"

Each description was coded, as appropriate, for its chosen or implied origin (e.g., top left, bottom right etc., where it lay at one of the corners of the basic 6-by-4 grid, otherwise its coordinates from bottom left where a corner was not used or, usually in the case of figural descriptions, whether it enlisted either a feature marked on the maze, a group of boxes that formed some shape or a pattern of spaces formed by missing boxes), the use of cardinal or ordinal numbers in enumerating boxes and the order in which the axes were introduced (i.e., X-axis first or Y-axis first). This information provided convergent evidence for the classification of ambiguous descriptions where it was not immediately clear from the description and its context what category applied. Overall, the transcripts generated a corpus of 975 descriptions, each classified into one of the description types.

During the process of coding the descriptions a difficulty with the application of Garrod and Anderson's criteria became apparent. The switch to a paper version of the maze task necessarily altered the exact nature of the task. In particular, subjects appeared to be less constrained by the exact configuration of the maze. In the electronic version, the presence or absence of passages between boxes represents an important restriction on the way the maze is conceptualised since subjects must move their figure between boxes and they consequently pay relatively close attention to the layout of connections between them. The paper version of the task does not require any movement between boxes, thus reducing the importance subjects attach to the pattern of connections. In terms of coding, the net effect of this

difference was to undermine the confidence with which some Path-type descriptions could be classified. Amongst the descriptions that could be unambiguously classified as Path, for example, because of an interdependence in the enumeration of the axes, subjects frequently offered descriptions that followed routes between unconnected boxes. This reduced confidence in categorisation of other descriptions that did not unequivocally meet the additional criteria for classification as Path-type.

In contrast to previous studies with the maze task, few matrix description types were identified in the corpus. Only 1.2% (11 out 975) fell into this category in the current study, compared with 23.4% reported in Garrod and Anderson (1987) and, averaging across conditions, approximately 40% in Garrod and Doherty (1994). Rather than discard these descriptions from the analysis, they were combined with the Line-type category to form a general category of 'Abstract' description types in the sense that, relative to Path and Figural descriptions, their interpretation relies less on the configuration of any particular instance of the maze.

Averaging across all six trials, the relative proportions of descriptions of each type were; Figural: 25%, Path: 53% and Abstract 22%. Although the proportion of Path-type descriptions remained relatively constant across trials; 51%, 55%, 55%, 52%, 55% and 52% respectively, the relative distribution of Figural and Abstract description types shifted, with a fall in the proportion of Figural description types across trials shadowed by a rise in the proportion of Abstract description types. Calculation of Pearson's product-moment correlation suggested a strong negative relationship between the two description types with r = -0.90. The shift in distribution is illustrated in figure 4.5 with trials 1&2, 3&4 and 5&6 combined to give low, medium and high levels of experience as before.

The reliability of this pattern was assessed firstly by calculating an omnibus χ^2 for the raw frequencies of Figural and Abstract description types across all six trials. This proved significant with $\chi^2_{(5)}=13.28$, p=0.020. Secondly, a focussed

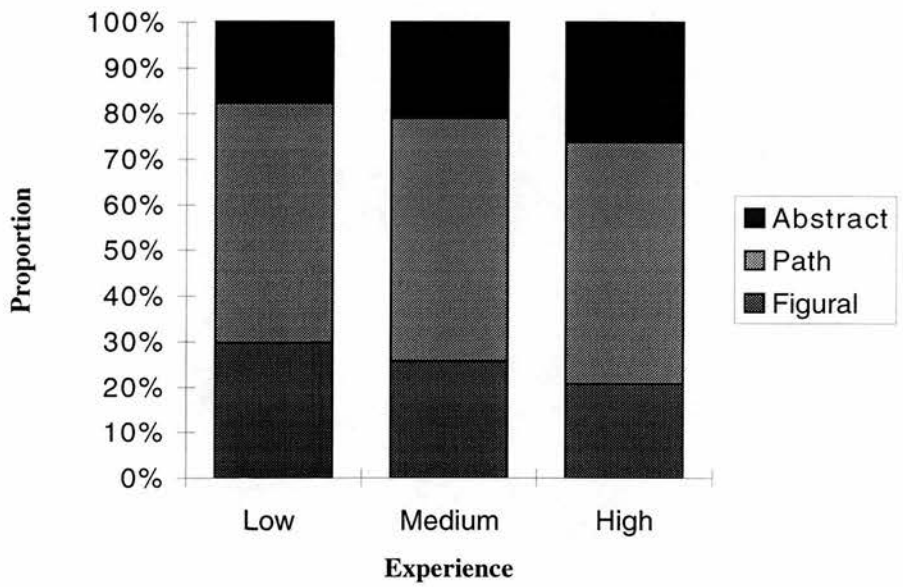

Figure 4.5: Relative Distribution of Description Types According to Experience

comparison was made of the relative frequencies of each description type on trial 1 versus trial 6. This also proved reliable with $\chi^2_{(1)}=10.66$, p=0.001.

The experimental manipulation of group composition was analysed by calculating χ^2 for the frequencies of Figural and Abstract description types in the homogenous and mixed dyads in trial 6. This was significant: $\chi^2_{(1)}=11.00$, p=0.000, with homogenous dyads using Abstract description types for 40% of target locations and Figural description types for 14%, compared with mixed dyads who produced Abstract description types for 19% and Figural description types for 24%. Additionally, comparison of the frequencies of Abstract and Figural decription types used by mixed dyads in trial 6 with those used by all dyads in trial 1 revealed no significant difference: $\chi^2_{(1)}=0.944$, p=0.331.

Turning to local patterns of coordination, entrainment scores, following the method described in Garrod and Doherty (1994), were also calculated for each member of a dyad in each trial. This is an index, varying between one and zero where 1=perfect entrainment, of the tendency for individuals to generate descriptions of the same type as those their partners have just generated. It is calculated as the number of description types produced by an individual that match the preceding description type produced by their partner, divided by the total number of exchanges of description types in that trial. Entrainment scores for trial 1 were not calculated as the average number of transitions was very low (2.4). Logically, the scores obtained for each member of a dyad are independent: one individual could always choose a different decription type from their partner, while their partner always matched description type. However, the entrainment scores for members of a pair displayed a strong, positive, relationship, Pearson's product-moment correlation, r=0.51, indicating that the tendency to match description types by each member of a dyad was related to the degree to which their interlocutors were also matching description types.

The average degree of entrainment displayed remained fairly consistent over trials 2-6: 2:0.49, 3:0.39, 4:0.46, 5:0.44, 6:0.51, with values close to the chance level of 0.40 (calculated as the sum of the squared proportions of each description type on trial one). The scores for each pair[5] were entered into an analysis of variance with trial number as a between-subjects factor. This was not reliable: omnibus $F_{(4,116)}$=0.610, p=0.656. The prediction that entrainment should increase across trials was not supported by a linear trend analysis (unweighted means): $t_{(116)}$=0.479, p (one-tailed)=0.316.

A comparison was also made of the experimental manipulation of group composition. The average entrainment scores on trial 6 were: Mixed: 0.46 and Homogenous: 0.63. The scores for each pair were entered into an analysis of variance with group composition, mixed versus homogenous, as a single, between-subjects, factor. This was not significant: $F_{(1,22)}$=2.34, p=0.139.

4.2.3 Discussion

Drawing on the results of both experiments 1 and 2 it seems warranted to conclude that both proportion of errors and number of items attempted constitute poor measures of semantic coordination between members of a dyad. Although they clearly do vary as a function of experience at the task they are not sensitive to the manipulation of group composition. In the case of errors this may be due, in part, to the lack of feedback subjects receive concerning the accuracy with which they have identified the target location. The only direct cue they have that something is wrong arises where they are unable to interpret all or part of a description given by their partner. This contrasts, for example, with the electronic version of the maze task in which an error is more likely to be detected since it will usually have consequences for subsequent actions. The lack of a similar check on

[5]It was necessary to analyse by pair as the positive correlation between the scores for members of the same dyad undermines the assumption of independence for analysis of variance.

accuracy in the paper version of the task makes the interpretation of errors more equivocal with some undetected errors arising from fundamental asymmetries in interpretation of descriptions and others stemming from 'accidental' sources such as left-right confusions or miscounting. All of this undermines the sensitivity of errors as a dependent measure. This problem feeds into the interpretation of the total number of items attempted. Since subjects were less likely to detect trouble in their interpretation of a description we might expect fewer delays due to cycles of repair, making the total number of items attempted less sensitive to the degree of coordination in interpretation by members of a dyad.

Interpretation of the failure to find any effects of local entrainment, either across trials or between the mixed and homogenous groups, is more vexed. The expected increase in degree of entrainment over trials was not observed, nor was any difference in entrainment found between the mixed and homogenous dyads in trial 6. One possible reason for this derives from differences between the electronic and paper versions of the maze task, which could be implicated in a number of ways. The average degree of entrainment observed here was lower than that reported in previous studies –an average of approximately 0.9 in Garrod and Doherty (1994) compared with 0.5 here. This may be partially due to the fact that Garrod and Doherty found strong convergence, in their community group, on the matrix scheme, a scheme practically absent from this corpus. This is the most abstract description type and its presence may well be a cause, as much as a consequence, of a high degree of entrainment. Also, Garrod's and Doherty's subjects had a greater opportunity for convergence, participating in nine trials, as opposed to six, each of which involved a greater number of transitions between speakers. The tighter constraints that the electronic task places on accuracy also provide a possible explanation since this might well be expected to influence the extent to which individuals coordinate.

Despite these reservations, the overall distribution of description types does

provide strong evidence of increasing convergence on the use of particular description types across trials. The change in relative proportions of each type indicates a move from relatively specific, context dependent forms, such as Figural and Path descriptions, to more abstract types, such as Line and (occasionally) Matrix, that provide a scheme which generalises well to new instances of the maze. Figural descriptions are highly context dependent, drawing on specific details of configuration or features that are unlikely to be repeated from one item to another. Conversely, Line type descriptions effectively preserve invariant information about the grid common to the different maze configurations, providing a scheme according to which the description of a range of target locations can be generated (cf. Garrod & Anderson, 1987). While this pattern fits well with the predictions of an explanation based on increasing experience with the task, it does not account for the difference between the mixed and homogenous dyads in trial six. The contrast between these groups suggests that the observed convergence depends to a critical degree on membership, even though unacknowledged, of a particular sub-community. The contrast in the distribution of description types between the mixed and homogenous dyads shows that the degree of convergence achieved over the course of five trials can be readily disrupted by transfer outside a subgroup, the observed distribution of description types produced by mixed dyads not differing significantly from those of dyads attempting the task for the first time.

Interestingly, this presents something of a contradiction since, given the consistently low, almost chance, level of entrainment across trials, this pattern cannot, without elaboration, be attributed to input-output coordination. Full discussion of this issue is deferred until section 4.3.3. Similarly, although some aspects of the data are amenable to analysis in terms of the collaborative model of dialogue, it cannot account for the contrast between the mixed and homogenous groups. The question then arises as to how both the shift in the overall distribution of description types and its sensitivity to group composition can be accounted for.

A further study was conducted in order to try to resolve some of these issues.

4.3 Experiment 3

In order to investigate further the mechanisms that drive the shift in patterns of description type, and to provide a check on their reliability, a third experiment was run, again utilising the paper version of the map task.

4.3.1 Methods

The methods adopted for this study were basically the same as those for experiment 2 with some modifications. Some of the difficulties in the interpretation of the previous study stemmed from the relatively low number of items attempted by each individual and the low incidence of Matrix description types in the corpus. In order to try to compensate for this, the materials were substantially altered and the number of items in each trial, as well as the time allowed for completion, were increased.

Materials

The transcripts from experiment 2 suggested several sources of difficulty for subjects. The use of screen dumps from the original task had resulted in the inclusion of a number of features, such as switch points, gates and the window border, that are strictly irrelevant to the completion of the task on paper. These were frequently utilised in Figural descriptions and their presence appeared to promote the use of this description type, possibly at the expense of the more abstract schemes. This observation prompted the removal of these features from the maze configurations. It was also noted that because both the maze and its background were white, a number of subjects, particularly on early trials, fell prey to figure-

ground ambiguities leading to problems in identifying which elements of the maze were to be interpreted as boxes. To overcome this, the background was shaded grey, making the pattern of boxes more apparent. As well as removing some of elements that favoured Figural descriptions, the size of the basic grid was increased to 5-by-6, increasing the range of possible targets, and enhancing the effectiveness of Line/Matrix description types relative to Path and/or Figural which should, on average, become more complex for larger grids. An example of the resulting configurations are given in figure 4.6.

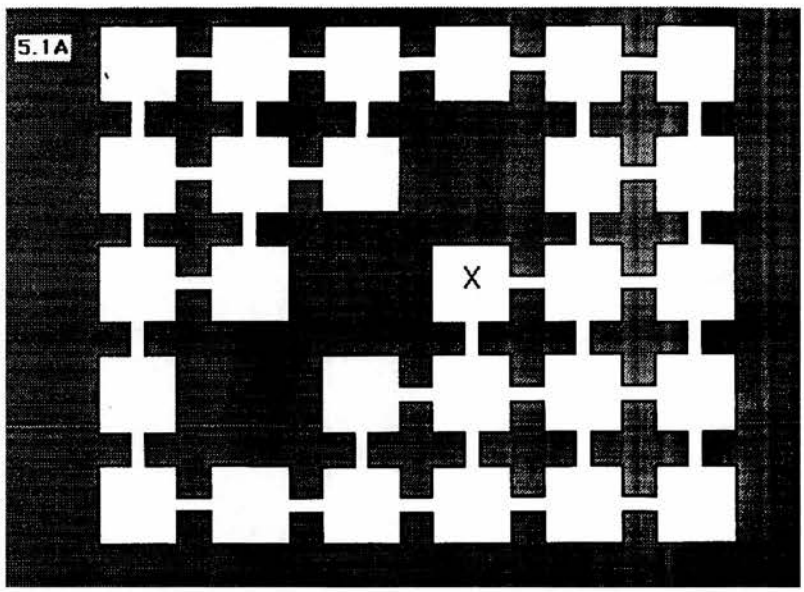

Figure 4.6: Example Configuration

Following the same strategy as in experiment 2, 120 basic configurations were generated and then made up into two sets, one with a location marked (this time by an X) and one without. These were divided up into twelve sets of 20 items, each forming a pair of booklets with target locations on alternate pages paired with the appropriate unmarked configuration on the corresponding page in the

other booklet. This increased the number of descriptions each subject aimed to generate up to a possible maximum of ten.

