Optimal Design in Language Production

Sarah L. Haywood

Ph.D.

The University of Edinburgh

2004

For Shauneen Mary Haywood

24 May 1952 – 6 August 1998

Writing a story is not simply a matter of writing lines of words, but calls on the writer to assemble the sentences in such a way that the reader receives them in the right order for stacking in the mind.

(Oliver Postgate, *Seeing Things*)

Contents

Tables and Figures

Declaration

This thesis has been composed by myself, and the research presented herein is my own. No portion of the work has been submitted for any other degree or professional qualification.

Sarah Haywood

Abstract

Psycholinguistic accounts of language production have traditionally been informed by evidence from highly constrained, non-interactive experimental tasks, such as picture description and sentence completion. These studies are informative about the mechanisms and representations that underlie production, but they tell us little about the impact of communicative *context* on those basic processes. This thesis examines language behaviour in more naturalistic situations, where the speaker is talking to a co-present addressee. This kind of setting more closely reflects production outside the laboratory, where speakers need to make themselves understood if communication is to be successful. In particular, the thesis investigates whether speakers follow a principle of 'optimal design' at the level of grammatical encoding. Optimal design can be interpreted in different ways; speakers may say things that are easy to produce, maximising efficiency for themselves. Alternatively, they might aim to produce messages that are easy for an audience to understand (or they might trade off between these goals). The thesis focuses on whether speakers take addressees' perspectives into account when they formulate syntactic structure and word order.

Referential communication paradigms were used to investigate language production during collaborative tasks. Speakers described picture cards or other objects so that an addressee could pick out the intended referent from an array. The structure of the array was manipulated such that particular syntactic structures or word orders would be easier for the addressee to interpret than others. The research suggests that grammatical stages of language production *can* be sensitive to information about an addressee's perspective. Speakers show evidence of optimal design in their choice of syntax and word order, but only when it is obvious how they can make their utterances easy to understand. Optimal design for the addressee's benefit seems to have less of an impact on production than other, more automatic processes (such as a tendency towards syntactic alignment with an interlocutor). The likelihood of a speaker attending to an addressee's perspective can also be influenced by the addressee's own language behaviour.

Acknowledgements

The research reported in this thesis greatly benefited from the help and advice of many people. Special mentions go to: Martin Pickering and Holly Branigan, for supervision and support; Sandie Cleland, Manon Jones, and Susie Flett, for acting as confederates in Experiments 4, 5 and 6; Janet McLean, for providing invaluable advice on statistical analysis; Hannah Reeve, for recruiting participants for Experiment 3; Nivja de Jong, for taking the digital photographs used to create instruction cards for Experiment 6; Davy Wilkinson, for providing assistance with audio recording and equipment; and Roy Welensky and Jamie Pearson, for help with printing and graphics.

CHAPTER 1
Introduction

1.0 Chapter Overview

This chapter gives a brief introduction to the notion of optimal design in language production, and outlines why it might be important for successful communication. I suggest reasons why optimal design has received relatively little attention in psycholinguistic accounts of production to date, and introduce the issues to be addressed by the thesis.

1.1 Taking Perspectives in Conversation

This thesis is fundamentally about how speakers communicate during conversation. More specifically, it is about language production, and particularly how speakers refer to objects in the world around them. Clark, Schreuder and Buttrick (1983) proposed a *Principle of Optimal Design* for language, which essentially suggests the speakers formulate utterances with an interlocutor's perspective in mind. In the optimal design literature, "perspective" is often used as short-hand for "… another person's time, place, and identity, or their conceptualisations, conversational agendas, or knowledge." (Hanna, Tanenhaus, & Trueswell, 2003, p.51). In other words, a person's perspective is taken to include not only their perceptual or physical vantage point in space and time, but also their mental and emotional state: what they know (or do not know), think, believe and feel. In conversation, speakers and addressees almost certainly have different perspectives, even when the differences are subtle. One of the challenges in speaking, then, is to engage in optimal design – to step out of one's own perspective in order to say things in such a way that they are understandable for an interlocutor.

Why is optimal design so important? Without it, communication is likely to be hard work. Imagine that Angela suddenly turns to Nick and asks *Did you hear that he got it?* If Nick knew that Angela had just been thinking about a job that their nephew had recently interviewed for, then the question would make perfect sense. But presumably Nick is not a mind-reader, and when the question comes out of the blue, he is likely to have to ask for more information (Who's *he*? And what's *it*?) before he can answer. In Gricean terms, Angela has been "uncooperative" by violating the Maxim of Quantity (Grice, 1975); she didn't give Nick enough information to interpret her question, and in this sense she has failed to communicate successfully. If Angela routinely produced such uncooperative utterances, she would probably be considered a very bad communicator (Donaldson, 1978). Indeed, psychologists have claimed that some disorders in which communication skills are impaired (specifically, autism and Asperger's syndrome) stem from a failure to understand that other people have perspectives that are different from one's own (e.g., Baron-Cohen, 1995).

Being able to tailor utterances to an interlocutor's perspective is clearly an important skill to have, but it is one that has to be acquired. Young children are notoriously bad at optimal design (see Nadig & Sedivy, 2002, for a review), a fact which Piaget (1926) put down to a fundamental inability to 'decentre' or think outside their own perspective. In fact, some research suggests that even young children show some sensitivity to their addressee's knowledge state (e.g., Nadig & Sedivy, 2002; O'Neill, 1996; Perner & Leekham, 1986). However, although they might be aware that information known to themselves is not shared by the addressee, they do not always manage to convey such "privileged" information successfully (Maratsos, 1976). For example, children as old as 7 or 8 sometimes produce references that are underspecified or ambiguous (Deutsch & Pechmann, 1982; Glucksberg, Krauss, & Weisberg, 1966; Sonnenschein & Whitehurst, 1984), and they are poor at monitoring conversation for potential misunderstandings (Garrod & Clark, 1993). Clearly, then, learning to engage in optimal design is an important part of the development of communication skills. In this thesis, we are concerned with the end point of this developmental process: to what extent do adults tailor what they say to an addressee's perspective, and how is this manifested in adult language processing? Our focus is on optimal design in spoken language production.

1.2 Optimal Design in Psycholinguistic Accounts of Production

What we know about optimal design has been informed by work in philosophy (Austin, 1962; Grice, 1975; Searle, 1969), sociology (Goffman, 1981), sociolinguistics (Bell, 1984), conversation analysis (Sacks, Schegloff, & Jefferson, 1974), theoretical linguistics (Hawkins, 1994) and phonetics (Lindblom, 1990). However, standard *psycholinguistic* theories of language processing have traditionally had little to say about the role of optimal design. This is probably due (at least in part) to the fact that psycholinguists have generally taken a "monologue" approach to studying language (Pickering & Garrod, in press). In order to isolate the basic cognitive representations and mechanisms that underlie production, sentences and utterances are usually analysed out of the contexts in which they would ordinarily be produced. In other words, psycholinguists have decontextualised language for the most part, by studying production independently from comprehension, in situations where speakers have few or no communicative goals (Pickering & Garrod, in press). For example, in production studies, participants are typically asked to give picture descriptions (Bock, 1986) or to complete sentence fragments (Hartsuiker & Westenberg, 2000; Pickering & Branigan, 1998) which are provided one-by-one by the experimenter, outside of a meaningful dialogue. Where there is any kind of "context" for production at all, it is usually extremely limited (e.g., Prat-Sala & Branigan, 2000).

Given the focus on monologue, it is hardly surprising that psycholinguistic models of language production have rarely addressed optimal design. Whether or not speakers take an addressee's perspective is simply not an issue in the experiments upon which such models are based; in many production studies there is no addressee for the speaker to take into account in the first place. Although the monologic approach to psycholinguistic research has helped to understand the cognitive processes that support language production, it might not tell us very much about how people actually *use* language in conversation (Pickering & Garrod, in press). Moreover, theoretical models based only on monologue may not be able to account for the full range of (contextual) influences on language production. After all, "Human communication is not just a transfer of information like two fax machines connected by a wire; it is a series of alternating displays of behavior by sensitive, scheming, second-guessing, social animals." (Pinker, 1994, pp. 229-230). A full account of situated language processing will therefore need to consider the extent to which optimal design affects language production, and the circumstances under which speakers do or don't take their addressees' perspectives into account. A growing body of literature has begun to address these questions, and this thesis aims to add to that literature.

1.3 Questions to be Addressed in the Thesis

This thesis experimentally addresses a number of issues relating to optimal design in spoken language production. The methodology used here is an adapted form of what has become known as the *referential communication paradigm* (see Chapter 2). Pairs or groups of participants play a game in which they give and follow instructions, or describe a set of visual stimuli to each other. They have a goal which can only be achieved by verbally collaborating on the task. For example, one participant helps their partner arrange a set of picture cards in a particular order. This kind of task has traditionally been used to study the processes and strategies of semi-naturalistic conversation (e.g., Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1989; Hupet, Seron, & Chantraine, 1991; Krauss & Weinheimer, 1964; 1966). In the experiments reported here, the addressee's perspective is manipulated such that they will find some kinds of utterance easier to understand than others. By recording and analysing the kinds of descriptions produced by the speaker, we can infer whether knowledge of the addressee's perspective is being taken into account during production.

The thesis investigates four broad questions concerning the role of optimal design in language production. Chapter 3 explores the effects of cognitive load on a speaker's ability to take addressees' conceptual perspectives into account in conversation. Chapter 4 asks whether there is evidence for optimal design at the level of syntactic (grammatical) encoding. Chapters 5 and 6 explore the relationship between optimal design and another process known to have pervasive effects on grammatical stages of language production: syntactic priming, or the tendency to re-use a syntactic structure that has recently been processed. Chapters 5 and 6 also investigate whether the likelihood of a speaker taking an interlocutor's perspective into account is affected by that interlocutor's own language behaviour. Does optimal design have automatic effects on production, or is it used more strategically? More specifically, are speakers more likely to engage in optimal design when their conversational partner is doing the same?

We will see that optimal design has pervasive effects on grammatical and conceptual aspects of utterance production. These experiments suggest that speakers can tailor not only what they say, but also how they say it, to meet the needs of their addressee.

CHAPTER 2

Literature Review

2.0 Chapter Overview

This chapter provides an introduction to the optimal design literature, outlines theories about its role in language production, and discusses some of the empirical evidence used to support those theoretical approaches. In 2.1 I discuss various conceptions of what makes for "optimal design" in language production, and introduce some important theoretical concepts, such as common ground, mutual knowledge, copresence heuristics, and addressee needs. 2.2 discusses how optimal design might manifest itself in speech, and relates this issue to "standard" models of production (e.g., Bock, 1995; Garrett, 1975; 1980; Levelt, 1989).

2.3 describes methods for investigating optimal design, and introduces the referential communication paradigm. 2.4 – 2.6 outline a range of theoretical models of optimal design in production, and present some of the empirical data used as evidence for and against these models. Three broad types of model are indentified: those suggesting that production is fundamentally "addressee-optimal", or guided by a model of the interlocutor (2.4), those suggesting that production is "speaker-optimal", or geared towards efficiency (2.5), and those suggesting that production is constrained by ordinary cognitive limitations (2.6).

2.1 What is Optimal Design?

Whenever the language production system generates an utterance, choices have to be made (Ferreira & Dell, 2000). For example, the average adult has a vocabulary of around 70,000 words (Altmann, 1997), many of which overlap in meaning with at least one other word (e.g., *mouse* and *rodent*). One of the fundamental choices the production system has to make is which words to use to express a message. Other choices concern the syntactic structure of the utterance, such as the choice between the simple present tense (*the cat chases the mouse*) and the present progressive (*the cat is chasing the mouse*) or between an active sentence structure and a passive (*the mouse is chased by the cat*). Utterances can be considered "optimally designed" to the extent that the choice made by the production system is somehow better than the alternative(s). What makes one choice "better" than another? In the following sections, two different conceptions of optimal design in production are introduced, one based on what is easy for the addressee, and one based on what is easiest for a speaker.