Design

The design of this study was unchanged from experiment 2, employing the same, counterbalanced, ordering of materials and composition of experimental groups. As before, this resulted in a simple factorial design with the experimental manipulation as a between-subjects independent variable of dyad composition (mixed vs. homogenous) and a within-subjects variable of number of trials completed. The descision to abandon number of items attempted and proportion of errors per pair as dependent measures allowed a reduction in the sample size to 24.

Subjects

Twenty-four subjects were recruited from amongst the staff and students of various disciplines at the University of Edinburgh. They consisted of thirteen males and eleven females with ages between 20 and 47 (average age: 25 years). Each was paid £3 for participating.

Procedure

Two substantive changes were made to procedure. Firstly, the length of time allowed for completion of each trial was increased to 3 minutes to promote completion of a larger number of items. Secondly, the instructions were altered with each item now referred to as a grid rather than a maze. This was principally motivated by the very weak resemblance each item now bore to the original mazes but was also intended to provide an extra prompt toward the use of more abstract description types.

4.3.2 Results

The changes to procedure and materials had the desired effect of producing a consistently higher average number of items attempted by each pair across trials: 1:13.4, 2:16.0, 3:16.4, 4:17.2, 5:18.6 and 6: 19.3, with 28 out of 72 dyads completing all twenty items within the time allowed. The number of items attempted by each dyad were entered into an analysis of variance with trial number as a single, between-subjects factor. This confirmed that the pattern of increase across trials was reliable: omnibus $F_{(5,66)}$=5.234, p=0.000, linear trend; $t_{(66)}$=3.503, p (one-tailed)=0.000.

Transcriptions

71 three minute dialogues were transcribed (one dyad excluded due to a failure to record) and analysed, following the same criteria as before, for the presence of the four main description types: Figural, Path, Line and Matrix. As well as recording information about each description's origin, order in which the axes were introduced and type of enumeration employed (cardinal versus ordinal), each description was also coded as 'challenged' where it was subject to clarification or repair and 'accepted' where a pair either moved straight on to the next item on completion of a description, a description prompted only a simple acknowledgement such as "okay" or "right" or where it was subject to a verbatim repeat that did not prompt any further exchange other than an additional acknowledgement.

The dialogues generated a corpus of 1,207 descriptions with all four description types represented: Figural:9%, Path:36%, Line:26% and Matrix:29%. The observed proportion of Matrix description types is comparable to those reported in Garrod and Anderson (1987), Garrod and Doherty (1994), confirming the effectiveness in the changes to the materials and procedure in promoting this scheme. As before, the relative proportions of each description type displayed a shift across trials, illustrated in figure 4.7. Trials 1&2, 3&4 and 5&6 are pooled to give low,

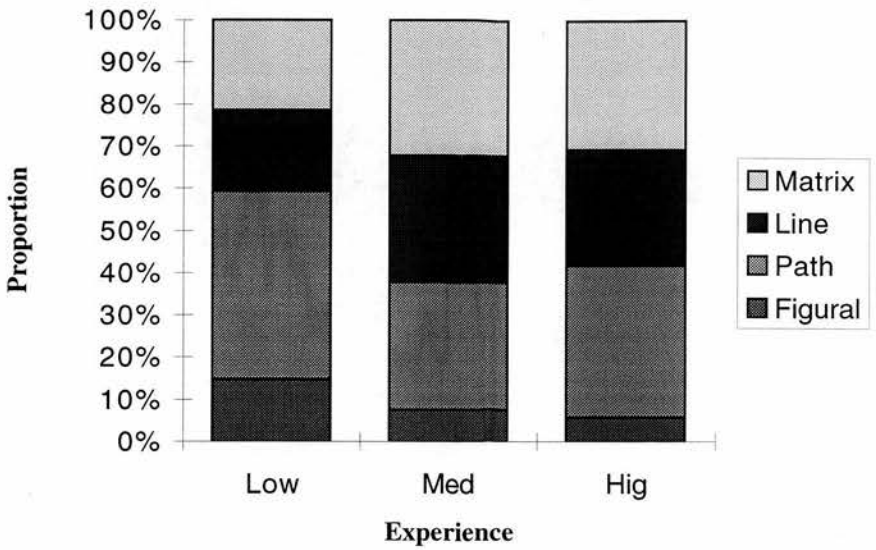

Figure 4.7: Relative distribution of Description Types According to Experience

medium and high levels of experience with the task.

The overall pattern is very similar to the previous study with Figural description types falling across trials, matched by a rise in the more abstract Line and Matrix description types. Pearson's product-moment correlation calculated between the proportions of Figural and Matrix-type descriptions in each trial indicates a strong negative relationship, r= -0,74, and there is a reliable difference in the pattern of raw frequencies of Figural and Matrix-type descriptions across trials: $\chi^2_{(5)}$=35.068, p=0.000.

Taking the Figural and Matrix description types separately, the proportion produced by each dyad, ignoring the mixed dyads in trial 6, was entered in an analysis of variance with trial as a between-subjects factor. For Figural-type descriptions omnibus $F_{(5,59)}$=1.59, p=0.17, and linear trend analysis (unweighted means) confirmed the presence of a regular decrease across trials: $t_{(65)}$=2.10,

Type:	Figural	Path	Line	Matrix
Mixed:	16%	40%	39%	5%
Trial 1:	18%	48%	15%	19%

Table 4.5: Description types in Trial 1 and Mixed (Trial 6)

p(one-tailed)=0.019. The parallel analysis for Matrix description types gave omnibus $F_{(5,59)}$=0.543, p=0.74, with weaker support for the pattern of increase across trials; linear trend analysis, $t_{(59)}$=1.540, p(one-tailed)=0.064.

Turning to the experimental manipulation, the frequencies of all description types, in the homogenous and mixed groups, were reliably different: $\chi^2_{(3)}$= 129.62, p=0.000, the relative distribution is illustrated in figure 4.8. Although comparison of the frequencies of all description types in the mixed dyads on trial 6 and all dyads on trial 1 were reliably different: $\chi^2_{(3)}$=26.28, p=0.000, both groups display a similar preference for Path and Figural description types while differing in the relative proprtions of Line and Matrix-type descriptions (see table 4.5). When the frequencies of Line and Matrix-type descriptions are pooled to form a single category of 'Abstract' descriptions, as for experiment 2, no reliable difference is found: $\chi^2_{(2)}$=3.34, p=0.187.

The proportion of descriptions subject to repair or clarification, ignoring the mixed dyads in trial 6, displayed a steady pattern of decrease across trials illustrated in figure 4.9. Ignoring descriptions that were incorrect but not subject to clarification or repair (for the reasons discussed in experiment 2), this data was analysed in an analysis of variance, with trial number as a between-subjects factor. This confirmed that the trend was reliable with omnibus $F_{(5,59)}$=1.020, p=0.41, linear trend analysis; $t_{(59)}$=2.05, p (one-tailed)=0.022.

The frequency with which descriptions were subject to repair or clarification versus accepted proved to be reliably different for the mixed and homogenous dyads: $\chi^2_{(1)}$=6.543, p=0.010, with homogenous dyads repairing or clarifying 16% of descriptions and mixed dyads repairing or clarifying 37%. Also, frequency of

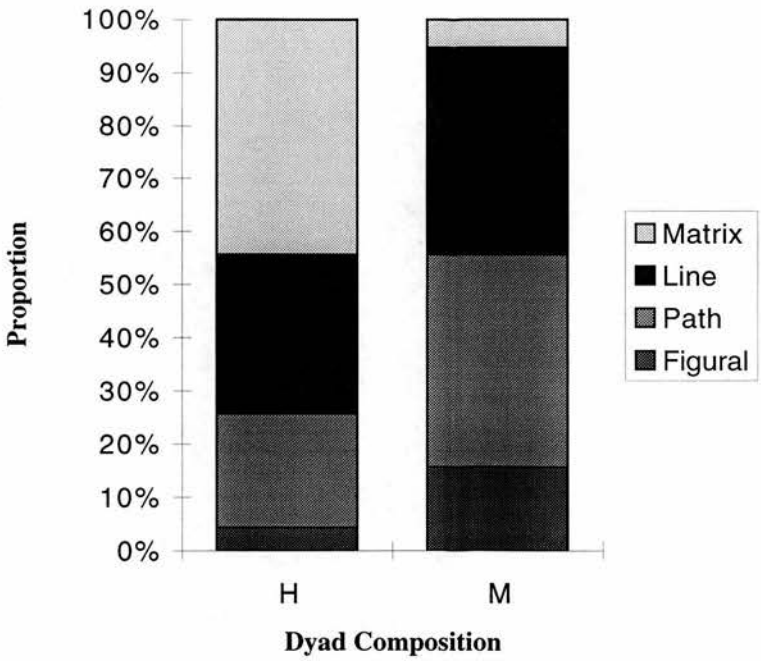

Figure 4.8: Distribution of Description Types in Homogenous (H) and Mixed (M) Dyads

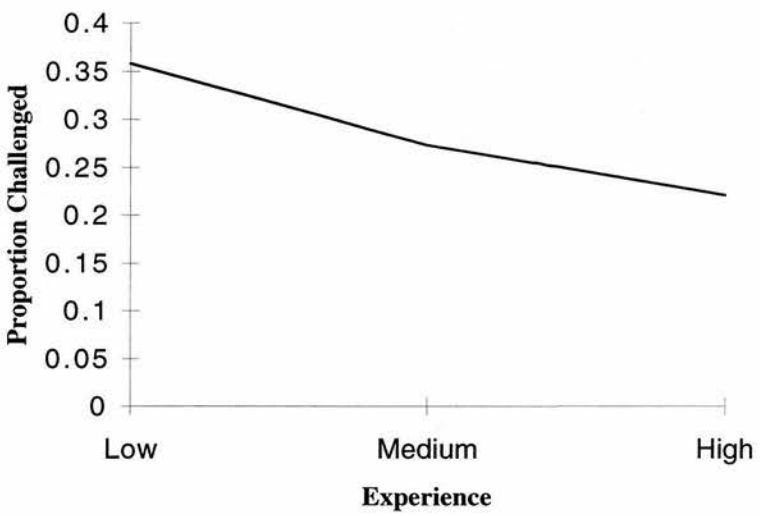

Figure 4.9: Proportion of Descriptions Repaired or Clarified with Experience

repair/clarification in the mixed dyads (trial 6) was not reliably different from that observed in all dyads on trial 1: $\chi^2_{(1)}$=1.135, p=0.287.

The entrainment scores, calculated for each subject, showed some interesting contrasts with experiment 2. The chance level of entrainment was lower; 0.32 compared with 0.40 previously (this is expected given the greater range of description types observed in the corpus). Conversely, the average degree of entrainment was higher than in experiment 2: 0.57 compared with 0.46 previously (although still lower than the average of 0.9 found by Garrod and Doherty (1994)). Across trials, the degree of entrainment was, again, relatively constant: 1:0.58, 2:0.58, 3: 0.54, 4:0.58, 5:0.56 and 6:0.54, but, unlike experiment 2, consistently above chance. Examination of the entrainment scores between members of each dyad revealed a similar, but substantially higher, positive correlation than before: Pearson's product-moment: r=0.90, but there was, again, no reliable increase in entrainment across trials. An analysis of variance on the average entrainment score for

each pair gave an omnibus $F_{(5,66)}$=0.054, p=0.99, linear trend: $t_{(66)}$=0.311, p (one-tailed)=0.378.

4.3.3 Discussion

Smaller sample size and changes to materials and procedure notwithstanding, these results do replicate the observations of experiment 2, providing extra support for the analysis. Dyads show the same shift from Figural toward more abstract description types across trials, a shift which is, again, dependent on membership of a sub-group. Even though experience with the task is equivalent for all individuals in the sixth trial, dyads composed of individuals drawn from different sub-groups show a reliable difference from dyads composed of individuals from the same sub-group. This difference occurs despite the fact that subjects receive no cues concerning the partitioning into different sub-groups and no one, in either experiment, reported detecting this manipulation. Where the homogenous dyads conform to the trend of producing fewer Figural description types and more Matrix on trial 6, the patterns observed in the mixed dyads are, in several respects, no different from those observed amongst individuals attempting the task for the first time. Although they still manage to attempt a high number of items and make fewer errors, the description types employed, and the frequency of repair or clarification of descriptions, are notably similar to those attempting the task for the first time.

The trend, across trials, to make fewer errors and deal with more items in the time allowed can be unproblematically attributed to greater expertise at the task, where it is understood as communicating target locations to a range of different individuals. Although the shift in patterns of description type produced, and frequency of repair/clarification, across trials also seem to admit an expertise-based explanation, the contrast between the mixed and homogenous dyads, which depends on membership of sub-groups, does not. This contrast also raises difficulties

for any explanation that appeals to the pairwise establishment of mutual beliefs since every individual, on every trial, meets a new partner and, according to this type of model, should have the same requirement for re-establishing the pertinent mutual beliefs on each occasion regardless of whether they are members of mixed or homogenous dyads.

A more promising candidate is Garrod and Anderson's (1987) principle of input-output coordination which does offer an account of group-based effects and does not depend on every individual establishing mutual beliefs with each other individual about the appropriate way to refer to a target location (see section 3.2.4). However, as noted earlier there also seem to be problems for this explanation.

The gross degree of coordination observed in experiments 2 and 3, as measured by entrainment scores, was effectively constant across trials and, although higher than chance in experiment 2, appears to be independent of the shifts in description types found here. To an extent, this is unsurprising since entrainment scores calculate the degree of matching regardless of the type of description produced. Thus, perfect entrainment on Figural descriptions is, on this index, equivalent to perfect entrainment on Matrix descriptions. However, an important element of the argument here has been that we can invoke a partial ordering of description types according to the degree of coordination they imply: matching of Figural descriptions represents the weakest degree of coordination and Matrix the strongest. The reasoning is that the generalisations possible from one Figural description to another are very weak whereas Matrix descriptions can potentially invoke the same order of axes, the same origin and the same counting scheme. This is supported by the fact that where Matrix description types tend to be highly elliptical, utterances often amounting to just two numbers, e.g., "three four", Figural description types are almost always extended, produced in installments over a number of turns and involve several stages of checking for comprehension. Effectively, Figural descriptions are the lowest common denominator, calling only on the pre-established

linguistic coordination that each individual brings to the task in the first place whereas Matrix descriptions call on local, more specific, conventions established during the course of the task (cf. Garrod & Anderson, 1987).