2.1.1 Optimal design "for" the addressee

The term "Optimal Design" was first introduced by Clark, Schreuder and Buttrick (1983), although it has precursors in work by Clark & Marshall (1981), Clark and Carlson (1981; 1982) and Clark and Murphy (1983). Clark and Marshall (1981) introduced a hypothesis about how interlocutors solve the seemingly-intractable "Mutual Knowledge Paradox" for definite reference. Standard linguistic theories claim that the use of a definite reference (e.g., *the movie at the Roxy*) presupposes that there is some object or entity which is mutually known as the unique referent of the expression by both speaker and addressee (e.g., Lakoff, 1974). So, to use and understand definite references felicitously, speakers and addressees have to assess what they mutually know (Lewis, 1969; Schiffer, 1972). However, fully assessing mutual knowledge is computationally impossible, because it requires an infinite number of recursive inferences about who knows what (e.g., Pickering & Garrod, in press). "The idea is that for a speaker to talk about *the sofa* (as opposed to *a sofa* or *some sofa*), she must believe that her listener can determine which particular sofa she is talking about… Technically, this isn't sufficient; the listener must also believe that the speaker believes that the listener knows which particular sofa is under discussion, and the speaker must believe that the listener believes that the speaker believes that the listener knows which particular sofa is being referred to… and so on." (Schober & Brennan, 2003, p.137), ad infinitum. If assessing mutual knowledge requires an infinite number of such recursive inferences, how on earth do speakers ever use definite reference, given finite cognitive resources, and a finite amount of time for processing?

Clark and Marshall's suggestion was that interlocutors work out what constitutes mutual knowledge by consulting their *common ground* (Stalnaker, 1978), or a model of the knowledge, beliefs and assumptions that they share. Common ground, under Clark and Marshall's view, is based on three kinds of information: perceptual evidence, linguistic evidence, and community membership. Perceptual evidence comes from the physical context of communication: information is in common ground if it comes from some perceptual event that interlocutors have experienced together. For example, if two friends are walking along and having a conversation, perceptual evidence for common ground might come from seeing an unusual bird fly by, hearing a dog bark, or smelling freshly-baked bread as they pass a baker's shop. Linguistic evidence comes from what has been said and heard already, in this and previous conversations.

Community membership involves knowledge that can be assumed to be shared by interlocutors, given the social groups to which they mutually believe they belong. For example, imagine that two psychologists meet for the first time at a party. One says *Are you a fan of Piaget?* Using a proper name (*Piaget*) to a complete stranger without first introducing the referent would ordinarily be considered infelicitous. However, since Piaget is a well-known figure in the field of psychology, the speaker can reasonably assume that this referent is part of the common ground she shares with her addressee. The speaker makes this assumption, according to the collaborative model, because she believes that both parties are part of same community – psychologists – and that most psychologists will know just one "Piaget".

Collectively, perceptual evidence, linguistic evidence and community membership are known as *copresence heuristics*. Clark and Marshall proposed that interlocutors routinely use these heuristics to create a model of their conversational partner. Speakers and addressees, they suggest, constantly consult and update such models as they talk and understand together. Modelling the listener involves taking the addressee's perspective, where "perspective" is simply short-hand for the addressee's knowledge, beliefs and needs (e.g., Hanna, Tanenhaus, & Trueswell, 2003; see also Fussell & Krauss, 1992, and Schober, 1998). Clark, Schreuder and Buttrick (1983) formalised these ideas in the *Principle of Optimal Design*:

> *The speaker designs his utterance in such a way that he has good reason to believe that the addressee can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground* (Clark, 1992, pp. 80-81).

By this definition, Clark et al.'s original notion of Optimal Design is similar to (indeed, almost indistinguishable from) what Clark and Carlson (1982) called *Audience Design*: the process of tailoring utterances according to what the speaker knows, believes, and/or supposes about hearers in various roles (addressees, side-participants, overhearers, and so on). Audience design is taken to be a fundamental property of utterances (though note that this does not necessarily imply that adaptations to the audience are made automatically and require no processing resources; see Polichak & Gerrig, 1998, for discussion). Audience design is a notion related to, but subtly different from, the *Principle of Least Collaborative Effort*, which was introduced by Clark and Wilkes-Gibbs (1986) in their characterization of how speakers refer to things in dialogue. According to Least Collaborative Effort, participants in a dialogue have a shared responsibility to make sure that what has been said in a conversation has been mutually understood. In situations where the addressee is free to provide feedback, the responsibility for efficient communication is shared; both (or all) parties can contribute to ensuring that the conversation proceeds smoothly.

Importantly, in describing the Principle of Least Collaborative Effort, Clark and colleagues acknowledge that speakers don't always get it right first time (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Dialogue is an interactive process in which the speaker and addressee share equal responsibility for making sure that what they have said to each other has been understood well enough for current purposes (Clark & Wilkes-Gibbs, 1986). This means that speakers do not necessarily have to do all the donkey work with respect to optimal design; in certain circumstances (such as when they are under pressure to start speaking quickly), they may produce utterances that are not *maximally* optimal from the addressee's point of view. Where this is the case, the speaker should try to mark their contribution as tentative, or provisional (Clark & Wilkes-Gibbs, 1986). But they should still be able to get their message across, so long as the addressee can give feedback or ask for clarification (e.g., Krauss et al., 1977; Krauss & Weinheimer, 1966). Moreover, there may be some circumstances in which the conversation as a whole is more efficient if the speaker does *not* try to adapt to the addressee's perspective (see 2.1.3 below).

To sum up, under Clark and colleagues' conceptualisation, speakers adhere to the Principle of Optimal Design when they formulate an utterance (as far as they can), and addressees assume that speakers have followed this principle when they interpret utterances. Language processing is optimal, then, in the sense that both speaking and understanding are circumscribed by what is assumed to be mutually known. Speakers tailor what they say to fit the common ground they share with the current addressee; they actively engage in *audience design* (Clark & Murphy, 1983) – adapting utterances to fit the addressee's perspective. Conversely, addressees interpret what they hear with reference to information that is in common ground (Clark & Carlson, 1981). So, the Principle of Optimal Design accounts for how speakers and addressees make use of common ground during both production and comprehension.

2.1.2 Optimal design "for" the speaker

Clark and colleagues' conception of optimal design is based on the idea that speakers consider their interlocutor's perspective when they can. Other authors have suggested that language processing is optimal for a different reason: because it prioritises the speaker's "needs" (for fluent and efficient speech) in the first instance. Under this view, "… the design of a language processing system is "optimal" not because it guarantees mutual understanding, but because it provides adequate real-time understanding [for addressees, and adequate real-time production, for speakers] at a minimal cognitive cost." (Barr & Keysar, 2002, p.392). From the speaker's perspective, when the production system is faced with a choice, one alternative might be easier or more efficient than another. For example, frequently used words are known to be more accessible to the production system than words that are used infrequently (e.g., Jescheniak & Levelt, 1994), so choosing *mouse* over *rodent* might make for more efficient production of an utterance containing that word.

Sometimes what is optimal for the speaker is also optimal for the addressee; what is easiest or most efficient from the speaker's point of view happens to overlap with the addressee's communication needs (Brown & Dell, 1987; Dell & Brown, 1991). This is often the case with lexical choices: high-frequency words are not only quicker to produce, they are also quicker to be understood (e.g., Morton, 1969), so choosing *mouse* over *rodent* is optimal for both speaker and addressee. But sometimes an utterance which is optimally designed from the speaker's perspective is not the easiest thing for the addressee to understand.

Consider an utterance like *Angela shot the man with the gun* (Keysar & Henly, 2002). The speaker who produces this utterance presumably knows what it is supposed to mean; after all, they generated the original pre-verbal message that they wanted to articulate. But this utterance is globally ambiguous – it has (at least) two possible interpretations. The utterance could either be about Angela using a gun to shoot someone, or about Angela shooting a man who happened to be carrying a gun. From the addressee's point of view, ambiguous utterances are not optimal, because they leave room for misinterpretation. The language comprehension system might prefer speakers to explicitly disambiguate syntactic structure, for example by saying something like *Angela shot the man <u>who was carrying</u> the gun*, but presumably this would entail more cognitive work for the speaker. As we will see in Chapter 6, research to date suggests that speakers do not systematically avoid producing syntactically ambiguous utterances, perhaps because checking each and every planned utterance for potential ambiguities would be a laborious task for the production system (e.g., Ferreira & Dell, 2000). In this sense, there is evidence that production is designed to be optimal for the *speaker* rather than the addressee (see 2.5 below).

Other evidence that production is designed to maximise ease of processing for the speaker comes from the *priming* literature. Priming phenomena are reviewed in more detail in Chapter 5, but briefly, priming refers to the tendency for speakers to repeat linguistic structures (words, phrases, sentence structures) that they have recently either heard or said themselves (e.g., Bock, 1986; Branigan, Pickering, & Cleland, 2000; Cleland & Pickering, 2003). Explanations for syntactic priming effects are sometimes framed in terms of ease of processing for a speaker.

Priming is assumed to make production easier, presumably, re-using a syntactic structure that has recently been processed is much more efficient than generating a new syntactic structure from scratch (Hartsuiker & Kolk, 1998; Hartsuiker & Westenberg, 2000; Pickering & Branigan, 1999), so priming allows speakers to efficiently produce fluent utterances at a minimal cognitive cost. Experimental evidence seems to support this idea. For example, speakers are quicker to produce a syntactic structure with which they have been primed, compared to the alternative, non-primed, structure (Corley & Scheepers, 2002; Smith & Wheeldon, 2001). Even Broca's aphasics, who are impaired in many aspects of production, show strong syntactic priming effects (Hartsuiker & Kolk, 1998), as do children (Huttenlocher, Vasilyeva, & Shimpi, 2004). In this sense, we can think of priming as a manifestation of optimal design from the speaker's perspective. The language processing system seems to be set up to take advantage of what makes production easy.

Priming could also ease the burden of comprehension for the addressee. Many theories of language processing assume that linguistic representations are shared between the comprehension and production subsystems (e.g., Pickering & Garrod, in press). Given this parity between speaking and understanding, the comprehension system could in principle use the fact that speakers tend to repeat linguistic structures to its advantage, by employing the same syntactic structure to parse adjacent utterances. This would be particularly useful in resolving global ambiguities; parsing a semantically unambiguous sentence (e.g., *Angela kissed the man with the flower…*) might help the addressee work out the intended meaning of a subsequent utterance with an ambiguous syntactic structure (*…and shot the man with the gun*). In this sense, priming might be optimal for the interaction as a whole, because it minimises the joint or collaborative effort that speakers and addressees expend in communicating messages successfully.

2.1.3 "Optimal design" as it is used in the thesis

So optimal design can mean different things from the speaker's and addressee's perspectives, and the two extreme positions described above take very different views as to what constitutes an optimally designed utterance. Clark and colleagues' view emphasises what is easy for the addressee, while an alternative conception (associated with Keysar and colleagues, amongst others) emphasises what is easy for the speaker. In practice, of course, speakers may in fact trade off between these different constraints. For example, consider a speaker who has low spatial ability but knows that his interlocutor has good spatial skills (Schober & Brennan, 2003). In describing some object's location in a room, he might choose to give a description from his own spatial perspective (*on my right*) rather than his addressee's, on the assumption that this will make for more efficient communication overall. Essentially, the speaker's utterance reflects what is optimal for the interaction as a whole (Schober & Brennan, 2003). In this sense, the speaker has produced what appears on the surface to be an egocentric utterance, but he has done so for a good reason – to minimise the joint effort that he and the addressee will expend. So this is an example of a speaker who does not appear to be engaging in audience design, but who is nonetheless adhering to the Principle of Least Collaborative Effort. Conversely, utterances that *look like* they are adapted to the addressee's perspective may in fact have been planned to facilitate production for the speaker, and without reference to the addressee's needs at all (see 2.5 below). This clearly complicates the notion of optimal design, and how we go about assessing the evidence for it.

With this caveat in mind, I will use the term optimal design here in the way that Clark, Schreuder and Buttrick (1983) originally suggested it; i.e., as a process of tailoring utterances to the addressee's needs, so far as the speaker is reasonably able. I use the term *Optimal* Design (rather than Audience Design) because, as we will see, speakers in the experiments I describe in Chapters 3, 4, 5 and 6 may not in fact be consulting a model of the addressee in order to shape their utterances appropriately. I therefore use the term optimal design to cover linguistic choices that are co-operative or helpful, in the sense that they make communication easier and more efficient for the participants, irrespective of whether the speaker consults a model of the addressee in making those choices. In the following section, I outline how optimal design might manifest itself at different levels of the production system.