As a result, entrainment scores provide only an approximate index of coordination which is not sensitive to the reliable pattern, observed in the mixed dyads, of switching toward more primitive description types, since both individuals tend to make the same switch. It seems that, as indexed by entrainment, input-output coordination cannot easily account for this switch. However, this is only one element of the mechanism considered by Garrod and Doherty. They also propose that where conflicts arise, in their case where a pair have coordinated on different description schemes in their previous game, they are resolved by a shift to the description type most commonly used by both players in all previous games. This is supported by their data from their community group which conform to this pattern in 8 out of 9 cases of possible conflict, the apparent exception being an artifact of the coding scheme.[6] However, this assumption does not hold true for the mixed dyads, where conflicts are most likely to occur, in experiments 2 and 3. Of the 18 mixed dyads 7 went against this pattern and 11 conformed. Furthermore, ignoring those dyads that had used the same description type in the previous game, the figures are 7 against and 8 conforming.[7] It appears that mixed dyads were most likely to shift not to the most commonly used previous scheme, but to the most basic scheme, with the two frequently coinciding.

An additional problem arises concerning the patterns of repair/clarification

[6] For this pair, although Matrix was the most common previous scheme type, they had adopted versions with different, conflicting, labelling systems for the axes.

[7] The chance level of independently switching to the same scheme is difficult to calculate precisely here but, intuitively, will be high since there were only 3 common description types in experiment 2 and 4 in experiment 3. Therefore if both players shift there is only one possible alternative in experiment 2 and two possible alternatives in experiment 3. If only one shifts there are 2 and 3 possible alternatives respectively. Of course, much depends on the specificity of the coding into types.

observed in experiment 3. There are two aspects to this: firstly, the reliable fall in repair across trials was not matched by a corresponding increase in entrainment. Secondly, while the mixed dyads engaged in reliably more repair/clarification than the homogenous dyads (roughly twice as frequently) their degree of entrainment was almost identical: 0.53 and 0.54 respectively. The occurence of repair appears to be independent of input-output coordination. Similarly, the additional proposals for group-based mechanisms do not account for this pattern, once pairs have shifted schemes in response to conflict there is no reason, on the input-output coordination model, to expect that repair should increase.

The problems with extending the group-based mechanisms proposed in the input-output coordination model to the results reported here suggests that the processes giving rise to coordination are approximated by this model rather than explained by it. That is, whatever the mechanisms are which are responsible for coordination, they result in input-output coordination rather than being caused by it. If this holds, the outstanding problem is to identify a mechanism by which individuals achieve the degree of coordination implied by the shift in description types.

4.4 Coordination Through Repair

Drawing on the discussion of previous chapters and the experiments reported above, the suggestion advanced here is that the mechanism of semantic coordination is located in the process of clarification and repair. There are several steps to developing this claim. The central assumption, one developed throughout this thesis, is that idiolectical variation is pervasive. In the context of the maze task this becomes the claim that the interpretations of a particular description by any

given pair of individuals will differ to some arbitrary degree.[8] Drawing on the discussion of Garrod and Anderson, it is reasonable to suppose that individuals generate descriptions of a target position according to their current conceptualisation of the maze/grid and, similarly, interpret descriptions according to the same conceptualisation. Given that these conceptualisations will vary, how do individuals manage to coordinate their actions using them?

One, apparently natural suggestion is that individuals overcome this problem by discussing their interpretations and arriving at an explicit, negotiated solution. However, this strategy would not be expected to succeed since it faces a marked bootstrapping problem. If we take idiolectical differences to obtain in both the 'meta-language' (in this case some dialect of English) and the 'object language' (in this case the expressions relating to locations in the maze, such as "rows", "columns" etc.) there is no guarantee that asking "what do you mean by rows?" will not itself receive a response that is open to misinterpretation. The empirical findings fit with this: Garrod and Anderson (1987), Garrod and Doherty (1994) both observe that explicit negotion does not seem to be effective in improving coordination. Negotiation is relatively rare and where it does occur the immediately succeeding description frequently deviates from whatever agreement had been reached. To avoid the threat of regress in the strategy of explicit negotiation, the proposed solution here is that, all things being equal, individuals apply a principle of charitable interpretation. They proceed with the attempt to coordinate behaviour, assuming that their descriptions are being interpreted as they intend, until some evidence of trouble arises. Where trouble does occur, they engage in local, minimal, repair which, with respect to the constraints provided by the current context of the task, appears to resolve the problem and then move on. While this does not provide any guarantee that there will not be subsequent problems, it

[8]Of course, their membership of a (sub)community of English speakers means they are unlikely to be completely orthogonal.

does offer a mechanism by which local coordination can improve, with each cycle of problem and repair moving the dyad towards better coordination. The limit on the degree of convergence achieved, is set by the constraints the task places on the coordination of behaviour and allows that their interpretations may still diverge to a substantial, but lower, degree.

So far, this basic proposal treats the steady fall in repair/clarification across trials observed here as a consequence of individuals' increasingly meeting the limits of accuracy required by the task. It also provides a plausible explanation of the differences between these studies and the previous experiments on the maze task. The stronger convergence on the Matrix scheme in Garrod's and Doherty's (1994) community group can be viewed as consequence of the tighter constraints the dynamic task places on the accuracy of interpretation; disparities in interpretation are more likely to be detected, in subsequent moves, increasing the likelihood of repair and, according to the current proposal, therefore greater coordination. Conversely, the weaker constraints on accuracy of interpretation in the current study allow for a greater degree of residual ambiguity, a greater range of description types persisting in later trials and a lower degree of coordination. All things being equal, entrainment will be higher where the premium placed by the task on convergence in interpretations is higher.

The more substantive problem is to account for the differences between the mixed and homogenous dyads in the current study and the differences between the community, control and isolated pairs in Garrod and Doherty (1994). Assuming a background of idiolectical variation in a population, the relative balance of asymmetries in interpretation will differ from pair to pair. For example, one pair may differ more widely on the interpretation of "row" relative to a another pair who differ more widely on the interpretation of "column". To the extent that the constraints imposed by the task highlight their differences, each pair will find different aspects problematic and engage in different kinds of repair, generating

uneven patterns of convergence. Importantly, where the pairs form a coherent sub-group that perform successive trials with individuals drawn from the same pool, this explanation predicts a regression towards some 'average' degree of asymmetry between members of the group. As the history of interactions with common individuals increases so the range of asymmetries between each pair will reduce. The detection and repair of task-relevant differences in interpretation over trials will drive the group toward convergence on some group-based optimum with respect to the task.

A second consequence of this explanation is that the type of convergence that emerges in a sub-group will be different for different groups; all things being equal, the initial balance of asymmetries in any subset of individuals will be different, giving rise to different patterns of convergence in each group. The semantic resources built up within a group are thus expected to be specific to that group. Consequently, transfer outside a subgroup will, on average, confront individuals with a situation in which the problem of coordinating on the interpretation of descriptions is similar to that they faced on the first trial, giving rise to a marked disturbance in coordination. This explains both the increased repair observed in the mixed dyads and the shift toward description types that rely on membership of the broader linguistic community from which the sample was drawn as opposed to the linguistic sub-community of the group from which the individuals transferred. This also provides a reason why individuals faced with a conflict within a community group shift to the scheme most commonly used by both individuals in previous games (Garrod & Doherty, 1994) whereas individuals faced with a conflict deriving from transfer between sub-groups tend to shift to the most basic scheme. In the former case the community group will already have achieved some degree of convergence and, when faced with a conflict, can still utilise this, shifting to a scheme that takes advantage of it. By contrast, the mixed dyads have no common resource beyond membership of the wider linguistic community from

which they are drawn at the start of the experiment. As a result they retreat to description types which draw on the common ground that does exist between them, the dialect of English shared prior to the task.

Turning to Garrod and Doherty's data, the isolated pairs are predicted to show a lower degree of coordination, and a wider range of description types, as a consequence of the fact that they are only faced with accommodating one set of asymmetries. Performance of the task precipitates repair of problems specific to the differences in their interpretations but leaves untouched those disparities that do not become apparent. With different partners, different problems are likely to arise, provoking convergence in different areas. As a result isolated pairs will show limited convergence and maintain a wider range of description types. This reasoning also explains why the non-community group, in virtue of their exposure to a succession of different partners, are predicted to develop a higher degree of coordination. However, since there is almost no common history of performing the task with other individuals, and the consequent process of accommodating the problems that arose in coordinating with them, the degree of coordination is not predicted to reach that of a full community group.

Overall, this explanation fits well with the data and resolves some of the tensions between these and previous findings. To the extent that it is successful in dealing with the empirical findings it gives additional support to the central assumption on which it depends, namely, that idiolectical variation is pervasive and creates problems for the maintenance of mutual-intelligibility. Importantly, this explanation only extends to the semantic coordination demanded by a co-operative task. Other factors that also affect performance, such as experience with task procedure, role differentiation and possible effectiveness of ratified versus unratified observers do not fall within its scope and the predictions of existing models with respect to them are unaltered.

Chapter 5

A Channel Theoretic Model

The preceding chapters have developed the claim that formal models of natural language semantics, by idealising to a single ontology or set of semantic primitives, imply a code theory of mutual-intelligibility; an implication explicitly realised in contemporary models of communication. Idiolectical variation, it has been argued, undermines this idealisation and any model which aims to account for the mutual-intelligibility of dialogue must accommodate disparities in interpretation between different parties to it. In particular, it has been proposed that an adequate formal semantic model that addresses communication between different agents must accommodate a degree of ontological pluralism between them. *Prima facie*, attempts to naturalise meaning by appeal to cognitive states seem promising candidates for acheiveing this since they can allow the cognitive states associated with interpretation to vary between different individuals. However, it has also been argued that this approach is, itself, undermined by the arguments that meaning cannot be directly reduced to cognitive states without violating important intuitions relating to its normative character. The distributed, or broad, nature of semantic content appears to defy analysis in terms of groups, or subgroups, of individuals' mental states (narrowly understood); a consequence which substantially weakens

any claim models which take this approach have on providing a genuine semantic theory.

This chapter aims to make progress in developing a semantic analysis of idiolectical variation while respecting the intuition that the content of any utterance is determined by reference to the socio-physical context in which it occurs; achieving this in a way that preserves the ambition of naturalising semantics. To do so it must reconcile the tension between idiolectical variation on the one hand and distributed content on the other.

The importance of developing a semantic framework resides in the adjunct it can provide to principle-based accounts of communication. While empirical models such as the collaborative model and the input-output coordination model offer principles which operate to improve semantic coordination between the parties to a discourse, they are not equipped to provide a detailed anatomy of when and why coordination fails. The emphasis, of both accounts, on characterising successful coordination has the consequence that, where the principles they offer are violated, little can be said about the nature of the violation or its likely effects on the subsequent conduct of dialogue (although see Garrod & Doherty, 1994). A parallel concern arises with theoretical accounts which, covertly and overtly, appeal to principles of charity in interpretation to explain the way idiolectical variation is often discounted in ordinary discourse (e.g., Putnam, 1981; Schutz, 1973). Again, these are principles which apply in the ideal case and would be considerably strengthened if they could be supplemented by some indication of under what conditions interpretation should be charitable and when it should not. A similar weakness is apparent in the repair-based model of coordination proposed in Chapter 4. While, arguably, it offers greater coverage of the empirical data it gives no more than a general specification of the conditions that might provoke repair, nor any indication of likely patterns of response.

Progress with any of these issues ultimately requires a model of idiolectical

variation at a level of analysis that addresses the specific conflicts that may arise between idiolects and characterises their consequences for the maintenance of mutual-intelligibility. With this goal in mind, the following analysis draws on the dialogues generated by the maze task experiments of the previous chapter, and aims to give a semantic characterisation of some of the phenomena observed. A promising framework in which to attempt this, and one that does not require assumptions that run contrary to those made by this thesis, is channel theory.

5.1 Channel Theory

Channel theory is a development of situation theory that aims to provide a formal, structural, characterisation of natural regularities which reconciles two key properties; their *reliability* and their *fallibility* (Barwise & Seligman, 1994, 1993; Seligman & Barwise, 1993). Taking up a theme explored in the original formulation of situation theory (Barwise & Perry, 1983), channel theory is concerned with the nature of the 'reliable connections' that make our knowledge of the world (sometimes) dependable. Reliability is important in accounting for successful representation, knowledge, truth and inference. Fallibility is important in explaining misrepresentation, error, falsity and defeasible inference. From the outset, this focus makes channel theory a promising framework for dealing with issues relating to the finite, error prone, information processing normally cited in cognitive models of human performance. Additionally, there is an explicit concern with the problems of naturalising intentional states, a concern shared by the discussion of Chapter 2. Overall, the goal of channel theory is to develop a naturalised theory of information flow, in the sense of Dretske (1981), that provides a structural account of the reliability and fallibility of regularities and is general enough to accommodate both formal logics and situated, defeasible, reasoning.

5.1.1 The Analysis of Regularities

An enduring problem for the attempt to naturalise regularities has been the difficulty of finding a plausible way to specify background conditions that underwrite their reliability. Because regularities are normative, stated *ceteris paribus*, they need to be exception bearing, however, the specification of conditions under which a regularity does in fact hold, for example by determining necessary and sufficient conditions for its application, has proved highly problematic. Barwise and Seligman (1994) argue that the most popular solution to this problem, which involves analysing regularities as conditionals, actually undermines the project of naturalising the relationships captured. The basic reason for this is that the semantics of conditionals are usually analysed in terms of relations between sets of possible worlds. This is quite successful in accounting for the entailments between conditional statements but leaves open the question of how the actual world might determine truth conditions like those generated by possible worlds. There appear to be two alternatives. The first, endorsed by Lewis (1972a), is to suggest that possible worlds are, in fact, real; a response strongly counter to naturalistic intuitions. The alternative, pursued by Stalnaker (1987), is to treat possible worlds as epistemic objects, instantiated by the beliefs of the individual reasoning about particular conditionals. The problem for this account is that the cognitive capacities invoked in this way must in turn be given a naturalistic reduction, raising the problems discussed in Chapter 2. A direct reduction of the meaning of conditionals to cognitive states will violate the intuitions associated with broad content.

5.2 Basic Apparatus

Instead of attempting to refine the conditions under which a regularity holds, formulating progressively finer necessary and sufficient conditions that determine when it applies, the strategy followed in channel theory is twofold. Firstly, it

emphasises the distinction between tokens, or parts of the world, and the connections between them that support a particular regularity, and types, and relations between them, that express the regularity. Secondly, taking up a proposal in Seligman (1990), it relativises the notion of regularity to classification domains, providing local 'theories' of particular domains and avoiding commitment to a total or definitive theory that exhaustively captures all possible regularities.