2.2 How Might Speakers Adapt to Addressees?

2.2.1 Generic-listener versus Particular-listener adaptations

What does optimal design look like? Or, how do we know that a speaker is designing what they say (or how they say it) "for" their addressee? Brown and Dell (1987; Dell & Brown, 1991) draw an important distinction between what they call *particular-listener* and *generic-listener* adaptations. Particular-listener adaptations are associated with the individual characteristics of the current addressee, or the context of the conversation. "For example, when the listener is far away the speaker shouts, or if he is taking notes the speaker talks more slowly." (Dell & Brown, 1991, p.106). Generic-listener adaptations, on the other hand, simply reflect choices that speakers often make, because they facilitate comprehension in general, no matter who the addressee might be. For example, speakers might repair speech errors in such a way that the addressee can easily infer where the error occurred (Levelt, 1983). For the strongest evidence of optimal design, we need to look at the adaptations speakers make for particular addressees, or in particular contexts. In the following section, I outline a current model of language production which will set the scene for a discussion of where and how optimal design might be expected to manifest itself.

2.2.2 Standard model of language production

Although there is still debate over the precise architecture and fine details of each process, current models of language production (see Bock, 1995; Bock & Levelt, 1994; Dell, 1986; Garrett, 1975; 1980; Levelt, 1989) assume that there are three major stages involved in planning an utterance: message generation (or conceptualisation), grammatical encoding, and phonological encoding. *Conceptualisation* is the pre-verbal stage where the message to be expressed is generated. *Grammatical encoding* is broken down into two sub-stages, called functional processing and positional processing respectively. During *functional processing*, the lexical items used to encode the pre-verbal message are selected, and syntactic functions are assigned to those words. A syntactic framework is built which specifies the role each lexical item will play in the final utterance (e.g., grammatical subject or object). Functional representations are assumed not to contain the phonological forms of the words that will end up in the final utterance. Rather, this stage is thought to involve representations known as lemmas (e.g., Kempen & Huijbers, 1983), which encode abstract grammatical information such as word class (verb, noun, adjective, and so on). *Positional processing* takes the output of functional processing and assigns the various constituents (the morphemes) to a serial order that satisfies the grammatical rules of the language. The final stage, *phonological encoding*, involves retrieving specific phonological forms for the various components of the utterance, and spelling out prosody, word duration, and so on. The output of the phonological encoding stage (a fully specified linguistic plan for the utterance) is then fed forward for motor programming.

FIGURE 1

Tangram figure (from Clark & Wilkes-Gibbs, 1986)

In principle, optimal design could have an impact on any (or all) of the stages described above (Schober & Brennan, 2003). At the message level, speakers could conceptualise ideas and objects in a way that they know will make sense for the addressee. For example, if one person has previously referred to the ambiguous shape in Figure 1 as *an ice skater*, then their conversational partner could use the same conceptualisation, *ICE SKATER*, in a subsequent reference to it (e.g., Clark & Wilkes-Gibbs, 1986). At the grammatical encoding stage, optimal design could manifest itself in several ways. For example, the speaker might choose particular words that make the message more understandable for the addressee. They might also choose word orders and syntactic structures that make the utterance easy to parse and interpret (this is assumed to be the case in speech directed at children, e.g., Newport, Gleitman, & Gleitman, 1977; Snow, 1972, and at non-native speakers, e.g., Hatch, 1983; Long, 1983). At the level of phonological encoding, they might generate prosodic cues that reflect the intended structure or meaning of the message, and aim to speak clearly, such that individual words are carefully articulated and maximally intelligible. Speakers might also organise their entire discourse in order to be as explicit as possible (Schober & Brennan, 2003).

Of course, the likelihood that speakers actually make these kinds of partner-adaptations is potentially open to a range of influences (see Schober & Brennan, 2003), but assuming that the speaker is motivated and able to take their partner's perspective, which of the adaptations described above do they make? Some of the empirical evidence will be presented in 2.4 – 2.6 below. But first I introduce some of the ways in which psychologists have set about looking for evidence of optimal design in conversation.

2.3 How is Conversation Studied?

2.3.1 Corpus approaches versus laboratory-based paradigms

Insights into the mechanics of conversation have mainly come from two kinds of research: corpus studies and laboratory experiments (Schober & Brennan, 2003). Transcribed corpora of unconstrained, naturally-occurring conversation are ecologically valid. In (free) conversation, interlocutors set their own agendas and have real communicative goals. In this sense, corpus approaches have an advantage over laboratory studies, where interlocutors most often have the goals of an interaction imposed upon them by the experimenter. However, transcription approaches bring their own set of methodological problems. For example, corpus analysts typically have no independent, objective evidence about the intentions and mental states of participants in a dialogue; instead, these must be inferred or "read into" a transcript of what was said (Schober & Brennan, 2003). Clearly this is a serious issue in studying some kinds of optimal design. For example, how can we decide whether a speaker took their addressee's spatial perspective into account if we have no way of knowing what the addressee's perspective actually was at that precise moment in the conversation?

Laboratory-based approaches to the study of language production have the advantage that the experimenter can put constraints on what the participants talk about, usually by giving them some kind of collaborative goal or task. The speaker's intentions can be inferred, because they are constrained by the task; what the addressee understands the speaker to mean can be inferred from overt behaviours like looking at (e.g., Barr & Keysar, 2002; Metzing & Brennan, 2003), or reaching for some object (e.g., Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Goal-directed tasks afford the experimenter much greater control over the dialogue, allowing them to manipulate certain factors of interest, such as who takes part (adults, children, non-native speakers, etc.), how well the participants know each other (close friends, acquaintances, complete strangers), what they know about each other's perspective, and how they communicate (face-to-face, via telephone, electronically, over some kind of the computer-mediated dialogue system, and so on). Giving speakers and addressees a collaborative goal also builds in an independent measure of how successful their interaction is – i.e., how quickly and accurately they complete the task they have been set (Brennan & Ohaeri, 1999; Lantz & Stefflre, 1964). The focus of the remainder of this chapter is on experimental paradigms, because these seem to afford the best opportunity to understand the cognitive processes that underlie language use in dialogue contexts (Schober & Brennan, 2003).

2.3.2 Referential communication paradigms

Krauss and colleagues developed a laboratory-based method which has come to be known as the *referential communication paradigm* (Krauss & Bricker, 1967; Krauss, Garlock, Bricker & McMahon, 1977; Krauss, Vivekananthan & Weinheimer, 1968; Krauss & Weinheimer, 1964; 1966). Broadly speaking, a referential communication task can be anything that allows (or requires) interlocutors to talk about some object or set of objects out in the world. This kind of task can be traced back to Piaget's perspective-taking studies with children (such as the "three mountains" experiments; Yule, 1997), and similar methodologies were used by Carroll (cited in Osgood & Sebok, 1954) and Maclay (1962; see Krauss & Weinheimer, 1966). Referential communication tasks were widely used during the 1960s, particularly amongst researchers interested in the development of communication skills in children (e.g., Glucksberg, Krauss, & Higgins, 1975; Glucksberg, Krauss, & Weisberg, 1966; Krauss & Glucksberg, 1969; Maratsos, 1976). These tasks are appealing to developmental researchers because they can be adapted for even very young children to understand. Describing a picture or an object is a relatively easy task for a child; it requires no literacy skills, is constrained enough to be motivating and interesting, but is open-ended enough that they can complete the task using whatever vocabulary they know. Another great advantage of tasks involving conversation between participants is that they "preserve many aspects of face-to-face conversation, which is arguably the primary site for language use." (Hanna et al., 2003, p.44).

In a referential communication task, pairs or groups of participants typically play a game in which they give and follow instructions, or describe a set of visual stimuli to each other. They have a goal which can only be achieved by verbally collaborating on the task, because one player has information that their partner needs. The person who has the crucial information has been called the speaker (Krauss & Weinheimer, 1966; Kraut, Lewis, & Swezey, 1982), the director (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992), the sender (Krauss & Weinheimer, 1964), the information or instruction giver (Bard & Aylett, in press; Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Garrod & Anderson, 1987), the instructor (Lloyd, 1991), the explainer (Blakar, 1984), the expert (Kraut, Miller, & Siegel, 1996), or the describer (Branigan, McLean, & Reeve, 2003). The other player has been known as the listener (Krauss & Weinheimer, 1966; Kraut, Lewis, & Swezey, 1982), the matcher (Clark & Wilkes-Gibbs, 1986), the receiver (Krauss & Weinheimer, 1964), the information or instruction follower (Anderson, Bader, Bard, Boyle, Doherty, Garrod, Isard, Kowto, McAllister, Miller, Sotillo, Thompson, & Weinert, 1991; Garrod & Anderson, 1987), the instructee (Lloyd, 1991), the follower (Blakar, 1984), the worker (Kraut, Miller, & Siegel, 1996), the model builder (Markman & Makin, 1998), and the helper (Hanna & Tanenhaus, 2004). Following Clark and Wilkes-Gibbs (1986), I use the terms 'Director' and 'Matcher' to describe participants' roles in referential communication tasks.

Many studies require the Director to describe some kind of visual stimulus so that the Matcher can select a matching object from an array. Alternatively, both players have an array in front of them and the goal is for the Director to help the Matcher organise the array in some "target" order (e.g., Clark & Wilkes-Gibbs, 1986). Director and Matcher may have exactly the same array in front of them (Clark & Wilkes-Gibbs, 1986) or they may have mismatching information (Bard & Aylett, in press; Bard et al., 2000), or only partly overlapping arrays (Horton & Gerrig, 2002) Some studies require the Director to help the Matcher draw a route on a map (Bard & Aylett, in press; Bard et al., 2000), build a model (Markman & Makin, 1998), re-tell a story (Bavelas, Coates, & Johnson, 2000; Brown & Dell, 1987; Gregory, Healy, & Jurafsky, submitted; Lockridge & Brennan, 2002) or describe a video clip such that their partner can answers questions about it (Kraut, Lewis, & Swezey, 1982). Other experiments have the added complication that the Director must help an addressee to select the correct referent, whilst simultaneously misleading a third participant about the object they are referring to (Clark & Schaefer, 1987).

The objects in the 'referent array' (Olson, 1970) may be unusual shapes and unfamiliar to both players (as in experiments using *tangram figures* as stimuli; Clark & Wilkes-Gibbs, 1986; Hupet, Seron, & Chantraine, 1991; Wilkes-Gibbs & Clark, 1992; see Figure 1), or they might involve more familiar objects (Brennan & Clark, 1996; Horton & Gerrig, 2002), such as pictures of furniture and household appliances (Bortfeld & Brennan, 1997; Fussell & Krauss, 1992). Referential communication experiments have also involved Directors describing real objects (Barr & Keysar, 2002; Keysar et al., 2000; Metzing & Brennan, 2003), photographs of landmarks and places (Clark & Schaefer, 1987; Isaacs & Clark, 1987) or children (Schober & Carstensen, 2001), stories (Brown & Dell, 1987; Dell & Brown, 1991), maps (Bard & Aylett, in press; Bard et al., 2000; Brennan, 1990; Wilkes-Gibbs, 1986) and video clips (Kraut, Lewis, & Swezey, 1982). The experimenter can manipulate the array such that the object being described by the Director is very different from all the other objects (and therefore easy to pick out) or similar to one or many of them (Hupet et al., 1991).

The experimenter also has control over the interactivity of the task. Can the Matcher give feedback (e.g., Krauss & Weinheimer, 1966)? Do the players switch roles during the game, and if so, how often? Both of the players may be real participants, or one might be a confederate of the experimenter (Barr & Keysar, 2002; Brown & Dell, 1987; Hanna et al., 2003; Horton & Keysar, 1996; Metzing & Brennan, 2003). Participants might be adults (Clark & Wilkes-Gibbs, 1986), children (Glucksberg, Krauss, & Weisberg, 1966), friends (Clark & Schaefer, 1987; Fussell & Krauss, 1989), acquaintances (Fussell & Krauss, 1989; Schober & Carstenson, 2001) or complete strangers (Wilkes-Gibbs & Clark, 1992). They may be experts or novices in some domain (Isaacs & Clark, 1987; Schober, 1993), such as native speakers of a language or second-language learners (Bortfeld & Brennan, 1997). They may be physically co-present and sitting next to each other, or looking at the same array but from slightly different spatial perspectives (Schober, 1993; 1995). They might be in the same room but unable to see what their partner is doing (e.g., Wilkes-Gibbs & Clark, 1992). They may be able to see the partner's face but not their array or actions (Boyle, Anderson & Newlands, 1994). Alternatively they may be talking over an intercom (Krauss et al., 1977), through a video-mediated dialogue system (Anderson & Boyle, 1994; Doherty-Sneddon, Anderson, O'Malley, Langton, Garrod, & Bruce, 1997), or they may be watching someone's actions via CCTV (Wilkes-Gibbs & Clark, 1992).

The main disadvantage associated with experimental studies, as noted above, is a lack of ecological validity. As with any laboratory-based paradigm, critics argue that this approach can only be generalised to a limited range of language settings (i.e., other task-based or goal-oriented situations; see Schober & Brennan, 2003). So insights into language production gleaned from referential communication tasks may not apply across the board to more spontaneous forms of conversation. Moreover, experimental approaches to dialogue tend to focus almost exclusively on reference, which is clearly only one aspect of language processing. A further concern is that analyses of speakers' behaviours tend to be restricted to a fairly coarse grain – spoken utterances, the end point of the production process. "As a consequence, the time course of effects of common ground in conversation, and its interaction with fine-grained aspects of other contributory linguistic processes, is poorly understood." (Hanna et al., 2003, p.44). Nonetheless, referential communication tasks have provided the data for some influential theoretical approaches to optimal design in conversation. Some of these approaches, and the empirical evidence used to support them, are outlined in 2.4 – 2.6.

2.4 Communication as Collaboration

Clark and colleagues' Collaborative framework (e.g., Clark & Wilkes-Gibbs, 1986) describes how interlocutors go about contributing to a conversation (Clark & Schaefer, 1989), and gives an account of the form of referring expressions produced during task-oriented dialogues (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). Under this view, speakers and addressees are equally responsible for making sure that the addressee understands what the speaker meant, at least well enough for their current (mutual) purposes. Interlocutors jointly participate in a process of *grounding* referring expressions (entering them into the common ground), and between them, they try to minimise the amount of *collaborative* effort they spend on this grounding process. For the speaker, minimising collaborative effort sometimes means prioritising the addressee's needs above their own. For example, they may have to put more effort into their initial formulation of a noun phrase than would be maximally efficient for themselves. They may be more likely to do this when there is an asymmetry between their own knowledge and that of the addressee, such as when a native speaker talks to a non-native speaker (Bortfeld & Brennan, 1997), or when a city local gives directions to someone who is clearly from out of town (Kingsbury, 1968; Isaacs & Clark, 1987; see Schober and Brennan, 2003).

Evidence for the Collaborative framework comes from an analysis of the kinds of referring expression produced in task-oriented dialogues. Consider this example from Clark and Wilkes-Gibbs's (1986) corpus of descriptions produced during a referential communication task. These utterances were used by a Director to describe the same ambiguous figure (see Figure 1 above) on six consecutive rounds of the game.

(1) *All right, the next one looks like a person who's ice-skating, except they're sticking two arms out in front*
(2) *Um, the next one's the person ice-skating that has two arms?*
(3) *The fourth one is the person ice-skating, with two arms*
(4) *The next one's the ice skater*
(5) *The fourth one's the ice skater*
(6) *The ice skater*

This example illustrates two well-established findings concerning referential communication. The first is that having used or heard a particular expression once, speakers will tend to use the same expression again. This tendency to repeat previously used expressions has been termed *entrainment* (Brennan & Clark, 1996; see also Garrod & Anderson, 1987; Pickering & Garrod, in press), a process which results in interlocutors settling on shared "description schemes" (Garrod & Anderson, 1987) and "conceptual pacts" (Brennan & Clark, 1996) for the objects that they are talking about. These conceptual pacts are essentially conceptual perspectives that have been ratified (perhaps following some initial modification) by both of the interlocutors.

The second robust finding illustrated by the example above is that once a conceptual perspective has been set up, speakers are free to express that perspective using shorter referring expressions (e.g., Brennan & Clark, 1996; Hupet et al., 1991; Krauss et al., 1977; Krauss & Weinheimer, 1964; 1966). In the example above, once the tangram figure has been conceptualised as an ice skater, the referring expressions used to refer to that figure get shorter and less elaborate, and the speaker starts to use a name with a definite article (*the ice skater*) rather than indefinite descriptions (*looks like a person who's ice-skating*).

Lexical entrainment suggests that language use is geared towards efficiency; the more often a speaker refers to an object, the shorter and less elaborate their references become (in terms of number of words, complexity, physical duration, and so on). From this point of view, conceptualisation and reference seem to be optimal for the speaker, because efficiency and fluency are maximised. But lexical entrainment also seems to be optimal for the addressee. Eye-movement data suggest that addressees identify an object more quickly the second time they hear a reference to it if the speaker re-uses the same referring expression that they used the first time (Barr & Keysar, 2002; Metzing & Brennan, 2003). In fact, if the speaker suddenly breaks a well-established conceptual pact, and uses a brand new expression to refer to a familiar object, then it takes longer for the addressee to look at the intended referent (Metzing & Brennan, 2003). Crucially, however, speakers can and do go against the general tendency towards efficiency in order to tailor what they say to the perceived needs of an addressee.

As we saw above, referring expressions tend to get shorter with repeated mention, but apparently only when the speaker can assume (based on the common ground they have acrrued) that the addressee will understand a shorter, less elaborate reference. For example, Wilkes-Gibbs and Clark (1992) had pairs of participants play a referential communication game using tangram figures similar to the one in Figure 1. By the end of the sixth round, Directors were using referring expressions that were shorter and more figurative than the ones they started out with, just as previous studies had suggested (Clark & Wilkes-Gibbs, 1986; Krauss & Weinheimer, 1964; 1966). After matching the set of tangrams six times with one partner, a second person took over the role of Matcher. The length and complexity of the Director's references in round 7 reflected what they knew about their new partner's knowledge of the tangrams. When the second Matcher had been able to both see and hear what was going on during the first six rounds (either because they had been sitting next to the Director, or because they had been watching the action over closed-circuit television), Directors produced shorter descriptions for the cards. They were also more likely to use definite references (*__the__ stair climber*) than when they changed to an entirely naïve partner (who had been waiting in a different room), or someone who had been sitting at the back of the room, able to hear but not see what was going on (cf. Horton & Gerrig, 2002).

Bard and Aylett (in press) found evidence for a similar effect in the Map Task corpus (Anderson et al., 1991; see 2.6.1 below). Referring expressions which named landmarks on a map tended to be syntactically simpler the second time they were produced by a particular speaker, but only when they were directed towards the same addressee on both occasions. When the second mention of the landmark introduced it to a new addressee, syntactic form was not simplified, compared to the first mention. Conversely, Brennan and Clark (1996) found that Directors tended to stick to conceptual pacts that they had established with a particular partner, even if that meant producing references that were overly specified with respect to the context. If a speaker had grounded an expression like *pennyloafer* with one partner, in the context of a game where there were several pictures of shoes to talk about, then they carried on using that expression, even when the array of picture cards changed so that there was only one shoe to describe. However, if a new, naïve Matcher was introduced, then the speaker was more likely to switch to a less specific expression like *shoe.*

Isaacs and Clark (1987) demonstrated a similar effect for situations in which speakers *inferred* common ground with an addressee. In this study, participants were either native New Yorkers (Experts) or non-New Yorkers (Novices), and the aim of the game was for the Director to help the Matcher arrange 16 postcards of New York City landmarks in a particular order. Neither the Director nor the Matcher were explicitly told whether their partner was a New York 'expert', yet speakers tailored their descriptions to the addressee's expertise from early on in the experimental session. Expert Directors tended to introduce a landmark by simply using its proper name (*the Citicorp Building*) when they were talking to a fellow Expert, even in the first round of the communication task, but when the Matcher was a Novice, they were more likely to add some extra description about what the postcard looked like (*just one huge building pointed at the top, Citicorp Center*). Isaacs and Clark suggested that Directors were tailoring their descriptions according to their beliefs about the Matcher's expertise. When Directors believed their addressee to be a Novice (with respect to New York City landmarks), they assumed that a proper name would not be enough to help the Matcher identify the correct referent, and therefore provided some extra information in the form of a description of what the postcard looked like (see also Kingsbury, 1968).

2.5 Utterance Planning is Egocentric

Section 2.4 introduced an account of language processing which assumes that utterances are designed with the addressee's needs in mind. At the other end of the theoretical spectrum lie accounts which assume that production optimises efficiency for the speaker, at least during initial utterance planning. In effect such accounts suggest that language production is inherently egocentric. Two particular models will be discussed here: Horton and Keysar's (1996) Monitoring and Adjustment account, which builds on the work of Brown and Dell (1987; Dell & Brown, 1991), and Pickering and Garrod's (in press) Interactive Alignment approach, which develops ideas originally proposed by Garrod and colleagues (Garrod & Anderson, 1987; Garrod & Clark, 1993; Garrod & Doherty, 1994). What Monitoring and Adjustment and

Interactive Alignment have in common is that they both assume no central role for (explicit) listener modelling in ordinary, day-to-day utterance planning.

2.5.1 Monitoring & Adjustment

Brown and Dell (1987; Dell & Brown, 1991) and Horton and Keysar (1996) proposed a production model in which information about the addressee's perspective does not constrain the initial planning of utterances. According to this approach, speakers initially make use of information available to themselves even when that information is not part of the common ground shared with the current addressee. However, once planned, an utterance can be monitored for "pragmatic adequacy" (Levelt, 1983). If it violates common ground, then it can be revised, but this monitoring and revision work is labour-intensive, and the monitoring process may have to compete with other processes for limited cognitive resources (Bard & Aylett, in press). Crucially, both Brown and Dell and the Monitoring and Adjustment models see knowledge about an addressee's perspective as high-level, "top-down" information, or meta-linguistic knowledge. Although this knowledge might help to avoid misunderstandings, using it is assumed to be computationally expensive.

Moreover, beliefs about an interlocutor's mental state may be uncertain (Hanna et al., 2003), and since interlocutors are often co-present during conversation, information available to the speaker is typically also available to the addressee. Consequently, an utterance based on the speaker's perspective will generally end up being perfectly adequate for the addressee even if common ground has not been used in its design. As Horton and Keysar (1996, p.112) put it,

> *If the system works like the model suggests, then there is no audience design in utterance planning. Some utterances might involve post hoc corrections, and some might involve no active audience design at all. Yet on the whole, utterances might still be appropriate for the audience, and they might look like they were pre-designed for that audience.*

Of course, using the speaker's own perspective as a proxy for the addressee is a something of a risky strategy (cf. Carletta & Mellish, 1996), because in situations where the addressee's knowledge is *not* the same as the speaker's it could potentially lead to confusion. But for the most part, taking the risk that the addressee's perspective is sufficiently similar to the speaker's pays off. Besides which, if the speaker and addressee *are* co-present, then the addressee can always ask for clarification, so all is not lost (e.g., Krauss & Fussell, 1991).

Brown and Dell originally proposed their modular version of the model on the basis of evidence from two experiments involving a story-telling task (Brown & Dell, 1987; see also Dell & Brown, 1991). These experiments were designed to investigate whether speakers syntactically "package" information according to an addressee's needs (Lockridge & Brennan, 2002). Participants silently read 20 three-sentence stories and retold them in their own words. The first sentence in each story mentioned an action that implied the use of an instrument. For example, one story began: *The robber hid behind the door and when the man entered the kitchen he stabbed him in the back.* *Stabbed* is the critical action here, and the second sentence of the story explicitly mentioned the instrument used in that critical action (*he wiped the blood off the knife/ice pick*). Sometimes the instrument was "typical" for the action, and sometimes it was "atypical". In this example, a knife is typically used for stabbing, whereas an ice pick is not. In the first experiment, speakers explicitly mentioned atypical instruments more often than typical instruments when they retold the stories in their own words. This happens to be helpful for listeners, because atypical instruments are harder to infer. But do speakers mention atypical instruments more often *because* this is helpful for a listener, or simply because atypical instruments are more salient or newsworthy for themselves? A second experiment suggested that the typicality effect is based on the speaker's perspective rather than the addressee's.