5.2.1 Classifications

Intuitively, a classification is a way of grouping a set of things into some category. Thus the type "Fiction" might, amongst other things, group together all books and short stories about imaginary people and events. By contrast, the type "Autobiography" might group together books and short stories about real lives and events. Of course, some autobiographies are calculated works of fiction and some works of fiction are highly autobiographical. This is an important point, the same things can be classified in many different, even conflicting, ways. People can be simultaneously classified as terrorists and freedom fighters. The same day can be simultaneously classified by a date, as the first day of the sales, as a birthday and as a national holiday. Although classifications can be derived on the basis of a range of criteria, not all are felicitous. People cannot, equivocation and nonce useage aside, naturally be classified as national holidays or literary works. Channel theory does not assume that there is any universal or 'correct' classification, rather it draws on the intuition that there are a range of possible classifications for any object, some of which are more natural than others. There is no attempt to define what makes one classification more natural than another, rather a pluralistic, intuitive view is taken on what classifications can apply to what things.

Formally, a classification A is modelled as a triple, $A = \langle T, S, \models \rangle$, where T is set of types $\{T_1, T_2, T_3 \ldots T_n\}$, S is a set of tokens $\{s_1, s_2, s_3, \ldots s_n\}$, referred to as sites, and $\models$ is a relation on $S \times T$ where $s_1 \models T_1$ is read as s_1 is classified as

being of type T_1. Following Seligman (1990), this structure can be conveniently represented in a two level diagram as in figure 5.1, where the set of types associated with a classification are rendered as elliptical bubbles and the set of tokens are rendered as a plane. The dotted line between a type and a token indicates that the relation $\models$ holds between them.

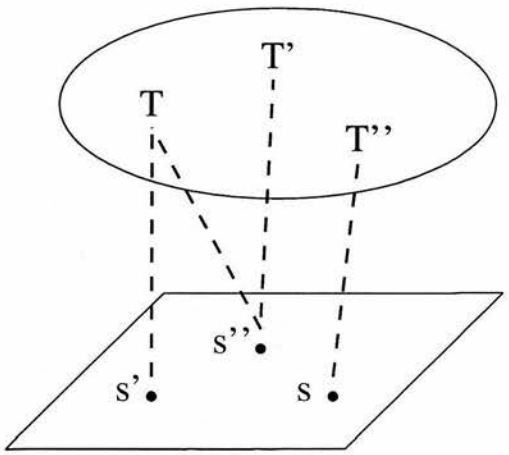

Figure 5.1: Simple Classification

Taking the example of the maze task, a classification might include the description types; "two three" (Matrix), "third row, second box" (Line), "two up, two along" (Path) or "above the space" (Figural) and the corresponding set of tokens might consist of the various screen displays classified by the description types. The same token screen configuration can be classified by several description types. For example, one position could correspond to all the types above and each description type can classify a range of token screen configurations.

In order to characterise informational relations between classifications, Barwise and Seligman introduce the notion of *infomorphism* which provides a mapping

between classifications.[1] Given two classifications $A = \langle T_A, S_A, \models_A \rangle$ and $B = \langle T_B, S_B, \models_B \rangle$ an infomorphism $f : A \rightleftarrows B$ is a pair of functions, $f^\vee : S_B \to S_A$ on the tokens and $f^\wedge : T_A \to T_B$ on the types such that $f^\vee(s_b) \models_A T_A$ iff $s_b \models_B f^\wedge(T_A)$.

Barwise (1995) illustrates this idea with the example of a translation of a classical, logical language L_1 in L_2 in which each sentence A in L_1 is associated with a sentence $f^\wedge(A)$ in L_2 and each truth assignment, or structure s, for L_2 is associated with a truth assignment or structure $f^\vee(s)$ for L_1. The reason that the relation $f^\vee$ goes in the 'reverse' direction is that, where the translation is an infomorphism, the structures for L_2 must be rich enough to preserve the entailment relations defined for L_1. More generally, infomorphisms are homomorphisms that preserve subclassification relations.

In the maze task, a translation of Path descriptions into Matrix descriptions associates each Path description type with a Matrix description type and associates each location classified by a Matrix description type with a location classified by a Path description type. This would be an infomorphism just in case each location classified by the Matrix description type M maps to a location classified by a Path description whose translation is M. This formulation allows for situations in which, a location corresponding to several description types in the Path classification translates to a single description type in the Matrix classification. For example, assuming an origin of bottom left, a location classified by the Path description types "two up, two along" and "three along, one up" can both be translated as the Matrix description type "two, three". Similarly, a single location in the target classification of the translation may correspond to several locations in the source classification of the translation. For example, a translation of Path descriptions into Figural descriptions might map locations classified by the Path

[1]The definitions used here are those presented by Barwise (1995) and differ somewhat from earlier treatments.

description types; "one in" "two in, one up" and "two in, two up", to a single location classified by the figural description "Left leg".

Treating translations between description types in terms of direct infomorphisms between the appropriate classifications is too restrictive. In practice, there are many informational relationships between classifications that are better modelled in a way that imposes weaker constraints. For example, it is not plausible to assume that all the subclassification relations of Path description types are preserved by translations into, say, Figural description types. To provide a more general notion of information flow, Barwise and Seligman elaborate the notion of infomorphism to give a characterisation of information channels.

An *information channel*, $\mathcal{C}$, is an indexed family of infomorphisms $\{f_i : A_i \rightleftarrows C\}_{i \in \Sigma}$ with a common target, illustrated in figure 5.2.

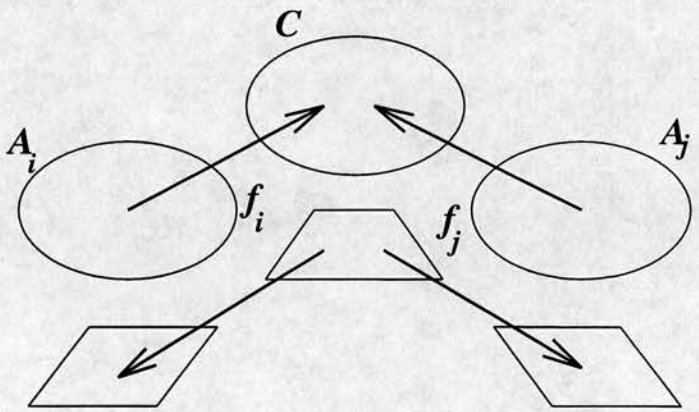

Figure 5.2: A Channel

C is a classification which forms the *core* of the channel $\mathcal{C}$ and its tokens are referred to as *connections*. An information channel, $\mathcal{C}$, between the classification, P, $\{p_1 \models P_1, p_2 \models P_2, p_3 \models P_3 \ldots p_n \models P_n\}$ of displays classified by Path description types and the classification M, $\{m_1 \models M_1, m_2 \models M_2, m_3 \models M_3, \ldots m_n \models M_n\}$

of displays classified by the Matrix description types consists of the two infomorphisms; $\{f_1 : P \rightleftarrows C\}$ and $\{f_2 : M \rightleftarrows C\}$. Although the classifications connected by a channel are infomorphic with respect to the core, they are not necessarily infomorphic with respect to each other. In the simplest terms, this is because an infomorphic mapping needn't be symmetric (cf. isomorphism, where this is the case).

Information flow along a channel, $\mathcal{C}$, between classifications P and M derives from two aspects of the channel. Firstly, the presence in the channel classification of connections, or *signalling relations*, $s_i \in \mathcal{C}$, written $p_j \rightsquigarrow_{s_i} m_k$ defined so that $f_1^\vee(s_i) = p_j$ and $f_2^\vee(s_i) = m_k$. Secondly, the presence of *indicating relations*, $T_i \in \mathcal{C}$, between P_j and M_k, written as $P_j \Rightarrow_{T_i} M_k$ defined so that $f_1^\wedge(T_i) = P_j$ and $f_2^\wedge(T_i) = M_k$. With respect to the maze task these two relations can be interpreted as the presence of connections between tokens of screen displays, in this case the physical and temporal connections between tokens of a single screen, or on different displays, and the presence of indicating relations understood here as translations between particular Path and Maze description types.

A channel, $\mathcal{C}$, is *sound* or *reliable* where if $p_j \models_P P_j$, $p_j \rightsquigarrow_{s_i} m_k$ and $P_j \Rightarrow_{T_i} M_k$ then $m_k \models_M M_k$.

Thus, if the assumption that Path and Matrix classifications are translatable captures a reliable regularity this can be interpreted as the claim that; if a) a particular location is classified by a description "two up, two along" b) the connections between hardware and software project the same display onto another screen and c) descriptions of the type "two up, two along" translate as Matrix descriptions of the type "two, three", then the display on the second screen is correctly classified by the description "two, three".

This apparatus may appear somewhat complex but this complexity provides for characterisations of a number of ways in which errors may arise. The example Matrix-Path inter-translation, $\mathcal{C}$, consists of the infomorphisms, with a common

target, between all Path and Matrix type descriptions. However, there could be a situation in which two screens are not, in fact, connected in the appropriate way. Faults in the hardware or software that link two displays might distort them relative to each other so that, although there clearly are still connections between the screens, they do not support all the regularities captured by $\mathcal{C}$. Under these circumstances, a situation may arise in which the fact that one screen is appropriately classified by "two up, two along" and this type of Path description translates as a Matrix description type "two, three" does not guarantee that the other screen can be appropriately classified as "two, three". Here, the fact that $p_j \models_P P_j$, the existence of a connection $p_j \rightsquigarrow_{s_i} m_k$, and the indicating relation $P_j \Rightarrow_{T_i} M_k$ does not ensure that $m_k \models_M M_k$. Barwise and Seligman term this kind of error an *exception*, s_i is an exception to the indicating relation T_i in the channel $\mathcal{C}$, more compactly $s_i \not\models T_i$. Although the faulty connection between the two screens may be an exception to the indicating relation $P_j \Rightarrow_{T_i} M_k$ it may still support other indicating relations in the channel. For example, if the faulty connection has resulted in the deletion of the second box up in the first column then if $m_1 \models_M$ "two, three" and T_1 is the indicating relation; "two, three" $\Rightarrow_{T_1}$ "two up, two along" and T_2 is the indicating relation; "two, three" $\Rightarrow_{T_2}$ "three along, one up" then the connection $m_1 \rightsquigarrow_{s_1} p_1$ is an exception to the indicating relation T_1 but not to T_2.

Another source of error considered by Barwise and Seligman concerns cases in which there is no connection $p_j \rightsquigarrow_{s_i} m_k$ in the channel formed from two classifications. In this situation it may be true that $p_j \models P_j$ and that $P_j \Rightarrow_{T_i} M_k$ but the lack of an appropriate connection means that no information can flow from one classification to the other. As a result of a procedural error, the players may be looking at screens connected to different machines, not each other. Thus the fact that one screen is appropriately classified as "two up, two along" has no consequences for whether the other screen is appropriately classified as "two, three",

although, accidentally, it may indeed be appropriate. Barwise and Seligman term this sort of error a *psuedosignal*.

This sketch of the basic structures in channel theory illustrates how the appeal to different channels and different forms of error provides a framework in which both the reliability and fallibility of regularities can be characterised. Exceptions to a channel do not entail that no information can flow, rather, they entail that some inferences, characterised by the indicating relations, will fail while others succeed. Channel theory provides an elaborate framework for modelling various possible relations on classifications, and the foregoing provides only a sketch of this apparatus (see e.g., Barwise & Seligman, 1994). However, this outline provides a sufficient foundation for the purposes of characterising communication in the maze task.

5.3 Modelling the Maze Task

The example channel, $\mathcal{C}$, which characterised some of the informational dependencies suggested by an inter-translation between Path and Matrix descriptions, was derived from the perspective taken in the discussion on the maze experiments of Chapter 4. Obviously, this is not a perspective available to the participants in those studies. They did not know what their partners were looking at any particular point, nor could they be sure of how their own descriptions were being interpreted. The channel $\mathcal{C}$ represents a theorist's characterisation of information flow between description types; what is required for a model of the maze task is a characterisation of the patterns of information flow determined by each individual's interpretation of various description types with respect to their own understanding of the maze. To the extent that this is possible, communication can then be analysed in terms of the information flow between individuals.

The goal of analysing the information flow in communication requires the elab-

oration of a number of steps. The following section proceeds toward a characterisation of individuals playing the maze task (henceforth agents) in three stages. Firstly, by appeal to schemes of individuation made up of primitive classifications. Secondly, by developing these primitive classifications into a model of conceptualisations through the elaboration of informational structure on schemes of individuation. Lastly by combination of conceptualisations in order to derive a characterisation of idiolects. These structures form the basic model of an agent. Section 5.3.2 considers the consequences for information flow between agents in terms of connections between them which can then be elaborated into a notion of a language.[2]

5.3.1 Agents

An important condition to be met by this model is that it should allow an arbitrary degree of ontological variation to obtain between agents. In the framework of channel theory two restrictions are required to meet this criterion; firstly, the types and tokens of a classification must be relativised to agents. Secondly, the tokens, though still understood as tokens of things in the world, are individuated solely by reference to the types in the relevant agent's classification.

Scheme of Individuation

A *scheme of individuation* (cf. Barwise & Perry, 1983) characterises the basic ontology recognised by each agent. For modelling the maze task, two schemes of individuation are important. The first scheme consists of a set, Σ, of primitive location classifications $\{L1, L2, L3 \ldots Ln\}$ each of which consists of a token part of a maze display, l_i, and the type, or *concept*, L_i, which individuates that

[2]The spirit of this approach, although not its exact form, owe a great debt to an ongoing collaboration with Carl Vogel (see e.g., Healey & Vogel, 1994) and should be understood as the product of joint work.

point i.e., $l_i \models L_i$. Different schemes of individuation are assumed to carve up the maze with unequal degrees of acuity, picking out different chunks for any instance of the maze. For example, the concepts involved in a simple Figural scheme of individuation might pick out large portions of the display, classifying areas as 'jutting portion' or 'large square',[3] discriminating only relatively gross characteristics of the display. By contrast, the scheme of individuation associated with Matrix descriptions would individuate locations more finely. The second scheme of individuation important to the development of the model below, is that associated with the classification of utterances. This is also modelled as a set of primitive classifications, $\{U1, U2, U3, \ldots Un\}$, consisting of an utterance token, u_i, and its individuating type U_i i.e., $u_i \models U_i$. More will be said about this below. However, again, the scheme of individuation is understood as determining the ontology of utterances, different schemes carving different boundaries. Some motivation for this derives from the familiar examples of perceptual confusions between different languages (see e.g., Clark & Clark, 1977), and the problems of even segmenting utterances in a language with which we are not conversant. Phonological distinctions vary from language to language and so do peoples' abilities to discriminate between them.

Conceptualisation

Given a basic scheme of individuation, we can develop a structure corresponding to the notion of mental model used in Garrod and Anderson (1987). The idea is to capture the informational relations between locations implied by mental models of the maze in virtue of their organisation into, for example, ordered lines of elements or interlocking paths.