In Experiment 2, participants retold the stories to a confederate addressee, who would be tested later on what they could remember (or so participants believed). Speakers were given a picture that was either informative about the instrument used in the critical action (the knife or ice pick could be seen in the robber's hand), or was not informative about the instrument (the robber's hand was hidden behind his back). Half of the speakers knew that the listener could see a copy of the same picture, while half knew that their listener could not. Presumably, mentioning an atypical instrument would be most helpful when the addressee could not otherwise infer it. So, if production is geared towards the addressee's needs, then speakers ought to have mentioned atypical instruments more often when the addressee could not see an informative picture. However, whether the addressee could see an informative picture or not seemed to have no impact on instrument mention. Speakers were no more likely to mention atypical instruments early in a sentence (i.e., in the same clause as the critical action) when the addressee did not have access to an informative picture than when the addressee could see an informative picture. However, they did mention atypical instruments in a separate clause after the verb more often when the addressee had no picture at all. Brown and Dell suggested that this was a coarse adaptation to the listener's needs, and that it ocurred at a relatively late stage in production. They proposed that when the first clause (which contains the critical action, *stab*) is monitored for pragmatic acceptability, the system realises that the addressee cannot infer the instrument used for the action, because they do not have access to an informative picture. A second clause which explicitly mentions the atypical instrument is then planned, as an addition.

A recent study adds a cautionary note to Brown and Dell's claim that speakers only take an addressee's perspective into account during a late monitoring stage of production. Lockridge and Brennan (2002) replicated Brown and Dell's Experiment 2, but rather than using a confederate addressee, they had speakers re-tell the stories to naïve partners who really were tested on their memory for what they had heard. Presumably a confederate who has heard the same information many times from lots of different speakers provides an unusual kind of feedback, and speakers might be more likely to show evidence of optimal design (with respect to instrument mention) when they talk to a naïve addressee with real communication needs. Like Brown and Dell, Lockridge and Brennan found that speakers explicitly mentioned atypical instruments more often than typical instruments. However, contrary to Brown and Dell's findings, speakers *did* seem to be taking the addressee's perspective into account. They were somewhat more likely to mention the instrument early (i.e., in the same clause as the verb) more often than when it was atypical and the addressee had no picture, compared to conditions where the instrument was typical, or where the addressee could infer it from an informative picture.

Horton and Keysar (1996) developed a version of Brown and Dell's model, which they called the Monitoring and Adjustment hypothesis. Evidence for Monitoring and Adjustment comes from a picture description task in which naïve participants played the role of speaker and a confederate played the role of addressee. The two players sat in front of a computer with a barrier between them such that the speaker could only see one half of the screen and the addressee could only see the other half. On each trial, the speaker saw two objects on their side of the screen, one of which (the "target" object, or the thing to be described by the speaker) moved across the barrier to the addressee's side. Sometimes the target object changed in some way as it moved from one side to the other. The speaker's job was to give descriptions such that the addressee could decide whether the target object had changed. The second object was stationary, and was designed to be a "context" object for the moving target. On experimental trials, this context object shared one or more features with the moving object. For example, a target circle was paired with a smaller or larger context circle. Half of the speakers were told that their addressee could see the context object on each trial (*shared context* condition). For these participants, the context object was always shared information with the addressee. The other speakers were told that the addressee *never* saw the context object, which was "privileged" information for the speaker (*privileged context* condition).

Horton and Keysar were interested in how often speakers used "context-relevant" adjectives like "large" and "small". For example, if a target circle was paired with a smaller context circle, would speakers describe it as *a large circle*? Most interestingly, would they call it *a large circle* even in the privileged context condition, when their addressee has no context circle to compare it against? If descriptions are designed with the addressee's perspective in mind, then we would expect speakers in the privileged context condition to use fewer context-relevant adjectives, because such adjectives would be irrelevant information for the addressee. Essentially, Horton and Keysar were interested in whether speakers would try to avoid being over-specific in the privileged context condition, by *not* including adjectives that they were likely to use in the shared context condition.

Half of the speakers were put under time pressure to start giving their descriptions within 1.5 seconds of the objects appearing on their screen. This manipulation was included to test the claim that speakers only consult a model of their listener during a late, monitoring stage of production. If this is true, and if monitoring is a labour-intensive process (as Brown and Dell suggested), then speakers under time pressure should show less evidence of optimal design than speakers who can start their descriptions at leisure. In the speeded condition there is less opportunity for the monitoring process to edit out any infelicitous (over-specific) adjectives, so time pressure should reveal the production system's "true colours". If utterances are not planned with the addressee's needs in mind, this should be most obvious when speakers have to start speaking quickly.

In the unspeeded condition, speakers did indeed use fewer context-relevant adjectives when the context object was privileged information for themselves. However, when speakers were under pressure to produce their descriptions quickly, they used context-relevant adjectives equally often whether the addressee could see the context object or not. Horton and Keysar interpreted this finding as strong evidence for Monitoring and Adjustment. The fact that speakers use context-relevant adjectives even when they convey irrelevant information from the addressee's perspective, they suggested, shows that the initial design of the utterance includes those adjectives, and that the monitoring process edits these out, but only when time and cognitive resources allow it.

Horton and Keysar concluded that putting speakers under time pressure reveals that initial utterance plans are not designed according to common ground. They made the assumption that time pressure reduces the production system's capacity to monitor for common ground violations, but does not affect the initial plan. Initial planning, they claim, is resource-free, while monitoring and adjusting utterances is thought to require additional cognitive resources. But as Polichak and Gerrig (1998) have pointed out, time pressure could just as well affect a speaker's ability to assess common ground in the first place. Clark and Marshall's (1981) original proposal was that objects are potentially in common ground if they are, have been, or could be co-present between speaker and addressee. Presumably deciding whether some object is, has been, or potentially could be co-present requires some kind of mental computation, and as such, this process could be disrupted by cognitive load. If this is the case, then Horton and Keysar's findings do not necessarily show that initial utterance planning is insensitive to knowledge about the addressee's perspective. In fact, Clark and Wilkes-Gibbs (1986) explicitly suggested that time pressure may have consequences for the kinds of references speakers produce; it is not clear that Horton and Keysar's results do anything beyond empirically demonstrating that point.

As a final point, it is worth noting that Horton and Keysar (like Brown and Dell, 1987) had speakers talk to a confederate rather than a second naïve participant. As Lockridge and Brennan (2002) showed, confederate listeners may not behave like 'genuine' addressees, and this could have consequences for whether or not speakers appear to tailor what they say to the addressee's perspective. The fact that Lockridge and Brennan's speakers tended to mention instruments 'early' (i.e., in the same clause as the critical verb) more often when the addressee could not infer them also suggests that beliefs about an addressee may not be relegated to only a late role in production. A model of the addressee's perspective may in fact be able to guide initial planning of an utterance, depending on how quickly information about the addressee is available to the production system. This point is taken up in 2.6.2. But for now we turn to another "speaker-optimal" model of production: the Interactive Alignment account.

2.5.2 Output-Input co-ordination & the Interactive Alignment account

Pickering and Garrod's *Interactive Alignment* theory of dialogue (in press) suggests that speakers tend to produce utterances which are easy for an addressee to understand simply because interlocutors converge on similar ways of talking about things during conversation. Under this view, speakers and addressees aim to align their "situation models" in order to communicate successfully, and saying things which look to be addressee-oriented is basically a convenient by-product of this alignment process. The Interactive Alignment account is developed from ideas originally formulated by Garrod and colleagues (Garrod & Anderson, 1987; Garrod & Clark, 1993; Garrod & Doherty, 1994), based on evidence from a referential communication task known as the maze game.

Garrod and Anderson (1987) had pairs of interlocutors play a collaborative computer game that involved moving around in a maze-like grid, and found that individual pairs hit upon different ways to conceptualise their spatial positions in the maze. Some pairs used "path" or "line" descriptions to refer to location, others took a co-ordinate approach, and others described parts of the maze figuratively. Garrod and Anderson called these different ways of referring to locations "shared description schemes", and suggested that these reflected different semantic conceptualisations of the maze.

Pairs of interlocutors seemed to converge on the same conceptualisation, over the course of the dialogue. For example, one player producing a path description was a good predictor of their partner using a path description in their next utterance. This could be taken as further evidence that speakers and addressees work together to communicate efficiently and avoid misunderstandings (c.f. Clark and colleagues' Collaborative framework). However, Garrod and Anderson suggested that using a shared description scheme is a much less deliberate process, driven by an automatic *output-input co-ordination mechanism* for language. Rather than collaborating to find mutually acceptable referring expressions, interlocutors are thought to produce aligned descriptions because linguistic material coming into the language processing system is 'recycled' in generating new utterances. In other words, referring expressions processed during comprehension are re-used in designing utterances for production, probably because this is a cognitively efficient strategy. Under this view, optimal design is simply a convenient side-effect of automatic output-input co-ordination, and overriding these automatic processes takes strategic effort on the part of the speaker (Garrod & Clark, 1993).

The notion of an output-input coordination mechanism is developed in the Interactive Alignment account (Pickering & Garrod, in press), which aims to give a broad account of language use in dialogue. Like Brown and Dell (1987), Pickering and Garrod suggest that what look like adaptations to the addressee's perspective usually have nothing to do with the speaker explicitly consulting a model of their listener. Instead, they claim that speakers tend to produce utterances that are easy for their addressee to understand because participants in a dialogue are *aligned* at various levels of mental representation (lexical, semantic, syntactic, and so on). How does this alignment come about? Essentially through Output-Input co-ordination.

On the assumption that there is parity between representations used in production and comprehension (although obviously not between comprehension and production *procedures*; e.g., Pickering & Branigan, 1999), then representations built during comprehension can be re-used for subsequent production. This is an optimally efficient strategy in the sense that it allows the production system to take short-cuts in formulating new utterances. Rather than generating each syntactic structure from scratch, the production process can simply recycle the structure of a previously encountered utterance wholesale. If the language processing system *does* capitalise on the parity between comprehension and production in this way, then alignment falls out of the process naturally. Speakers who re-use the same words and syntactic structures that they have heard their interlocutors use will automatically end up with the same (or similar) lexical and syntactic representations. This "local" alignment percolates through the system and results in increased "global" alignment between interlocutors' *situation models* (Johnson-Laird, 1983; Sanford & Garrod, 1981; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998), or mental representations of the topic under discussion. Ultimately, successful communication is dependent on interlocutors having aligned situation models, but crucially, the Interactive Alignment account claims that this comes about without speakers having to do any explicit modelling of what their addressee does or doesn't know. As Pickering and Garrod (in press, 2.2) put it:

> … *hearing an utterance that activates a particular aspect of a situation model will make it more likely that the person will use an utterance consistent with that aspect of the model. This process is essentially resource-free and automatic.*

Through alignment, then, interlocutors build up an *implicit common ground.* If we think of interlocutors' situation models as the circles in a Venn diagram, then the implicit common ground is the overlap between the two circles. At the beginning of a conversation, this overlap may not be very large (because participants come to the conversation with different situation models), but as the interlocutors start to align in terms of lexical choice, syntax, and so on, the amount of overlap increases – their situation models converge. So, the implicit common ground between participants in a dialogue grows via automatic alignment at other levels of representation. As a consequence, speakers may say things that appear to be tailored to their addressee. But under Pickering and Garrod's view, this has nothing to do with the speaker taking the addressee's needs or perspective into account. Instead, it is simply a convenient by-product of alignment, which is automatic and makes no particular demands on processing resources.

Pickering & Garrod (in press) contrast implicit common ground with "full common ground", or common ground in a Brown and Dell (1987) sense, i.e., an explicit model of the addressee's perspective. Establishing full common ground, they claim, is "… a specialised and non-automatic process that is used primarily in times of difficulty (when radical misalignment becomes apparent)." (Pickering & Garrod, in press, 4.1). As evidence for this view, Pickering and Garrod cite Horton and Keysar (1996) and Brown and Dell's (1987) findings from production (see 2.5.1), which suggest that utterance planning is not guided by a model of the listener in the first instance. Additionally, they point out that adults (like children) sometimes use definite reference infelicitously, using definite determiners to mark information that is New for an addressee as Given (Anderson & Boyle, 1994).

To show that interlocutors do not normally consult full common ground, Pickering and Garrod also cite comprehension studies which suggest that addressees do not adhere to the Principle of Optimal Design when they interpret referring expressions. For example, Keysar, Barr, Balin, and Paek (1998) found that addressees do not only consider objects in common ground as potential referents for a speaker's description. In fact, in the first moments after they hear the description, they are just as likely to look at a "secret object" that is only available to themselves (i.e., a potential referent in privileged ground) even when they know that the speaker is unaware of it (see also Keysar et al., 2000, and Keysar et al., 2003). This would seem to be strong evidence against optimal design in comprehension, but Keysar and colleagues have been criticised for setting up the experimental situation such that privileged information was more salient (Keysar et al., 2000) or more recently encountered (Keysar et al., 1998) than information in common ground (e.g., Hanna et al., 2003; Metzing & Brennan, 2003; Schober & Brennan, 2003). When such biases are removed, the evidence suggests that addressees do make rapid use of (full) common ground, or partner-specific information (Brennan & Metzing, in press). As we will see in 2.6.2 below, addressees may briefly look at objects hidden from the speaker, but they very quickly rule them out as potential referents (Hanna et al., 2003; see also Nadig & Sedivy, 2002). Moreover, addressees are momentarily confused when they hear a familiar speaker break a previously-established conceptual pact (Metzing & Brennan, 2003).

Of course, it's possible that addressees are using full common ground in an unusual way in the studies cited above (as Pickering and Garrod, in press, suggest). After all, the experimental situation is set up such that there is little (natural) interaction between speaker and addressee. Since there is no interaction, there is no opportunity for implicit common ground to develop, and addressees might use full common ground in a way that they wouldn't normally use it in everyday conversation. However, contrary to Pickering and Garrod's suggestion, it seems unlikely that consulting full common ground in these circumstances is entirely strategic. The eye-movement effects are rapid, and they occur even when the addressee is looking at a relatively complex visual display, where gaze is probably not under strategic control (e.g., Metzing & Brennan, 2003; see Brennan & Metzing, in press, for discussion).

2.6 Production is subject to ordinary cognitive constraints

So far I have introduced two classes of model which make very different claims about the role of the addressee (or rather, the speaker's beliefs about their addressee) in language production. Collaborative models (e.g., Clark & Wilkes-Gibbs, 1986) and "speaker-centric" accounts (e.g., Brown & Dell, 1987) represent two extreme points on a continuum of theoretical claims. In this section I introduce models that lie somewhere between the two. These models do not altogether rule out optimal design as a potential influence on initial planning, as the Monitoring and Adjustment hypothesis does. Instead, they suggest that speakers attend to optimal design as and when time and processing resources allow (e.g., Bard et al., 2000), or to the extent that common ground is useful and important for successful communication (e.g., Hanna et al., 2003).

2.6.1 The Dual Process hypothesis

The "Dual Process" hypothesis (Bard et al., 2000; Bard & Aylett, in press) suggests that there are (at least) two families of process at work in language production. The first family includes automatic priming processes that "... operate during any attempts to produce spoken utterances..." (Bard et al., 2000, p.17). The second family includes processes that involve an inference- or decision-making component, such as deciding what an addressee does or does not know. Like Pickering and Garrod (in press), Bard et al. (2000) see priming processes as fast, automatic, and essentially requiring little in the way of cognitive resources. In contrast, processes that require inferences to be drawn, or decisions to be made, are assumed to carry a greater cognitive burden than other aspects of production, and as such they are slower and more effortful.

Essentially, the Dual Process hypothesis concurs with the Interactive Alignment account as to why speakers sometimes seem to behave egocentrically when they talk. Firstly, complex processes like modelling the listener may be "... simply too slow to precede every attempt at speech production. They are therefore likely to be too slow to make their effects felt on a word-to-word basis in running speech." (Bard et al., 2000, p.17). These slower processes also have to compete for cognitive resources with other processes and activities associated with dialogue (or task) management. If there are not enough resources to go around, then optimal design suffers at the expense of other, more pressing mental activities. Secondly, most conversations take place between interlocutors who are either physically co-present, or can rely on other co-presence heuristics, such as belonging to the same community (cf. Clark & Marshall, 1981). So for most everyday purposes, speakers do not need to go to great lengths to model the listener – their own knowledge and perspective functions perfectly well as a proxy model of the addressee (see also Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996; Pickering & Garrod, in press).

Evidence for the Dual Process hypothesis comes from a study which compared different kinds of evidence for optimal design. In a series of analyses, Bard and Aylett (in press) studied the form and duration of referring expressions from the Map Task corpus (Anderson et al., 1991). In the Map Task, one participant (the Instruction Giver) helped their partner (the Instruction Follower) to mark a route on a map of an imaginary location. Each player had a map showing named landmarks. Some of these landmarks were 'shared' (i.e., they appeared identically on both the Instruction Giver's and the Instruction Follower's maps). Other landmarks were 'non-shared'; they mismatched on the two maps in some way. Speaker's first and subsequent references to a particular landmark were coded for accessibility or syntactic simplification on a scale from 0 (most accessible, or least simplified, e.g., indefinite NPs like *a mountain*) to 3 (least accessible, or most simplified, e.g., pronouns like *it*) (cf. Ariel, 1990). The physical durations of these referring expressions were compared to references produced by the same speakers in citation form.

What predictions does the Dual Process account make for these data? Bard and Aylett assume that speakers must "... consult a listener model at least once per intonational phrase" to be able to appropriately adapt utterances to the addressee's perspective. Dual Process hypothesis predicts that fast-cycling production processes which deal with small planning units should be more likely to be affected by fast, automatic influences (e.g., priming) and less likely to be affected by slow, computationally-demanding influences (like listener modelling). Conversely, production processes dealing with larger planning units might be more likely to show an influence of the addressee's perspective. Bard and Aylett's analyses of landmark references from the Map Task supported this hypothesis.

Articulation – a fast-cycling process which deals with small units – showed effects of repeated mention (priming) but not optimal design. On the other hand, referential form – a slower-cycling process that generates larger units – *did* seem to be sensitive to optimal design. Referential forms were *less* reduced at second mention if the landmark was New information for the current addressee. That is, speakers were less likely to use a syntactically simplified form (like a pronoun) if the speaker directed his first and second mentions of a particular referring expression to different partners, during two different conversations (compared to a control condition with the same addressee hearing both first and second mentions). However, the physical duration of a reference was significantly shorter at second mention even if the landmark was New for the addressee. This replicates a similar finding by Bard et al. (2000), who found that second mentions of landmark names were less *intelligible* (identified less accurately in isolation by naïve listeners) than first mentions, even when the second mention conveyed information that was New for the addressee for whom the reference was produced. So while referential form seems to be sensitive to the addressee's needs, articulation does not.

In contrast with Bard and colleagues' findings, and unpublished study by Gregory, Healy and Jurafsky (submitted) suggests that articulation might show some gross sensitivity to addressee needs. In Gregory et al.'s study, referring expressions (noun phrases and proper names) showed *less* phonological reduction at second mention when they conveyed New information for the addressee than when they conveyed Given information. Nonetheless, references were always shorter at second mention than at first mention. If shorter references are less intelligible, then articulation is certainly not entirely determined by considerations of the addressee (Bard & Aylett, in press).

2.6.2 A Constraint-based approach to optimal design

In 2.5.2 above I briefly introduced some evidence which suggests that addressees do not always take the speaker's perspective into account during comprehension. For example, Keysar et al. (2000; 2003) demonstrated that addressees consider a hidden object as a potential referent for a referring expression, even when they know that the speaker cannot see it (Keysar et al., 1998), or has been mis-led about its identity (Keysar et al., 2003). However, other studies suggest that addressees quickly rule out objects in privileged ground as potential referents.

For example, Hanna et al. (2003) tracked participants' eye movements as they listened to instructions for moving coloured shapes around on a board. On critical trials, they heard instructions that were ambiguous with respect to the objects in front of them. For example, the addressee might hear *Now put the blue triangle on <u>the red one</u>*, when there were two red triangles on the board. In the "common ground" condition, addressees knew that both of the red triangles were visible to the speaker. In the "privileged ground" condition, however, they believed that one of the red triangles was a "secret shape" that the speaker (actually a confederate) was unaware of. In effect the secret shape was privileged information for the addressee, in the same way that the context objects in Horton and Keysar's production study were privileged information for the speaker. The interesting question, then, is whether addressees would consider the secret shape as a potential referent for the expression *the red one* (i.e., whether they would fail to follow the Principle of Optimal Design), or whether they would simply ignore it, on the basis that it is not in common ground.

The eye movement data show that addressees did not entirely rule out the secret shape as a potential referent for the critical referring expression *the red one*, even though they believed that the speaker did not know it existed. This finding, along with those of Keysar and colleagues, seems to rule out the most extreme version of Optimal Design as it applies to comprehension. Addressees clearly do not *always* search only in common ground for a referent, at least not in the earliest moments of processing. Nonetheless, addressees in Hanna et al.'s study were *quicker* to look at a target object (the speaker's intended referent) when the competitor object was a secret shape than when it was in common ground (Nadig and Sedivy, 2002, report similar findings in 5-6 year old children). So, addressees were quickly ruling out the secret shape as a potential referent, and there was no evidence for an initial stage of processing that was completely oblivious to common ground. This challenges the most extreme version of the Monitoring and Adjustment model as interpreted by Keysar and colleagues.

A further eye-tracking experiment suggests that reference resolution is constrained not only by what an addressee knows about the speaker's perspective, but also by dynamic, task-relevant considerations, such as what the speaker is doing at the moment of referring to some object in the visual context. Hanna and Tanenhaus (2004) had participants play the role of "helper" to a "cook" (actually a confederate) in tasks like making bread or cupcakes. The confederate was scripted to produce requests like *Could you put the cake mix next to the mixing bowl?* Critical referring expressions (e.g., *the cake mix*) could potentially pick out either just one object (a cake mix in the helper's area, which was out of the cook's reach) or two similar objects, one in the helper's area and one in the cook's area (e.g., two cake mixes). At the critical moment of referring, either the cook's hands were empty, or she was busy doing something else, and her hands were full. Eye movement data demonstrated that helpers interpreted the critical references as referring to objects in their own area when the cook's hands were empty, but also immediately considered objects in the cook's area when her hands were full. This suggests that comprehension can be sensitive, even from the earliest moments of processing, to the pragmatic non-linguistic considerations associated with actions and abilities (cf. Chambers, Tanenhaus, Eberhard, Filip,

& Carlson, 2002). Helpers seemed to assume that the cook must be talking about something in their area if her hands were empty (*why would she ask me to move something that she can reach herself?*) but that she was referring to an object in her own area when her hands were full.

On the basis of this evidence, Hanna et al. (2003) suggest that common ground can best be understood as a probabilistic constraint on language processing, just as other language-related variables are considered to be probabilistic constraints (e.g., MacDonald, 1993; 1994; Tanenhaus & Trueswell, 1995). "Constraint-based models at a general level propose that alternative interpretations with some degree of likelihood are evaluated in parallel, based on the simultaneous and continuous integration of probabilistic evidence provided by multiple constraints." (Hanna et al., 2003, p.50). Under a constraint-based approach, optimal design is not an all-or-none phenomenon. Speakers and addressees should consider common ground as soon as information about their interlocutor becomes available to the language processing system, but only to the extent that it is important or relevant for communication. In comprehension, there is some evidence that this may be the case. For example, addressees' eye movements do not show the usual cohort effects (see Sedivy, Tanenhaus, Chambers, & Carlson, 1999) when they are listening to a speaker whom they believe to be "unreliable" (Grodner & Sedivy, 2003). This makes sense: for the addressee, there is no point assuming that your interlocutor has designed their utterances optimally if you know that they are unlikely to have followed the normal "rules" of interaction.