A *conceptualisation* is a classification formed by a channel, $\mathcal{C}$, from a scheme

[3]The phrases 'jutting portion' and 'large square' are intended as labels for the corresponding non-linguistic concepts.

of individuation and consists of an indexed family of infomorphisms between locations; $\{f_i : L_i \rightleftarrows C\}_{i\in\Sigma}$ that share a common target together with relations defined on the indicating relations.

The classification that forms a conceptualisation consists of a set of signalling relations, between points in various instances of the maze $l_j \rightsquigarrow_{s_i} l_k$ and a set of indicating relations between location types $L_j \Rightarrow_{T_i} L_k$. Modelled this way, a conceptualisation amounts to a collection of location classifications and the different frames of reference associated with different mental models are captured in a conceptualisation through the structures encoded by the types in the channel and relations on them. To illustrate this we can consider a display that consists of a 3 by 3 grid, , as in figure 5.3, classified, according to a scheme of individuation, as nine possible locations; $\Sigma : \{L1, L2, L3, \ldots L9\}$ consisting of a token of part of the display l_i classified by an individuating type L_i.

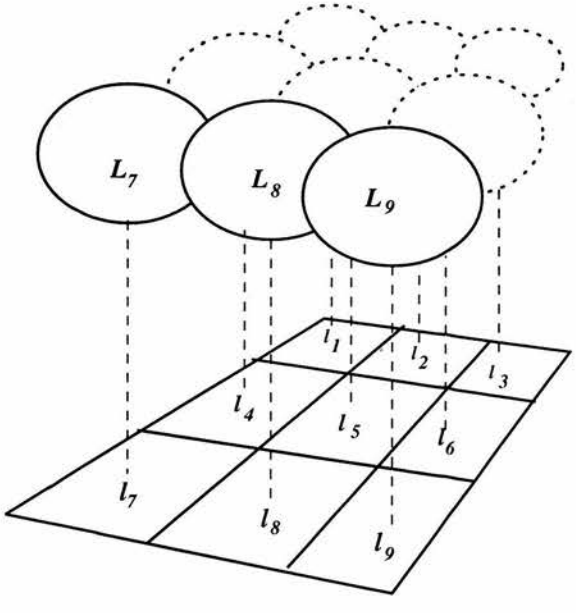

Figure 5.3: Scheme of Individuation for a Simple Maze Display

A mental model that imposes a Line based frame of reference for this display can be characterised as a conceptualisation *Line* consisting of the indexed homomorphisms $f_i : L_i \overrightarrow{\leftarrow} L$. The classification *Line* will include various types or indicating relations that characterise the basic structures for a Line based model with information flow determined by the connections classified by the indicating relations in particular instances. The grouping of locations into lines can be captured in the channel in terms of the component indicating relations in the conceptualisation *Line*. For the display in figure 5.3 the top row might be modelled as the indicating relation R_1 formed from the infomorphisms; $f_1 : L1 \overrightarrow{\leftarrow} R1$, $f_2 : L2 \overrightarrow{\leftarrow} R1$, $f_3 : L3 \overrightarrow{\leftarrow} R1$. Thus the indicating relation $\Rightarrow_{R_1}$ holds between the location types; L_1, L_2 and L_3 and classifies the various connections $\rightsquigarrow_{r_i}$, between l_1, l_2 and l_3. Similarly, the positions in a row can be modelled by other indicating relations in *Line*; $\Rightarrow_{P1}, \Rightarrow_{P2}$ and $\Rightarrow_{P3}$ with, for example, $\Rightarrow_{P1}$ holding between the types; L_1, L_4 and L_6 and classifying connections $\rightsquigarrow_{p_i}$ between l_1, l_4 and l_6.

Given the basic organisation of locations into rows and positions a vertical ordering of rows can be captured by a relation, $>$, on *Line* such that for all pairs of rows, R_i is above R_j if $\langle R_i, R_j \rangle \in >$. Similarly the horizontal ordering of positions could be captured by another relation, $\gg$, such that for all pairs of positions P_i is before P_j if $\langle P_i, P_j \rangle \in \gg$.

Similar considerations apply to conceptualisations formed on the scheme of individuation for utterances. This channel is considerably more complex as it must characterise an individual's conceptualisation of the relations obtaining between utterance types and the possible structures for this channel will clearly be complex, however, they are not directly relevant to the concerns of this analysis. It is assumed that they capture the relations, as understood by a particular individual, obtaining between utterance types, including segments of utterances, grouped according to similarity of form. For example, via types that capture notions like 'quantifier' or 'noun' or higher order structures corresponding to 'noun-phrase'

etc. Roughly, the syntax recognised by a given individual.

Clearly, the notion of conceptualisation admits a wide range of possible structures, each corresponding to alternative mental models, or frames of reference, on the locations picked out by the scheme of individuation. The question of which structure is appropriate, or what is the best way to characterise it, in any given instance is taken to be essentially an empirical one. However, we can also consider two formal properties of conceptualisations that can be used in comparing them.

A conceptualisation, $\mathcal{C}$, is *consistent* if it has no exceptions. This is the soundness condition; for each connection $s_i \in \mathcal{C}$ and each indicating relation $T_i \in \mathcal{C}$ if $L_j \Rightarrow_{T_i} L_k$ then for all $l_j \models L_j$ and all l_l such that $l_j \rightsquigarrow_{s_i} l_l$ then $l_l \models L_k$.

In the context of the conceptualisation *Line* considered above, this amounts to the condition that, for example, all the location types or concepts in the scheme of individuation are mutually consistent. This is because the conceptualisation is modelled in a way that entails connections such as $l_1 \rightsquigarrow_{p_1} l_4$, that fall outside the domain of an indicating relation such as $\Rightarrow_{R_1}$, have sites of a type appropriate to the indicating relation, i.e. $l_1 \models L_1$, and $L_1 \Rightarrow_{R_1} L_2$, requiring that $l_4 \models L_2$ if the conceptualisation is consistent. Less formally, the individuating types or concepts that contribute to a consistent conceptualisation, *Line*, should all correspond to something like a notion of 'box'.

The stringency of this condition is a function of how systematic the framework of a particular mental model is. For example, in a simple Figural conceptualisation, consistency is a weaker condition since, intuitively, they encode fewer relations between locations. Where a Line conceptualisation encodes relations between all locations in a display a Figural conceptualisation may only pick out the relations between one or two locations with respect to some coarse individuating type that picks out a feature such as the 'left leg' or 'large space'.

A conceptualisation is *lucid* if whenever $l_j \models L_j$ and $l_j \rightsquigarrow_{s_i} l_k$ and $l_k \models L_k$ then $L_j \Rightarrow_{T_i} L_k$.

A lucid conceptualisation of the maze displays is one in which every connection between individuated points in the maze is classified by the appropriate indicating relation between their individuating types. It amounts to the condition that all the connections, as given by the infomorphisms that characterise the conceptualisation, between token locations in a display are classified by some indicating relation i.e., the possible informational relations with respect to a conceptualisation are exhaustively mapped. This condition is met by the example conceptualisation *Line* discussed above since every connection is classified either by an indicating relation characterising position, $\Rightarrow_{P_i}$ or an indicating relation characterising a row, $\Rightarrow_{R_i}$. In a figural conceptualisation this condition will often not be met since it offers only a partial characterisation of the interrelations between locations. Although one, target, location will be determined with respect to some feature, or element of the configuration, in a display the *ad hoc* nature of this sort of model suggests that many other connections are not classified.

There is a tradeoff apparent between consistency and lucidity since the more connections that are classified under a conceptualisation the tighter the restriction on compatibility between the indicated types.

The properties of lucidity and consistency can be utilised to generate a partial order, $\prec_C$, on conceptualisations that characterises their relative internal-coherence. Firstly, they are ordered according to their lucidity, with the most exhaustive classification of connections corresponding to the most lucid. Where two conceptualisations are equally lucid they can be ordered according to their consistency. The most lucid, consistent classification is the most coherent, providing the most information about the inter-relations between locations and their individuating types. There is no restriction on agents to adopt only consistent, lucid classifications; a condition unlikely to be met on the grounds of performance limitations alone. However, agents are assumed to prefer more coherent to less coherent conceptualisations as characterised by $\prec_C$.

This comparison of conceptualisations according to their coherence does not depend on the relative acuity of the ontology or scheme of individuation they employ. That is, a relatively gross Figural conceptualisation can be as coherent, in this sense, as a Matrix conceptualisation. This is important because it avoids a commitment to any absolute preference for one ontology over another, a commitment that would be inconsistent with the arguments of Chapter 2. Such a preference could only be determined by reference to the uses to which a conceptualisation is put. With respect to some goal, an ontology may prove inadequate if it fails to make a discrimination appropriate to the task at hand. An absolute judgement of the adequacy of an ontology presupposes a definitive specification of the uses to which it will be put.

Idiolect

The final step in the characterisation of agents is to model the informational contingencies in the interpretation of utterances about the maze.

An *idiolect*, $\mathcal{I}$, for the maze task is a channel formed by an indexed set of infomorphisms, $\{f_i : T_i \overrightarrow{\leftarrow} I\}_{i \in \mathcal{C}, \mathcal{U}}$ between the conceptualisation of $\mathcal{C}$ locations employed by an agent and the conceptualisation, $\mathcal{U}$, of utterances.

An idiolect is modelled as indicating relations, $\Rightarrow_{I_i}$, between utterance forms and elements of the conceptualisation that classify connections $\rightsquigarrow_{i_i}$ between tokens of utterance forms and tokens of conceptualisation types. In the context of the maze task, the signalling relations in an idiolect can be thought of as the physical/temporal links between instances of displays and instances of utterances perceived by an agent. An idiolect characterises the interpretation function embodied by an agent.

An *interpretation* is a signalling relation, i_i, and an indicating relation, I_i in $\mathcal{I}$, such that $i_i \models I_i$. An idiolect characterises the semantics, for an agent, of expressions relating to the maze. Each type in the idiolect, such as $\Rightarrow_{I_1}$ provides

a relation pairing expression types such as "row" with types of structure in the maze conceptualisation such as R_i. This formulation still admits a number of possible idiolects even where the conceptualisations of maze and utterance forms are constant. Thus, the indicating relation between "first row" and lines of boxes in the display could correspond to an interpretation in one idiolect in which it picks out the bottom line of boxes and one in which it picks out the top line. By contrast, if "first row" is interpreted in one idiolect as a horizontal element and in another as a vertical element this difference could be modelled as due to divergent conceptualisations.

As for conceptualisations, a wide range of possible structures could be defined in any particular case. The question of which of these is appropriate is also understood to be a matter of fitting whatever empirical data is available. However, again, properties of consistency and lucidity can be formulated to provide a partial ordering.

An Idiolect, $\mathcal{I}$, is consistent if it admits no exceptions, i.e. for all $\rightsquigarrow_{i_i} \in \mathcal{I}$ if $c_j \rightsquigarrow_{i_i} c_k$ and $C_j \Rightarrow C_l$ and $c_j \models C_j$ then $c_k \models C_l$. For example, using the *Line* conceptualisation discussed above translates as the condition that a token of the utterance form "row" is connected to elements in the example grid each of which is classified by mutually compatible type. This condition would be violated where, for example, there was some equivocation in the interpretation of "row" as corresponding to horizontal or vertical elements.

An idiolect, $\mathcal{I}$, is lucid if for each connection $\rightsquigarrow_{i_i} \in \mathcal{I}$ such that $c_j \rightsquigarrow_{i_i} c_k$ there is an indicating relation $\Rightarrow_{I_i} \in \mathcal{I}$, between the types of c_j and c_k such that $i_i \models I_i$. This can be understood as the condition that there are no connections between utterance forms and elements of the maze that cannot be interpreted.

These two properties can be used to generate a partial ordering, $\prec_I$, on idiolects according to their coherence. Agents are assumed to prefer more coherent idiolects but are not restricted to completely consistent or lucid ones. The principle value

of this approach is that is allows the semantics of an idiolect, for both production and comprehension, to be characterised in terms of information flow.

5.3.2 Communication

The preceding section developed a model of agents that employed three levels of analysis; schemes of individuation, conceptualisations and idiolects. This was done in a manner that maintained a substantial degree of independence between each level, allowing that even where schemes of individuation are held constant, conceptualisations may vary and, similarly, even where conceptualisations are held constant, idiolects may vary. As a result, different agents can vary in the ontology they recognise for a domain, the mental model they use to organise it and the interpretational scheme they apply in relating utterances to their mental model. Given these possible asymmetries between agents, the next step is to characterise how communication can occur in the face of these disparities.

Communication

Two agents A and B are communicating iff there is an utterance token $u_j^A \models U_j^A \in A$ and an utterance token $u_k^B \models U_k^B \in B$ and there is a connection $u_j^A \rightsquigarrow_c u_k^B$.

This is communication in the literal sense requiring only a connecting route, c, between utterances classified by both parties. In the electronic version of the maze task this connection is the audiolink between players. In the paper version it is the transmission of sound waves through the air. Agents themselves cannot determine whether they are connected in the right way, they might be subject to hallucination or background noise might infect their classification of what they hear. In the former case there is, by our lights, no communication at all and in the latter there is communication but it is distorted. The connection between agents is a precondition imposed for communication to actually occur but this does not

imply that agents cannot be mistaken about with whom or exactly what they are communicating.

A connection c is a *clear signal* between agents A and B iff, for both agents, there is one $i_i \models I_i \in \mathcal{I}$ such that $u_j \rightsquigarrow_{i_i} c_k$, $U_i \Rightarrow_{I_i} C_k$ and, in their conceptualisation of the maze, $c_k \models C_k$.

This characterisation of successful communication requires only that agents arrive at unique interpretations for an utterance on their current understanding, not that this is, by our lights, the same interpretation. The definition is consistent with circumstances under which the interpretation arrived at in each agent's idiolect is quite different, for example A interprets "rows" as corresponding to horizontal elements whereas B interprets them as vertical elements. In this sense communication is a special case of misunderstanding, agents' interpretations may still diverge to an arbitrary degree, but can communicate successfully as long as this divergence is mutually indiscriminable.

To deal with unsuccessful communication and its likely effects on interpretation we need to analyse cases where agents are connected in the appropriate way but one or both does not have a clear signal, i.e., do not associate a unique interpretation with an utterance. The consequences for information flow in these cases can be characterised by two alternatives.

An agent A has a *multisignal* where $u_j \models U_j$ but there is more than one interpretation $i_i \models I_i \in \mathcal{I}$ for which $u_j \rightsquigarrow_{i_i} c_k$ and $U_j \Rightarrow_{I_i} C_K$.