In terms of production, a constraint-based approach leads to some interesting predictions about where we might see evidence of optimal design. We would expect speakers to pay more attention to their addressee's perspective when the communicative consequences of doing so (or *not* doing so) are important (Branigan, McLean, & Reeve, 2003; see Chapter 4). For example, tailoring the content of an utterance (in terms of lexical choice and conceptualisation) is probably more important than tailoring syntax and grammatical form, in most circumstances. After all, the choices a speaker makes about how to conceptualise a referent, and the words they use in expressing that conceptualisation, may well affect the likelihood of the addressee forming an aligned mental representation (to use Pickering and Garrod's terms). But grammatical choices often have no real consequence for the meaning of an utterance (although clearly this is not so in the case of globally ambiguous syntactic structures; see Chapter 6). For example, speakers are free to express a ditransitive relationship using either a prepositional object structure (*the pirate showing the book to the monk*) or a double object structure (*the pirate showing the monk the book*) without changing the meaning of their message – essentially both constructions will be interpreted in the same way (Branigan, Pickering, & Cleland, 2000). So tailoring grammatical choices for the addressee might turn out to be more trouble than it is worth.

This asymmetry between semantics and syntax might account for the general picture of results in the optimal design literature. Speakers are typically shown to behave "helpfully" in terms of lexical and conceptual choice (e.g., Bortfeld & Brennan, 1997; Isaacs & Clark, 1987; Wilkes-Gibbs & Clark, 1992), but are argued to make egocentric choices with respect to grammar and syntax (Branigan, McLean, & Reeve, 2003; Brown & Dell, 1987; Dell & Brown, 1991; Ferreira & Dell, 2000). Perhaps we would see more evidence of optimal design at the level of grammatical encoding if there were potential consequences for the addressee? Almost all of the experiments reported in the remainder of the thesis address this issue (see Chapters 4-6). However, first we turn to optimal design at the level of conceptualisation, and the issue of whether speakers are less likely to take their addressee's perspective when they are under increased cognitive load.

CHAPTER 3

Optimal Design under Cognitive Load

3.0 Chapter Overview

This chapter begins with a summary of the major findings from referential communication experiments. For example, speakers gradually produce shorter and less elaborate descriptions for a referent over time, but only when they can be fairly sure that their partner will understand a shortened reference (e.g., Wilkes-Gibbs & Clark, 1992). This is often taken as evidence for optimal design in language production. Much of this evidence comes from 'idealised' communication situations: one speaker and one addressee, who can interact freely for as long as they need to satisfy themselves that they are both talking about the same thing. But in everyday life, communication often takes place under less idealised conditions. Experiment 1 investigated whether speakers are able to attend to optimal design when they are put under cognitive load.

3.1 Referential Communication: Standard Findings

3.1.1 Referring expressions get shorter over time

In Chapter 2, I outlined the referential communication paradigm developed by Krauss and his colleagues (Glucksberg, Krauss, & Higgins, 1975; Glucksberg, Krauss, & Weisberg, 1966; Krauss, Garlock, Bricker & McMahon, 1977; Krauss & Glucksberg, 1969; 1977; Krauss, Vivekananthan & Weinheimer, 1968; Krauss & Weinheimer, 1964; 1966; 1967). In referential communication games, pairs or groups of participants work together to solve a problem. Typically the participants cannot see each other, and they have a goal that can only be achieved by verbally collaborating on the task. For example, one player (the Director) describes a set of picture cards so that their partner (the Matcher) can arrange their cards in some pre-specified order (e.g., Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992; Hupet, Seron, & Chantraine, 1991; Wilkes-Gibbs & Clark, 1992). Picture cards used for the matching task are chosen so that speakers will have to produce complex noun phrases to describe them, usually for one of two reasons. Sometimes the objects in the pictures are unusual or ambiguous shapes, and cannot easily be labelled (i.e., they have *low codability*; Lachman, 1973; Hupet et al., 1991). Alternatively, a simple noun phrase might not unambiguously pick out one referent (the objects in the referent array have *low discriminability*; Hupet et al., 1991). For example, in the context of 15 different kinds of chair, speakers would have to be specific about what *kind* of chair they were talking about if they wanted to get an addressee to select the intended picture. A simple noun phrase like *the chair* would be inadequate (Bortfeld & Brennan, 1997).

Early studies showed that referring expressions produced during referential communication tasks typically get shorter (in terms of number of words) with repeated mention (Krauss & Weinheimer, 1964; 1966). A Director's first reference to a picture card usually contains many words, and may include "nonstandard" expressions, such as "dummy" noun phrases like *whatchamacallit* and *thingamabob* that do not name a particular referent (Clark & Wilkes-Gibbs, 1986). Later references to the same card, though, are typically shorter in terms of both number of words (Clark & Wilkes-Gibbs, 1986) and the physical duration of those words (e.g., Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Fowler & Housum, 1987; McAllister, Potts, Mason, & Marchant, 1994; Samuel & Troicki, 1998). In other words, repeated references show *referential reduction* and *durational reduction.* In this chapter, I deal with referential reduction only, not the physical length of the words produced, which is outside the scope of the thesis (see Bard et al., 2000, for a review). Later referring expressions are also more likely to be marked as definite (rather than indefinite), and simple names (*the ice skater*) are more likely to be used than complex descriptions (*looks like a person who's ice skating, except they're sticking two arms out in front*; Clark & Wilkes-Gibbs, 1986). These are very robust findings in the referential communication literature, and have been replicated in many studies (e.g., Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Gregory, Healy, & Jurafsky, submitted; Horton & Gerrig, 2002; Hupet & Chantraine, 1992; Hupet et al., 1991; Isaacs & Clark, 1987; Krauss & Weinheimer, 1967; Markman & Makin, 1998; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992; see Krauss & Fussell, 1996, and Schober & Brennan, 2003, for reviews).

Experimental studies also suggest that the process of referring to novel objects is normally a collaborative one. It is not the case that the speaker produces descriptions as though they were writing things down for someone else to read later, and simply hopes that the addressee will understand them. Instead, speakers and addressees seem to work together to *ground* particular referring expressions, where grounding means making sure that each reference has been mutually understood well enough for current purposes (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). By working together to ground references, interlocutors tend to settle on shared "conceptual pacts" for the objects they are describing (Brennan & Clark, 1996), or flexible temporary agreements about how to conceptualise objects out in the world. Interlocutors mark having reached a conceptual pact by reusing the same or a similar term to refer to an object (Schober & Brennan, 2003) – a process known as *entrainment* (Brennan & Clark, 1996; Garrod & Anderson, 1987). Entrainment presumably comes about because if a referring expression (say *the ice skater*) has been used successfully once, the speaker can be pretty sure that it will be successful the next time around. Expressions that become entrained upon could be called *conversational precedents* (Horton & Gerrig, submitted), and they seem to make comprehension easier. When addressees hear the same expression twice, they are quicker to identify its referent the second time around whether the expression is produced by the same or a different speaker (Barr & Keysar, 2002; Metzing & Brennan, 2003). But if a "familiar" speaker suddenly breaks a conceptual pact without warning, and uses a completely new expression in place of an established conversational precedent, then comprehension is disrupted, and addressees take appreciably longer to work out what is being referred to (Metzing & Brennan, 2003).

3.1.2 Constraints on referential reduction

Interestingly, the standard findings mentioned above only hold when both interlocutors are free to make timely contributions to the dialogue. Anything that reduces the quality of the feedback speakers receive seems to affect the efficiency of the conversation (e.g., Bavelas, Coates, & Johnson, 2000; Pasupathi, Stallworth, & Murdoch, 1998; Tatar, 1997). For example, if a speaker gets limited feedback from their addressee, either because they cannot see (Boyle, Anderson, & Newlands, 1994; Krauss et al., 1977) or hear (Krauss & Weinheimer, 1966) them, then references are not shortened to anything like the same extent (see also Kraut, Lewis, & Swezey, 1982, and Schober, 1993). Similarly, if feedback is available but artificially delayed (Krauss & Bricker, 1967), then communication is less successful and efficient.

Studies comparing addressees with other kinds of listener underline the collaborative nature of reference. Overhearers and eavesdroppers who do not have an opportunity to interact with the speaker do less well at matching tasks than addressees who can give feedback (Kraut et al., 1982; Schober & Clark, 1989). Speakers seem to be sensitive to this, because if they have not grounded a particular conceptualisation with the current addressee, they tend to use longer, more descriptive expressions in the first instance (see Wilkes-Gibbs & Clark, 1992). They avoid references that are syntactically reduced, like pronouns (Bard & Aylett, in press), although they may not manage to avoid pronouncing repeated expressions with less clarity (Bard et al., 2000; but see Gregory, Healy, & Jurafsky, submitted, for discussion). In spatial perspective-taking tasks, speakers are also more likely to take the addressee's view-point if the addressee is not present, and therefore cannot give feedback (Schober, 1993).

All of the findings discussed so far relate to situations where one speaker communicates with one addressee, under 'idealised' conditions. The Director has as much time as they like to decide what to say, and they are usually free to converse with the Matcher as much as they need to (although some experimental situations manipulate feedback, as discussed above; e.g., Bavelas et al., 2000; Krauss & Weinheimer, 1966, and Kraut et al., 1982). Given that there is only one addressee, it should be reasonably easy for the speaker to keep in mind what is mutually known, what conceptual pacts have been established, and so on. But outside the psychology laboratory, communication often does not happen under these kinds of circumstances. Speakers (and addressees) may be under some kind of cognitive 'load', such as talking at the same time as doing something else (driving a car, mowing the lawn, or even carrying out another linguistic task, such as writing an email). The addressee may not be physically co-present with the speaker, as in telephone or computer-mediated conversations. Moreover, conversations are often multi-party, involving several listeners at any one moment, one or more of whom may be an overhearer or a bystander rather than an intended recipient of the speaker's message (Bell, 1984; Clark & Carlson, 1982). All of these things make the speaker's job more difficult if they are trying to design utterances optimally.

If the speaker is engaged in some other task at the same time as participating in the conversation, they may have fewer cognitive resources to recruit for the task of optimal design (e.g., Bard et al., 2000; Bard & Aylett, in press; Schober & Brennan, 2003). If the *addressee* is engaged in some other task, the quality of the feedback the speaker gets may be reduced, and reduced feedback, as we have seen, makes communication less successful (e.g., Bavelas et al., 2000; Krauss et al., 1977; Krauss & Weinheimer, 1966; Kraut et al., 1982). If there is more than one addressee, the speaker must keep track of the common ground they share with each of them; the mutual knowledge that they can assume with one addressee may not hold with the other. What effect does this have on speakers' abilities to produce utterances that are adapted to particular addressees' needs? The next section considers three studies that address this question experimentally.

3.2 Referential Communication under Increased Load

3.2.1 Effects of time pressure

Chapter 2 introduced Horton and Keysar's (1996) Monitoring and Adjustment hypothesis for the role of optimal design in language production. This proposal was based on their study of the effects of time pressure on referential communication. Participants described target objects (geometric shapes) to an addressee whose task was to decide whether their visual display matched the speaker's. In the *shared context* condition, both the speaker and the addressee also saw a context object that shared one or more features with the target object. In the *privileged context* condition, speakers knew that the context object on their screen was not seen by the addressee. Crucially, Horton and Keysar were interested in the extent to which speakers' descriptions included *context-related adjectives*, or adjectives that contrasted the target object with the context object (Schober & Brennan, 2003). For example, if the target object was a large sealed envelope, and the context object was a smaller sealed envelope, then the adjective *large* in a description like *a large sealed envelope* is a context-related adjective, because it refers to the property that distinguishes the target object from the context object. But the adjective *sealed* in the same description is context-unrelated, because it refers to a property that is shared by both of the objects.

Optimal design (or more accurately, Horton and Keysar's version of optimal design, which they called the *Initial Design hypothesis*) predicts that speakers should produce fewer context-related adjectives in the privileged context condition than in the shared context condition. Why? Well, if the context object is privileged information for the speaker, then describing the target as *a large sealed envelope* is overly specific from the addressee's point of view. Essentially, Horton and Keysar were interested in whether speakers would try to avoid being overly specific in the privileged context condition, by *not* including adjectives that they were likely to produce when the context object was shared knowledge with the addressee. When speakers could take as long as they wished to prepare a description before they started speaking, they showed evidence of optimal design. Fewer context-related adjectives were produced in the privileged context condition than in the shared context condition. But when speakers were under time pressure to initiate their descriptions quickly, they produced just as many context-related adjectives in the privileged and shared context conditions. Horton and Keysar interpreted this as evidence against the use of optimal design in initial utterance planning.