A multi-signal arises where some utterance token u_j is classified in the utterance conceptualisation but there is more than one element of the maze conceptualisation to which it could correspond. For example the utterance type "Leg" could plausibly indicate several parts of a display under a Figural conceptualisation.

An agent A has a *pseudosignal* where $u_j \models U_j$ and $U_j \Rightarrow_{I_i} C_k$ but there is no $c_k \models C_k$.

A psuedosignal would arise where, for example, an utterance is classified as

having a type "two up, one along" but, under the current conceptualisation there is a blank space at the position indicated by that description type.

The foregoing definitions of clearsignal, multisignal and pseudosignal characterised the information flow corresponding to interpretation in an idiolect. They can equally well be used to characterise the information corresponding to production. A clearsignal would capture the default case in which a particular part of the maze relates to a particular utterance. A multisignal would give multiple realisations for the same location. A pseudosignal would correspond to an 'ineffable' aspect of the configuration, i.e., one which has no realisation as an utterance. Intuitively, the default case in production would be a clear signal. The other two possibilities amounting to uncooperative or even pathological behaviour.

So far, this characterisation of communication provides an indication of how communication, in terms of information flow, can be modelled despite marked idiolectical disparities between the parties to a dialogue. It requires only that there is some physical connection, even a very indirect one, between interlocutors. This addresses, in at least a minimal way, the desire to develop a model of communication that does not depend on a shared code. It also indicates how the semantics for an idiolect could be constructed. A second aim of this Chapter is to give some analysis of the consequences of unsuccessful communication for semantic coordination. A full account of this process, dealing with cycles of repair, and more nebulous factors such as whether the agents are cooperative, would require a dynamic model beyond the scope of the model outlined here. However, some simple suggestions can be made on the basis of the characterisations of signal types and their consequences for the different levels of analysis: scheme of individuation, conceptualisation and idiolect.

The expectation is that, all things being equal, the different kinds of signal should make certain responses more likely than others. Given a clear signal, an agent would be expected, possibly after a confirmation, to move on to production

of the next utterance. Given a multisignal an agent would be expected to seek clarification e.g., "counting the first box as one?" or "which side?". Consideration of psuedosignals suggests more interesting possibilities. In situations where an agent has no interpretation available which provides a link between an utterance and a part of the display several options are open. Firstly, a refusal to accept the utterance, e.g., "I don't know what you mean" or "Sorry, no". Secondly, an adjustment in idiolect that results in a unique interpretation or clear signal. Thirdly an adjustment of the current conceptualisation of the display in order to find an appropriate location, in turn, altering the idiolect to arrive at a clear signal. Lastly an agent may revise their scheme of individuation, causing adjustments in both the conceptualisation and idiolect in order to arrive at a clear signal.

The repair based model of coordination in Chapter 4 appealed to some general constraints on the likely responses to difficulties in interpretation. The most general of these was that explicit negotiation is dispreferred on the grounds that it is susceptible to infinite regress. This translates to a suggestion that rejection of an utterance is the least preferred response to a psuedosignal. Appeal was also made to the claim that individuals would generally prefer to engage in minimal repair until the difficulty was resolved. The question of what constitutes minimal repair was given no further explanation. One approach to this is that it corresponds to minimal collaborative effort in the sense of Clark et. al. where each individual tries to limit the joint effort required to overcome whatever difficulty has arisen. Appeal to the structures proposed above admits an additional, individualistic criterion relating to the extent of the revisions an agent must engage in in order to arrive at a unique interpretation for an utterance. If we assume a general preference for adjustments that occasion the least revision for an agent then the model predicts that adjustments in interpretation will be preferred to adjustments in conceptualisation which are in turn preferred to adjustments in the scheme of individuation. Changing the scheme of individuation has consequences for both the

conceptualisation and interpretation, changes to the conceptualisation alter the idiolect but leave the scheme of individuation intact. Changes in interpretation cause the least disturbance. Although this is speculative, it does illustrate a way in which the model can be used to analyse semantic coordination.

The model of an agent characterises the semantics of idiolects in terms of information flow determined by the existence of connections between instances of things classified by the agent. This goes some way toward accommodating Putnam's argument that the meaning of a term, at least its reference, is partially determined by the actual nature of the stuff referred to. The appeal to classification of tokens 'in the world' has the consequence that the individuating types or concepts, and the indicating relations between them, depend for their reliability on the 'ultimate' nature of whatever is picked out. There is no suggestion that these individuating concepts determine the extension of the things they classify, only that the information flow and, thereby, semantics of the indicating relations between types depends on the constitution of the connections they classify. This element of indexicality feeds into the model of semantics for an idiolect. However, Putnam's argument also requires a characterisation of meanings in a language not just an idiolect. This characterisation is also necessary to deal with Burge's claims concerning the role of a linguistic community in fixing the meaning of terms in the language.

Translation

A *translation*, $\mathcal{T}$, is modelled as the channel formed by an indexed set of infomorphisms between interpretations in a set Σ of idiolects $\{f_i : I_i \rightleftarrows T\}_{i \in \Sigma}$. The types of this classification are indicating relations, $\Rightarrow_{T_i}$ between the types in each component idiolect. The connections $\rightsquigarrow_{t_i}$ are between the tokens of each idiolect.

This channel brings the discussion back to the characterisation of the theorist's perspective, outlined in section 5.2.1, which determined various equivalences be-

tween Matrix, Line, Path and Figural description types. The indicating relations that determine translations between these types are characterised by the types of $\mathcal{T}$. For example, a translation between Line and Matrix descriptions is a set of indicating relations, formed from the infomorphisms, between interpretations in Line based idiolects and interpretations in Matrix based idiolects.

Naturally, no theorist actually has veridical access to the structure of the component idiolects they aim to translate. Rather, this is inferred, as it was in the discussion of the maze task, from e.g., patterns of co-occurrence of particular expressions relating to the maze and comprehension as indexed by patterns in the marking of locations. A translation between idiolects uses the theorist's idiolect to determine the interpretations employed by each individual and the interrelations between them. Depending on how the component infomorphisms are defined, a number of translations are possible between idiolects.

The theorist's translation resolves the ambiguities and conflicts between idiolects by determining relations that provide mappings between all the different interpretations in the component idiolects, for example by treating "row" as indicating horizontally oriented elements in one and vertically oriented in another. This allows that a single interpretation in one idiolect may map to several in another.

Language

A characterisation of a language in some community involves a restriction of the possible mappings between interpretations to those that determine unique mappings between interpretations in each idiolect utilised in interaction.

A *language*, ℓ, is a translation between a set Σ of idiolects $\{\mathcal{I}^i, \mathcal{I}^j, \mathcal{I}^k \ldots\}$ such that for each type $S_i \in \ell$ and each $\mathcal{I} \in \Sigma$, there is only one type $I^i_j \in \mathcal{I}^i$ and only one $I^k_l \in \mathcal{I}^k$ such that $I^i_j \Rightarrow_{S_i} I^k_l$.

This condition imports the idea of successful communication as a clear signal

into the notion of a language. Interpretations may still vary between idiolects, but each indicating relation in the language relates only a single interpretation type from any idiolect. The types of a language represent the content of each expression and the tokens, as connections between idiolects naturalise this content in terms of the information flow supported in specific interactions. Importantly, the variation admitted between idiolects means that this content is not necessarily reducible to the associated interpretation type in any particular idiolect and it preserves the intuition that the meaning of many terms of a language is inherently vague. A term like "row" can relate interpretations that treat it as a horizontal, diagonal or vertical set of elements. It also allows that where interpretations converge as a result of repair so the content of a term as specified by the relevant type in the language becomes more specific.

The set Σ of idiolects that contribute to a language determines the linguistic community. This treats the choice of linguistic community as arbitrary depending only on the presence of interconnections between them. This allows for a characterisation of sub-languages that emerge for the different experimental communities as well as more general languages such as English.

5.4 Discussion

It was suggested in Chapter 2 that one possible reason for the persistence of implicit or explicit appeals to some form of shared semantic code in explaining mutual-intelligibility is the intuition that there must be one in order for communication to be possible at all. The model described above provides a possible account of communication that does not appeal to a code. There is no reliance on individuals' associating the same, or even commensurable, interpretations with utterances, only that there should be connections between them for communication to occur. This was reflected in the definition of successful communication

as a clear signal which depends on interpretations being mutually-indiscriminable in the context of their current dialogue. The only condition on success is that interlocutors can continue to coordinate their behaviour in a way appropriate to the constraints placed on them by a particular task.

The model attempts to defuse the tension between cognitivist semantics and broad content by isolating two different domains of semantic analysis: idiolects and languages. The semantics of idiolects are characterised by reference to the connections in the world that support information flow amongst the component classifications in the idiolect. As noted above, this imports a degree of indexicality into the notion of an interpretation that allows the types and indicating relations in an agent's classifications to embody only partial characterisations of the things they classify, and the relations between them. However, the appeal to information flow in virtue of classification of things in the world leaves room for the actual nature of what is classified to constrain the reliability of the indicating relations employed in an idiolect. This provides important leverage in meeting the difficulties raised by Putnam.

In order to characterise the semantics of a language the notion of a translation was introduced with languages defined as a special case. In the model, a language is a set of infomorphisms between idiolects that provides unique translations between interpretations arrived at in interaction. This is intended to provide a level of analysis which corresponds to what Putnam and Burge, and, to an extent, semantical realists, think of as the meaning of expressions in a language. The types of a language ℓ were treated as determining the content of various expressions and classify connections between interlocutors. This provides for a normative characterisation of content that need not be reducible to the interpretations employed by any specific individual or set of individuals. The content of a term like "friend" is modelled as an indicating relation between unique interpretations arrived at by each individual, interpretations which may differ to an arbitrary degree. The

only requirement is that the information flow between individuals is captured. As a result, the semantics of a language in this framework is a generalisation over interactions, and depends on how individuals coordinate their interpretations in use.

Situations of the kind considered by Burge can be analysed by treating the 'deviant' use of "arthritis" as an exception to the appropriate indicating relation between interpretations of arthritis in the patient's speech community. The associated signalling relation supports many other interactions between the patient in expressing other beliefs about arthritis, but is an exception in the case of the expression of beliefs relating to arthritis in his thigh. For this model, the shift to the counterfactual community, in which the 'deviant' interpretation is widely held, alters the derived indicating relation in a way which ensures that, in this case, it does classify the signalling relation.

With respect to the goal of modelling dialogue this model is incomplete in a number of ways. Although it does suggest a way of characterising the semantics of idiolects and languages it does not provide any dynamic mechanisms for modelling how these structures change during the course of a dialogue. Such mechanisms are necessary for an adequate account of how convergence in interpretations evolves over the course of a dialogue and for accounting for the group-based effects observed in the maze experiments. In addition to its limitations in accounting for semantic coordination, the model does not address other important aspects of the successful conduct of communication. For example, 'higher order' concerns such as planning models, intentions, game structures and speech acts (e.g. Airenti et al., 1993; Cohen & Levesque, 1990; Grosz & Sidner, 1990; Kowtko, Isard, & Doherty, 1991). Nonetheless, it does provide a semantic framework on which these structures could, at least in principle, be built. With respect to this aim, the model of communication advanced here could be advantageous. As Cohen and Levesque (1993) have noted, one of the problems in developing adequate models of dialogue

is the assumption that the literal meaning of utterances is automatically recognised by the parties to an interaction. This creates problems in accounting for phenomena, such as backchannel responses, which seem to function primarily as signals that the 'literal' meaning of an utterance has been understood (cf. Clark & Wilkes-Gibbs, 1990; Clark & Schaefer, 1989). The relaxation of this assumption in the current model may provide a way of overcoming these difficulties.

This model also has very little to say about more familiar semantic concerns such as compositionality, anaphora and negation. The issues it is designed to address do not speak directly to these concerns; however it is worth noting that the structures provided by channel theory can be adapted to provide a normal compositional syntax and semantics (e.g., Cooper, 1989, 1991). This could be applied both at the level of idiolects and at the level of languages as defined by this model.

Chapter 6

Diallage

In cognitive science, semantic theory has two broadly separate lineages. One tradition, exemplified by semantical realism, analyses language effectively abstracted away from users, treating it as an independent domain of investigation. The other tradition, exemplified by semantical cognitivism, interposes cognitive structures between language and the world, locating semantic analysis in the investigation of regularities between cognitive states and interpretation. Although it is generally acknowledged that the notion of meaning is bound up with the communicative function of language, both traditions, it has been argued, proceed to the analysis of semantic structures in a manner that overlooks this function. Both traditions assume that communication is possible in virtue of "speaking the same language" or possessing conceptual structures that are "sufficiently similar" but no explanation is given of what this consists of. This assumption implies a shared code model of mutual-intelligibility in which communicational success is apparently underwritten by the existence, at some level, of a common semantic model. In theories that do directly address communication, this assumption is frequently made explicit. It has not been claimed that this is a necessary feature of these models, rather that it has persisted as an unexamined, frequently tacit, assumption in theories

directed toward other concerns. Consequently, in models that deal explicitly with communication, a range of 'pragmatic' factors are countenanced as contributing to interpretational asymmetries between interlocutors but they take effect only after some literal or primitive semantic content of an utterance has been recovered (cf. Rommetveit, 1983).

Taylor (1992) suggests that this situation has arisen as a consequence of an adaption of informal discourse about communication into an argument which takes as a premise the observation that successful communication is a commonplace and reasons from this to the existence of some form of shared linguistic code. In theoretical semantics this code has found expression in realist, structuralist and naturalist forms but each variation draws support from the intuition that language is, after all, mutually-intelligible and that this must depend ultimately on some shared set of meanings. A central theme of this thesis has been that another commonplace observation, namely, that we do not all understand exactly the same thing by an utterance, ultimately undermines an appeal to a shared code. The main body of this thesis has been concerned with establishing an empirical case for the importance of idiolectical variation and then arguing from this to conclusions about the nature of semantic theory. It has been suggested that idiolectical variation is a semantic problem, that it cannot be discounted as 'noise' and that attempts to blunt the threat it poses by appeal to factors that might naturalise a code are inadequate.

One property that is frequently cited in discussions of meaning is that it is a normative notion: what an expression means is not a matter for individual determination. *Pace* Humpty-Dumpty,[1] meanings are not something that individuals are free to legislate as they please, rather as McDowell (1984) puts it, there is a 'contractual obligation' to conform to certain standards by which our use of

[1] " 'When I use a word' said Humpty dumpty in rather a scornful tone 'it means just what I choose it to mean –neither more nor less' " (Carroll, 1962, p.75)

an expression is judged correct or appropriate. For the semantical realist, this is because there is some determinate fact about what a word actually means, independent of our, possibly mistaken, beliefs on the matter. The application of this view in understanding communication is problematic since it is difficult to provide an adequate explanation of how such facts about correct interpretation could ever come to be internalised or perhaps apprehended by an individual. Semantical realists preserve a strong explanation of normativity at the expense of an account of communication.

The alternative view, that meanings are cognitive entities faces the converse problem, it does appear to explain intersubjectivity but faces problems in dealing with normativity. Chapter 2 concentrated on the problems idiolectical variation creates for determining an appropriate semantic ontology. It was noted there that normativity may result from the presence of an interpretive strategy which involves discounting these differences for 'most purposes'. However, there is also another important strand of argument, deriving from Wittgenstein's (1958) remarks on rule-following which argues, in principle, that cognitive states cannot be the sort of things which account for normativity. It is useful to consider how this issue relates to the concerns of this thesis.

Wittgenstein (1958) was concerned, amongst other things, with the problem that we can never determine whether an individual has settled on the correct interpretation of an expression, understood as an internalised rule, since their behaviour is always consistent with indefinitely many possible interpretations. An addressee's utterance of "four" in response to "two plus two" provides no guarantee that the interpretation of "plus" is correct, since it may transpire on more exhaustive examination that they offer 'deviant' responses where the numbers to be added rise above, say, one thousand. In fact, for any finite number of examples there are indefinitely many possible interpretations of "plus" consistent with answers that appear to conform to the accepted meaning of the rule for addition.

This is not just a point about induction; the problem is that nothing we can cite about internalised rules can successfully narrow down the open horizon of possible interpretations that could be consistent with them. Whatever mental state or rule I have, through learning, come to associate with addition I may still reinterpret in ways that are at odds with the normative meaning of "plus". For Kripke (1982), the upshot of this is that no sense can be made of the idea that an individual, considered in isolation, can be understood as following a rule since there is no way of giving substance to the claim that a rule or mental state has been (in)correctly interpreted. Kripke, (but not Wittgenstein, see e.g., Baker & Hacker, 1984; McDowell, 1984), diagnoses this as a problem with the conception of interpretations as correct or incorrect. His solution is to discard the idea that interpretation should be analysed in terms of truth conditions in favour of assertion conditions. What counts as a correct interpretation is thus something determined with respect to the practices of a particular linguistic community. The utterance "two plus two is four" is not determinately true or false, rather, it is merely justified by a community in which most individuals assent to its assertion. Consequently, even granted a set of rules for interpretation, they cannot do the work we hope since the rules must be applied in given instances and nothing can guarantee a correct application. Kripke's anti-realist response to this is to appeal to warranted assertions whose appropriateness is determined as a matter of "brute fact" in the community.

The appeal to community assent brings the communicative aspects of language into focus. It is through use in a community that the normative nature of meaning is cashed out. The Kripkensteinian arguments against interpretive realism are similar to aspects of Garfinkel's (1967) critique of the role of shared knowledge in the Parsonian model of social order. The Parsonian approach discussed in section 3.2.1 offers two main ways of accounting for intersubjectivity. In the basic case, individuals share knowledge in virtue of making convergent discoveries

about the objective world. As long as the knowledge is obtained, in each case, by something approximating scientific methods convergence is assured. The second case, more pertinent to the current concerns, deals with shared knowledge of socio-cultural institutions. Here, the appeal to objective discoveries according to scientific method is of no avail since, for Parsons, there is no appropriate objective structure to underwrite intersubjective convergence. Instead, Parsons appealed to the internalisation of existing institutional/cultural practices or norms that limit potential divergences in individual's perspectives, providing the necessary conditions for intersubjectivity. Communication, in particular, is achieved through the internalisation of norms relating to the institutional use of a system of symbols i.e., meanings established by the prior practices of the community.

Garfinkel (1967) objected to this view of intersubjectivity on the grounds that the notion of "common" or "shared" at work in the Parsonian model did not, in fact, provide an adequate guarantee of intersubjectivity. Garfinkel urged that even if it is granted that a set of norms or symbols for governing interpretation were somehow internalised this fact itself does not actually solve the problem of how an individual determines an appropriate application of those rules.

> "If no rule can 'itself step forward to claim its own instances' but always awaits contingent application 'for another first time' , it necessarily follows that rules *per se* cannot determine the specifics of actual conduct no matter how deeply internalised they are (Heritage, 1984; p124)

Garfinkel's proposal, following Schutz (1973), was that intersubjectivity should not be understood as an 'in principle' problem requiring a philosophical or conceptual solution but rather as a practical problem that individuals deal with on a day to day basis. Rather than trying to account for the possibility of shared knowledge,

> "'Shared' agreement refers to various social methods for accomplishing member's recognition that something was said-according-to-a-rule and not the demonstrable matching of substantive matters. The appropriate image of a common understanding is therefore an operation rather than a common intersection of overlapping sets." (Garfinkel, 1967; p.30)

The rejection of realist accounts of intersubjectivity and the difficulties with individualistically understood norms or rules leads both anti-realists and ethnomethodologists to a similar conclusion about the normative nature of meaning. Like Kripke, ethnomethodology and, latterly, conversation analysis rejects the idea that successful communication consists in some matching of internal states, turning instead to the public practices of the community as a means of grounding intersubjectivity and normativity. What marks the ethnomethodological approach off from Kripke's anti-realism is that it treats intersubjective understanding as a local, contingent, matter achieved for the first time in each interaction.

There is, however, a difficulty with this sort of appeal to communal practices in that the kind of normativity it licenses seems to be too weak. Communities, under the anti-realist conceptualisation look like arbitrary aggregations of individuals with a propensity to make certain noises under certain circumstances. This, seems to lose sight of the intuition that language somehow gains traction, not just on individuals but also the world. McDowell (1984) suggests that the notion of justified and unjustified assertions, is a "thin surrogate" of what is required by the intuitive notion of objectivity. It appears as though the problem has been moved rather than solved since we are still left with a residual question about what the standards of correctness are for a community. Without some way of linking the patterns of communal assent with the world it seems any arbitrary pattern of assent will suffice for normativity. In the case of conversation analysis we are offered a theory of the 'procedural infrastructure' of interaction but again,

no obvious way of interleaving this with the intuition that language relates to the world in some non-arbitrary way. (Heritage, 1984) regards this as part of a general "ethnomethodological indifference" toward the objectivity of language and argues that in ethnomethodological accounts the real world is temporarily 'bracketed' in order to pursue other questions. In recent approaches to the sociology of science and discourse analysis in social psychology the notion of an objective world has dropped out completely, being treated as a purely discursive construction (see e.g., Leudar, 1991; Button & Sharrock, 1994).

The in-principle problem of accounting for the objective aspects of the normativity of meaning is thus intimately bound up with the notion of communication. It seems we can maintain either the intuitive notion of objectivity or the intuitive notion of intersubjectivity but not both. Like informal discourse about meaning, as captured by the conduit metaphor, theoretical discourse seems to be driven by a basic dilemma. This thesis has attempted to resolve this dilemma by appealing to both objective and communal features of linguistic practice. The result has been a claim that there are two domains of analysis appropriate to an account of meaning; idiolects and languages, and that it is in the interaction between these two domains that the requisite properties emerge. In some respects this is similar to the ethnomethodological account, but there are some important discontinuities. Taken literally, Heritage's suggestion that the interpretation of normative rules is "achieved for the first time" in each interaction leaves no account of how different communities can ever converge toward particular interpretations, in turn, implying that social order is actually illusory. There seems to be no way of accounting for the fact that communities converge, in an apparently regular way, on certain patterns of making sense rather than others. It is not at all clear how the effects of interference between 'sub-communities' reported in chapter 4 can be accounted for on this picture. On the ethnomethodological view, intersubjectivity is held to rely on a symmetry of methods or procedures for making sense (see e.g., Heritage,

1984; p.179 and p.153) and in this respect appears similar to the appeal made here to processes of repair. However, there must also be an account of how the methods and procedures change to explain how individuals actually arrive at the situation of symmetry. As noted earlier, in CA the specification of procedures for the maintenance of intersubjectivity proceeds at a level for which cognitive or individualistic interpretations are explicitly disavowed. The suggestion here has been that what individuals are repairing is precisely something cognitive, they are trying to arrive at some situation in which their inferences about the links between utterances and states of affairs provide for more reliable coordination of behaviour with their interlocutors.

Although this account appeals directly to the cognitive states of individuals in characterising the competences that underpin interaction it does not rely on these alone in accounting for semantic content. Intentional content, as it is normally understood, has been treated as an emergent property of interaction, in particular, of the need to coordinate action through language. The fact that we consider utterances to be about things has been treated as a consequence of the ways in which we can use utterances, in concert with others, in order to achieve things. Effectively, aspects of both the realist and anti-realist accounts of normativity have been combined in an attempt to provide a more satisfactory model of the way meaning forms a link between expressions and states of affairs. This has been done in a way that does not treat this link as a code but nonetheless provides for the possibility of mutual-intelligibility.

Bibliography

Airenti, G., Bara Bruno, G., & Colombetti, M. (1993). Conversation and Behaviour Games in the Pragmatics of Dialogue. *Cognitive Science, 17*, 197–256.

Anderson, A., Brown, G., Shillcock, R., & Yule, G. (1984). *Teaching Talk: Strategies for Production and Assessment.* Cambridge: CUP.

Anderson, A. & Garrod, S. (1987). The dynamics of referential meaning in spontaneous dialogue. In Reilley, R. G. (Ed.), *Communication Failure in Dialogue and Discourse*, pp. 161–183. Amsterdam: Elsevier.

Anderson, A., Garrod, S., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in Narrative Text. *Quarterly Journal of Experimental Psychology, 35A*, 427–440.

Baker, G. & Hacker, P. (1984). Critical study: On misunderstanding Wittgenstein: Kripke's private language argument. *Synthese, 58*, 407–450.

Bar-Hillel, Y. & Carnap, R. (1953). Semantic information. *British Journal for the Philosophy of Science, 4*, 147–157.

Barr, A. & Davidson, J. (1981). Knowledge Representation. In Barr, A. & Feigenbaum, E. (Eds.), *The Handbook of Artificial Intelligence*, Vol. 1, pp. 141–222. California: William Kauffman.

Barwise, J. (1995). Reasoning about Complex Systems. Lecture notes *5th International Colloquium in Cognitive Science* San Sebastian.

Barwise, J. & Perry, J. (1983). *Situations and Attitudes.* Cambridge, MA: MIT Press.

Barwise, J. & Seligman, J. (1993). Imperfect Information Flow. In *Proceedings of the 8th Annual IEEE Symposuium on Logic in Computer Science.* IEEE Computer Society, Washington, DC.

Barwise, J. & Seligman, J. (1994). The Rights and Wrongs of Natural Regularity. In Tomberlain, J. (Ed.), *Philosophical Perspectives, Volume 8*, pp. 331–365. California: Ridgeview.

Bell, R. T. (1991). *Translation and Translating: Theory and Practice.* London: Longman.

Bower, G., Black, J., & Turner, T. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.

Burge, T. (1979). Individualism and the Mental. In Frence, P. A., Vehling, T. E., & Wettstein, H. K. (Eds.), *Midwest Studies in Philosophy, Vol.IV: Studies in Metaphysics.* Minneapolis: University of Minnesota Press.

Burge, T. (1986). Individualism and Psychology. *The Philosophical Review, XCV*, 3–45.

Button, G. & Sharrock, W. (1994). A Disagreement over Agreement and Consensus in Constructionist Sociology. *Journal for the Theory of Social Behaviour, 23*(1), 1–25.

Carey, S. (1988). Conceptual Differences between Children and Adults. *Mind and Language, 3*(3), 167–181.

Carroll, L. (1962). *Through the Looking-Glass.* London: The Folio Society.

Cherry, C. (1966). *On Human Communication: a review, a survey, and a criticism* (2nd edition). Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature Origin and Use.* New York: Praeger.

Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy, 78*, 67–90.

Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A., Webber, B. L., & Sag, I. (Eds.), *Elements of Discourse Understanding*, pp. 10–63. Cambridge: Cambridge University Press.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as A Collaborative Process. *Cognition, 22*, 1–39.

Clark, H. H. & Wilkes-Gibbs, D. (1990). Referring as a Collaborative Process. In Cohen, P. R., Morgan, J., & Pollack, M. E. (Eds.), *Intentions In Communication*, pp. 463–494. Cambridge, Mass: MIT Press.

Clark, H. H. (1979). Responding to Indirect Speech Acts. *Cognitive Psychology, 11*, 430–477.

Clark, H. H. (1985). Language use and language users. In Lindzey, G. & E., A. (Eds.), *Handbook of Social Psychology* (3rd edition)., pp. 179–231. New York: Harper and Row.

Clark, H. H. (1993). Language Use as a Joint Activity. Lecture notes *5th JCI Summer School in Cognitive Science, Edinburgh University.*

Clark, H. H. & Clark, E. V. (1977). *Psychology and Language: An introduction to Psycholinguistics.* San Diego: Harcourt, Brace, Jovanovich.

Clark, H. H. & Haviland, S. (1977). Comprehension and the given-new contract. In Freedle, R. O. (Ed.), *Discourse Production and Comprehension*, Vol. 1, pp. 1–40. Norwood N.J.: Ablex.

Clark, H. H. & Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science, 13*, 259–294.

Cohen, P. R. & Levesque, H. (1990). Persistence, Intention, and Commitment. In Cohen, P., Morgan, J., & Pollack, M. (Eds.), *Intentions in Communication*, pp. 33–70. Cambridge, Mass: MIT Press.

Cohen, P. R. & Levesque, H. (1993). Preliminaries to a Collaborative Model of Dialogue. In *Proceedings of International Symposium on Spoken Dialogue: New Directions in Human and Man-Machine Communication.* Waseda University, Tokyo. November 10–12.

Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour, 8*, 240–248.

Cooper, R. (1989). Information and Grammar. Tech. rep. DAI Research Paper No. 436, Department of Artificial Intelligence, University of Edinburgh.

Cooper, R. (1991). Three Lectures on Situation Theoretic Grammar. In Filgueiras, M. L., Damas, N. M., & Tomás, A. P. (Eds.), *Natural Language Processing, EAIA '90, 2nd Advanced School in Artificial Intelligence.*, pp. 102–140. Lecture Notes in Artificial Intelligence, 476. Berlin: Springer Verlag.

Davidson, D. (1977). Reality without reference. *Dialectica, 27*, 313–328.

Davidson, D. (1984). *Essays on Actions and Events.* Oxford: Clarendon Press.

Dowty, D. R., Wall, R. E., & Peters, S. (1981). *Introduction to Montague Semantics.* Dordrecht, Holland: D. Reidel.

Dretske, F. I. (1981). *Knowledge and the Flow of Information.* Oxford: Basil Blackwell.

Drew, P. (1990). Conversation Analysis. In Asher, R. & Simpson, J. (Eds.), *The Encylcopedia of Language and Linguistics.* Oxford: Pergamon Press.

Elffers, J. (1973). *Tangram: The Ancient Chinese Shapes Game.* London: Penguin Group.

Feyerabend, P. (1962). Explanation, Reduction and Empiricism. In Feigl-Maxwell (Ed.), *Minnesota Studies in the Philosophy of Science*, Vol. 3, pp. 28–97. Minneapolis: University of Minnesota Press.

Fodor, J. (1975). *The Language of Thought.* Cambridge, Massachusetts: Harvard University Press.

Fodor, J. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioural and Brain Sciences, 3*, 63–109.

Fodor, J. (1981). *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Brighton, Sussex: Harvester Press.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Fodor, J. & Lepore, E. (1992). *Holism: A Shopper's Guide.* Oxford: Basil Blackwell.

Foley, R. (1987). *Another Unique Species: Patterns in human evolutionary ecology.* Harlow: Longman Scientific and Technical.

Frege, G. (1892). Uber Sinn und Bedeutung. *Zeitschrift fur Philosophie und philosophische Kritik, 100*, 25–50. Translated as: On "Sense and Reference" in P.T. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, 1960, pp. 56–77, Oxford: Blackwell.

Galliers, Julia, R. (1989). A theoretical framework for computer models of cooperative dialogue acknowledging multi-agent conflict. Tech. rep. 172, Cambridge University, Centre for Computational Linguistics.

Gardenfors, P. (1993). The emergence of meaning. *Linguistics and Philosophy, 16*, 285–309.

Gardenfors, P. (1991). Conceptual spaces as a basis for cognitive semantics. In *Proceedings of the Second International Colloqiuim in Cognitive Science.* Universidad del Pais Vasco.

Garfinkel, H. (1967). *Studies in Ethnomethodology.* Englewood Cliffs: Prentice Hall.

Garrod, S. C. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Garrod, S. C. & Doherty, G. (1994). Conversation, coordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition, 53*, 181–215.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form.* New York: Academic Press.

Gentner, D. & Gentner, D. (1983). Flowing Waters or Teeming Crowds: Mental Models of Electricity. In Gentner, D. & Gentner., D. (Eds.), *Mental Models*, pp. 99–127. Hillsdale: Lawrence Erlbaum Associates.

Gopnik, A. (1983). Conceptual and Semantic Change in Scientists and Children: Why there are no semantic Universals. *Linguistics, 21*(1), 163–179.

Grosz, B. J. & Sidner, C. L. (1990). Plans for Discourse. In Cohen, P., Morgan, J., & Pollack, M. (Eds.), *Intentions in Communication*, pp. 417–444. Cambridge, Mass: MIT Press.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, *3*(2).

Healey, P. (1991). Evolution, Language and Interaction. Unpublished Masters Thesis, University of Edinburgh, Centre for Cognitive Science.

Healey, P. & Vogel, C. (1994). A Situation Theoretic Model of Dialogue. In Jokinen, K. (Ed.), *Gothenburg Papers in Theoretical Linguistics 71: Pragmatics in Dialogue Management*, pp. 61–79. Gothenburg: Gothenburg University.

Heritage, J. (1984). *Garfinkel and Ethnomethodology.* Cambridge: Polity Press.

Houghton, G. & Isard, S. D. (1987). Why to speak, what to say and how to say it: modelling language production in discourse. In Morris, P. (Ed.), *Modelling Cognition*, pp. 249–267. Wiley: UK.

Humphreys, K. (1993). Given and new information: a terminological minefield. Unpublished manuscript, University of Edinburgh, Centre for Cognitive Science.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, *77*, 187–222.

Hutchins, E. (1983). Understanding Micronesian Navigation. In Gentner, D. & Gentner., D. (Eds.), *Mental Models*, pp. 191–227. Hillsdale: Lawrence Erlbaum Associates.

Jackendoff, R. S. (1992). *Languages of the Mind: Essays on Mental Representation.* Bradford Books, MIT Press, Cambridge.

Johnson-Laird, P. (1983). *Mental Models.* Cambridge: Cambridge University Press.

Kamp, H. & Reyle, U. (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht: Kluwer Academic.

Karmiloff-Smith, A. (1988). The Child is a Theoretician, Not an Inductivist. *Mind and Language, 3*(3), 182–195.

Katz, J. (1972). *Semantic Theory.* New York: Harper and Row.

Katz, J. & Fodor, J. (1963). The structure of a semantic theory. *Language, 39,* 170–210.

Kay, M., Gawron, J. M., & Norvig, P. (1994). *Verbmobil: A Translation System for Face-to-Face Dialog.* Centre for the Study of Language and Information, Stanford University, CA. CSLI Lecture Notes Number 14.

Keil, F. C. (1981). Conceptual development and category structure. In Neisser, U. (Ed.), *Concepts and Conceptual Development: ecological and intellectual factors in categorization,* pp. 175–200. Cambridge: Cambridge University Press.

Kitcher, P. (1988). The Child as Parent of the Scientist. *Mind and Language, 3*(3), 217–228.

Kleinke, C. (1986). Gaze and eye contact: a research review. *Psychological Bulletin, 100*(1), 78–100.

Kowtko, J. C., Isard, S. D., & Doherty, G. M. (1991). Conversational games within dialogue. In *Proceedings of the ESPIRIT Workshop on Discourse Coherence.* University of Edinburgh, 4–6 April.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition.* Blackwell, Oxford.

Kuhn, Thomas, S. (1983). Commensurability, Comparability, Communicability. In *Proceedings of the 1982 Biennial Meeting of The Philosophy of Science Association*, pp. 669–688. East Lansing.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd, enlarged edition). Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago, IL: Chicago University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By.* Chicago, IL: Chicago University Press.

Langacker, R. W. (1986). *Foundations of Cognitive Grammar*, Vol. 1. Stanford, CA: Stanford University Press.

Leudar, I. (1991). Sociogenesis, coordination and Mutualism. *Journal for the Theory of Social Behaviour*, *21*(2), 197–220.

Levinson, S. C. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Lewis, D. (1969). *Convention: A Philosophical Study.* Oxford: Basil Blackwell.

Lewis, D. (1972a). *Counterfactuals.* Harvard: Harvard University Press.

Lewis, D. (1972b). General Semantics. In Davidson, D. & Harman, G. (Eds.), *Semantics of Natural Language*, pp. 168–218. Dordrecht: Reidel.

Linkey, H. & Firestone, I. (1990). Dyad dominance composition effects, nonverbal behaviours and influence. *Journal of Research in Personality*, *24*, 206–215.

Locke, J. (1690). *An Essay Concerning Human Understanding.* Glasgow: Collins, Fount Paperbacks.

Markova, I. (1990). Introduction. In Markova, I. & Foppa, K. (Eds.), *The Dynamics of Dialogue*, pp. 1–22. London: Harvester-Wheatsheaf.

McDowell, J. (1984). Wittgenstein on following a rule. *Synthese, 58*, 325–363.

Miller, G. & Johnson-Laird, P. (1976). *Language and Perception.* Cambridge, Mass: Harvard University Press.

Minsky, M. (1975). A framework for representing knowledge. In Winston, P. (Ed.), *The Psychology of Computer Vision*, pp. 211–277. New York: McGraw-Hill.

Morton, J. (1970). A functional model for memory. In Norman, D. (Ed.), *Models of Human Memory*, pp. 91–121. New York: Academic Press.

Moxey, L. M. & Sanford, Anthony, J. (1993). *Communicating Quantities: A Psychological Perspective.* London: Lawrence Erlbaum Associates.

Neisser, U. (1987). *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization.* Cambridge: Cambridge University Press.

Partee, B. (1979). Semantics – mathematics or psychology? In Bauerle, R., Egli, U., & von Stechow, A. (Eds.), *Semantics From Different Points of View: Proceedings of the Konstanz Colloquium on Semantics*, pp. 1–14. Sringer Verlag: Berlin.

Perrault, C. (1990). An Application of Default Logic to Speech Act Theory. In Cohen, P., J., M., & Pollack, M. (Eds.), *Intentions in Communication*, pp. 161–186. Cambridge Mass.: MIT Press, Bradford Books.

Pettit, P. & McDowell, J. (1986). Introduction. In Pettit, P. & McDowell, J. (Eds.), *Subject, Thought and Context*, pp. 1–16. Oxford: Oxford University Press.

Pollack, M. (1990). Plans as Complex Mental Attitudes. In Cohen, P., J., M., & Pollack, M. (Eds.), *Intentions in Communication*, pp. 77–104. Cambridge Mass.: MIT Press, Bradford Books.

Prince, E. F. (1969). Toward a taxonomy of given-new information. In Cole, P. (Ed.), *Radical Pragmatics*, pp. 223–255. New York: Academic Press.

Prince, E. F. (1988). Discourse analysis: a part of the study of linguistic competence. In Newmeyer, F. J. (Ed.), *Linguistics: The Cambridge Survey III: Extensions and Implications*, pp. 167–177. Cambridge: Cambridge University Press.

Putnam, H. (1970). Is semantics possible? In Keifer, H. & Munitz, M. (Eds.), *Languages, Belief and Metaphysics, Volume 1 of Contemporary Philosophic Thought: The International Philosophy Year Conferences at Brockport.* New York: State University of New York Press.

Putnam, H. (1988). *Representation and Reality.* Cambridge: Bradford Books, MIT Press.

Putnam, H. (1975). The meaning of meaning. In Gunderson, K. (Ed.), *Language, Mind, and Knowledge, Minnesota Studies in the Philosophy of Science, 7.* Minneapolis: University of Minnesota Press.

Putnam, H. (1981). *Reason, Truth and History.* Cambridge: Cambridge University Press.

Quine, W. (1953). *From a Logical Point of View.* Cambridge Mass: Harvard University Press.

Quine, W. (1960). *Word and Object.* Cambridge, Mass: MIT Press.

Quine, W. (1969). *Ontological Relativity and Other Essays.* New York: Columbia University Press.

Quine, W. (1992). *The Pursuit of Truth.* Cambridge, Mass: Harvard University Press. Revised Edition.

Reddy, M. J. (1979). The Conduit Metaphor: A Case of Frame Conflict in Our Language about Language. In Ortony, A. (Ed.), *Metaphor and Thought*, pp. 284–324. Cambridge: Cambridge University Press.

Roger, D. & Nesshoever, W. (1987). Individual differences in dyadic conversational strategies: A further study. *British Journal of Social Psychology, 26*, 247–255.

Rommetveit, R. (1983). In search of a truly interdisciplinary semantics. A sermon on hopes of salvation from hereditary sins. *Journal of Semantics, 2*(1), 1–28.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis* (2nd edition). Singapore: McGraw Hill.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English Verbs. In McClelland, J. L. & Rumelhart, D. E. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2 Psychological and Biological Models*, pp. 216–271. Cambridge, MA: MIT Press.

Sacks, H. (1984). Methodological Remarks. In M., A. J. & Heritage, J. L. (Eds.), *Structures of Social Action: Studies in Conversation Analysis*, pp. 21–27. Cambridge: Cambridge University Press.

Sacks, H. & Schegloff, E. (1979). Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction. In Psathas, G. (Ed.), *Everday Language: Studies in Ethnomethodology.* New York: Irvington.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language, 50,* 696–735.

Sadock, J. M. (1988). Speech act distinctions in grammar. In Newmeyer, F. (Ed.), *Linguistics: The Cambridge Survey III. Linguistic Theory: Extensions and Implications.* Cambridge: Cambridge University Press.

Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Schegloff, E. A. (1972). Sequencing in conversational openings. In Gumperz, J. J. & Hymes, D. H. (Eds.), *Directions in Sociolinguistics,* pp. 346–380. New York: Holt, Rinehart and Winston.

Schegloff, E. A. (1992). Repair after the next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology, 97*(5), 1295–1345.

Schmidt-Besserat, D. (1991). The Earliest Precursor of Writing. In Wang, W. S.-Y. (Ed.), *The Emergence of Language: Development and Evolution,* pp. 31–45. New York: W. H. Freeman and Company. First published in *Scientific American,* 1978.

Schober, M. F. & Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology, 21,* 211–232.

Schutz, A. (1973). Common-sense and Scientific Interpretation of Human Action. In Natanson, M. (Ed.), *Collected Papers, Volume 1: The Problem of Social Reality,* pp. 3–47. Martinus Nijhoff: The Hague.

Schweizer, P. (1991). Blind grasping and Fregean senses. *Philosophical Studies, 62*, 263–287.

Searle, J. (1969). *Speech Acts.* Cambridge: Cambridge University Press.

Searle, J. (1976). The classification of illocutionary acts. *Language in Society, 5*, 1–24.

Searle, J. (1990). Intentions. In Cohen, P., Morgan, J., & Pollack, M. (Eds.), *Intentions in Communication*, pp. 417–444. Cambridge, Mass: MIT Press.

Seligman, J. (1990). Perspectives in Situation Theory. In Cooper, R., Mukai, K., & Perry, J. (Eds.), *Situation Theory and its Applications, Volume I*, pp. 147–191. Centre for the Study of Language and Information, Stanford University, CA.

Seligman, J. & Barwise, J. (1993). Channel theory: toward a mathematics of imperfect information flow. *unpublished manuscript* (available by ftp from phil.indiana.edu/pub/SelBar93.ps).

Shannon, C. & Weaver, W. (1964). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition.* Cambridge, Mass.:MIT Press.

Stalnaker, R. C. (1987). *Inquiry.* Cambridge, MA: Bradford Books, MIT Press.

Taylor, T. J. (1992). *Mutual Misunderstanding: Scepticism and the Theorizing of Language and Interpretation.* London: Routledge.

van Dijk, T. A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension.* New York: Academic Press.

Wieder, D. (1974). *Language and Social Reality.* The Hague: Mouton.

Wilkes-Gibbs, D. & Clark, H. H. (1992). Coordinating Beliefs in Conversation. *Journal of Memory and Language, 31*, 183–194.

Wiser, M. & Carey, S. (1983). When Heat and Temperature Were One. In Gentner, D. & Gentner., D. (Eds.), *Mental Models.* Hillsdale: Lawrence Erlbaum Associates.

Wittgenstein, L. (1958). *Philosophical Investigations* (Second edition). Oxford: Basil Blackwell. Translated by G.E.M. Anscombe.

Wooffitt, R. (1990). On the Analysis of Interaction: An Introduction to Conversation Analysis. In Luff, P., Gilbert, G., & Frohlich, D. (Eds.), *Computers and Conversation*, pp. 7–38. New York: Academic Press.