In essence, Horton and Keysar showed that, under time pressure, speakers tend to produce descriptions that are overspecified from the addressee's perspective. In fact, this is not terribly surprising. We know that, under experimental conditions at least, speakers very often produce descriptions that would be considered over-specific in Gricean terms (e.g., Deutsch & Pechmann, 1982; Eikmeyer & Ahlsen, 1998; Ford & Olson, 1975; Mangold & Pobel, 1988; Pechmann, 1984; 1989; 1994). But unlike ambiguous references, overly specific references do not necessarily make comprehension more difficult. When people have to identify one object from an array on the basis of a written description, they identify that critical object just as quickly when the description is overspecific (in the current context) as when it is optimally informative (Noordman, Maes, & Arts, 2003; see also Mangold & Pobel, 1988; Sonnenschein, 1982; 1984). Perhaps Horton and Keysar's speakers produced overly informative descriptions because their own experiences outside the psychology laboratory told them that this would not make life particularly difficult for their addressee. Indeed, they might even have believed that extra information would be helpful (Dale & Reiter, 1995; Mangold & Pobel, 1988). To decide whether speakers pay less attention to optimal design under increased load, it may be less informative to look at how *much* information is produced ('just enough' versus 'too much'), but rather the *content* of speakers' descriptions (cf. Brennan & Clark, 1996). This is exactly what Horton and Gerrig (2002) did in a subsequent study.

3.2.2 Learning when and how to tailor utterances to addressees

Horton and Gerrig (2002) had Directors play a referential communication game with two different Matchers. The stimuli for this experiment were picture cards showing common living things (birds, fish, and flowers) and tangrams figures which looked a little like real objects (boats, people or rockets). In the first part of the experiment (rounds 1-3 of the game), both Matchers were present, but each of them only matched a sub-set of the Director's cards.

In the second part of the experiment, the Director played a further three rounds of the game with each Matcher in turn. Both Matchers now had the full set of cards, of which some were "new", in the sense that they had not seen them during the first three rounds. Of particular interest was how Directors described those "new" cards, compared to how they described cards that were "old" for the current Matcher. "Old" cards that an addressee had discussed with the Director during the first part of the experiment was termed "shared", and cards that were new were termed "non-shared". If speakers are generally sensitive to what is and is not in common ground with a particular addressee, then Directors should show less referential shortening and less entrainment for non-shared cards than for shared cards. Why? Well, if a particular card is non-shared (or new information) for the current addressee, the Director will probably need to give a more detailed description, or at least avoid a very idiosyncratic conceptualisation that they might have developed with the previous Matcher.

Horton and Keysar looked for evidence of optimal design in three different ways. Firstly, they counted the number of idea units (e.g., Fussell & Krauss, 1992) that Directors produced in their initial descriptions of shared and non-shared cards. Nouns, adjectives and prepositional phrases were counted as belonging to separate idea units. For example, they coded the following description as including nine idea units, which are underlined: *a fish, its fins are uh red, its body is blue and green, looking to the left, eyes sorta yellowish*. If Directors are sensitive to the Matcher's need for more detailed descriptions of non-shared cards, they should produce more idea units for non-shared than for shared cards. They also counted the number of *hedges* in Directors' initial descriptions for shared and non-shared cards. Hedges are expressions like *kind of*, *sort of*, and morphemes like *–y* and *-ish* that are assumed to mark the tentativeness or provisionality of a conceptualisation (Brennan & Clark, 1996; Brennan & Ohaeri, 1999). Horton and Gerrig interpreted hedges as an index of "… how certain Directors are that their description is adequate" (p.596). Directors should be more likely to mark descriptions of non-shared than shared cards with hedges, especially if they are using a new or expanded conceptualisation. They should also be more likely to abandon a conceptual pact and *reconceptualise* non-shared cards (more than shared cards), if they are sensitive to the fact that the current Matcher may not understand a very idiosyncratic conceptualisation that they have established with the other partner.

Together, these measures basically showed that Directors were sensitive to their addressee's needs. They tended to use more idea units, more reconceptualisations, and more hedges for non-shared cards than for shared cards. However, Directors showed *more* evidence of tailoring their descriptions for non-shared cards to the new addressee's needs at the second partner change (round 7) than at the first partner change (round 4), in the sense that they did better at avoiding the usual entrainment and referential reduction effects. This was particularly true for the tangram cards, which were more likely to be reconceptualised at round 7 than at round 4. Since tangrams are novel or unfamiliar to Directors (but everyday things like flowers, birds and fish are not), this suggests that speakers need experience in a particular conversational domain to learn *how* to tailor what they say to the addressee's needs. In effect, Horton and Gerrig suggested that feedback about adequate and inadequate references helps speakers to overcome the tendency to assume that their own perspective will be shared by an addressee (Fussell & Krauss, 1992; Nickerson, 1999; Nickerson, Baddeley, & Freeman, 1987).

3.2.3 Keeping track of who-knows-what

In Horton and Gerrig's (2002) experiment, speakers could keep track of which Matcher had what cards relatively easily, simply by remembering that one partner had all of the fish and frog cards, while the other had all the birds and flowers. In practice, of course, making inferences about the way the experimenter has designed the study is easier said than done, as Horton and Gerrig point out. But in principle, speakers could have exploited such cues. What happens when memory load for the speaker is increased, such that there is no longer a simple association between one Matcher and a particular category of stimuli? This is what Horton and Gerrig (submitted) set about to investigate in a follow-up study.

Essentially, this experiment was designed to find out what happens when partner-specific memory representations are harder to access. Under what conditions would a memory representation be considered less accessible? Horton and Gerrig (submitted) give a simple example, which I elaborate here to make it clear how this situation might come about. Imagine that Alex is looking through a book about ornithology with his friend, Oliver. Oliver is interested in bird watching, and has talked about birds with Alex many times. In fact, Alex and Oliver have looked at this book together before, and referred to one of the pictures in the book as *the firecracker bird.* For Oliver, the picture of this particular bird is 'old', and he has established a conceptual pact with Alex for that picture, which is part of the common ground they share. Another friend, Ned, has never seen Alex's bird book, and they have never talked about birds together before. Alex and Ned have no conceptual pact for this bird, and bird conversations in general are not part of their common ground.

In an 'ideal' situation, Alex should find it reasonably easy to remember that he has referred to the picture as the *firecracker bird* with Oliver, but not with Ned, because it should be reasonably easy to remember that he has never discussed birds with Ned before. He can make use of the conceptual pact he shares with Oliver in a subsequent conversation; it is reasonable to assume that Oliver will understand what Alex means when he hears *the firecracker bird.* However, *the firecracker bird* is a rather idiosyncratic description for the picture. If *Ned* had to pick out which picture Alex was talking about, it is unlikely that he would be able to understand this description unless Alex added some extra information, or somehow reconceptualised his description to make the reference more transparent. In other words, when referring to the picture for the first time with Ned, it would probably be better for Alex *not* to rely on the conceptual pact he created with Oliver, and under optimal conditions it should be easy for Alex to bear this in mind.

Compare this with a situation in which Alex has looked at the bird book with both Oliver and Ned in the past, but he has only referred to this particular picture as *the firecracker bird* with Oliver. When Alex comes to refer to the picture for the first time with Ned, it might be harder for him to remember whether or not he will easily understand an idiosyncratic reference. Alex will have to remember whether *the firecracker bird* is one of the conversational precedents that he has established with Ned, or with some other interlocutor. In this sense, having discussed birds with both Oliver and Ned makes it harder to access partner-specific memories about particular conceptual pacts.

Horton and Gerrig (submitted) compared these two kinds of situation experimentally. Directors in a communication game talked about different kinds of picture cards with two different addressees. For example, in their *orthogonal* condition, the Director might discuss birds and dogs with Matcher A, and fish and frogs with Matcher B. Put another way, the two Matchers were associated with completely different sets of picture cards. In the *overlapping* condition, Directors talked about the same types of picture cards with both addressees. The cards were distributed such that each Matcher had two of each kind of card (birds, dogs, fish and frogs). For example, both Matchers had two bird pictures in their array, but Matcher A's two birds were different pictures from Matcher B's birds. Horton and Gerrig suggested that these two conditions would put speakers under different memory demands. In the orthogonal case, speakers can use category information (e.g., 'birds') to associate a type of card with a particular addressee. There is no need to make specific partner-picture associations, as the category label is enough to remember which addressee has what cards. In the overlapping case, the speaker's task is more difficult. It is not enough simply to remember which addressee has what set of cards (e.g., 'Matcher A has the birds and dogs, Matcher B has the fish and frogs'). Instead, the Director must keep track of specific associations between, for example, Matcher A and a picture of a yellow fish with a green fin.

Directors played four rounds of a matching game with one partner, and then four rounds with a second partner. Then they played a final round with each Matcher in turn (rounds 9 and 10). In these final rounds, Matchers were given the complete set of cards, half of which were 'old' for that partner (in the sense that they had matched them already for four rounds), and half were 'new' (they had not matched those cards before). The interesting comparison is between a reference to a 'new' card (for the current Matcher) in round 9 or 10, and the Director's most recent description for the same card, which had been directed towards the other partner. By that point, the Director has already referred to the card four times, and will probably have established a conceptual pact for that picture with the other Matcher. However, the referring expression they most recently used for the card may not be sufficient to help the new Matcher uniquely identify the referent (just as Ned might have trouble picking out the *firecracker bird* from Alex's bird book in the example above). So if Directors are designing their utterances with a particular addressee in mind, we would expect them to avoid the usual tendency towards entrainment. Instead, they should reconceptualise a description that is new to the current Matcher so it will be more readily identifiable. If this is easier when conversational precedents were encoded under low memory load, Directors should produce more reconceptualisations for new cards in the orthogonal condition than in the overlapping condition – and this is exactly what Horton and Gerrig found.

In the critical rounds (rounds 9 and 10), Directors reconceptualised about 35% of their descriptions for 'old' cards (cards they knew the current Matcher was familiar with), compared to about 50% of descriptions for 'new' cards (cards they had not previously described to the current Matcher). This shows that Directors were somewhat sensitive to the need to avoid established, idiosyncratic precedents when they spoke to a new addressee (cf. Wilkes-Gibbs & Clark, 1992), although clearly they only remembered to do so about half of the time. Crucially, Directors showed more evidence of optimal design when it was easier to keep track of which cards had and hadn't been discussed with the current Matcher. In the orthogonal condition, Directors reconceptualised 35% of their descriptions for old cards, versus 57% of descriptions for new cards. This difference between the old and new cards was smaller in the overlapping condition (36% reconceptualisations for old cards versus 44% for new cards), suggesting that speakers were taking their partner's (conceptual) perspective into account less often under increased load. To put it another way, speakers showed less evidence for co-operation when they had not been able to associate card categories as 'belonging' to different addressees.

This finding suggests that speakers show less evidence of tailoring what they say to fit the addressee's perspective when it is harder to keep track of 'who knows what'. Essentially, Horton and Gerrig (submitted) put Directors in the overlapping condition under increased load at the point of establishing partner-specific memory representations for the various cards. What about increased load at *recall*; would speakers show similar difficulties if they were put under extra cognitive load at the point of remembering different sets of conversational precedents? Experiment 1 was designed to investigate this question. Participants in a referential communication task played the role of Matcher with two different Directors. Having played the matching game four times with each partner, participants took over the role of Director, and either described the entire set of tangram figures to one Matcher at a time (*Low Load* condition), or they alternated between Matchers, describing one card to each in turn (*High Load* condition). In effect, Directors in the High Load condition had to hold two conversations on the same topic concurrently, which should be harder than having just one conversation at a time. If so, we should see less evidence of optimal design in the High Load condition than in the Low Load condition.

EXPERIMENT 1
Effects of Cognitive Load on Optimal Design

3.3 Rationale and Predictions

Two different Directors (players A and B) described tangram cards to a Matcher (player C) for four rounds each (rounds A1-A4 and B1-B4). This was to give player C the opportunity to establish two different sets of conceptual pacts for the tangrams with their two partners. B was out of the room while A was the Director, and vice versa. Then the players swapped roles for one final round of the game (round 9); A and B became the Matchers, and player C described the tangrams to both of them. This was to test whether player C could remember the different conceptualisations associated with their two partners, and produce the appropriate description for the current partner. A and B were both present during round 9. In the *Low Load* condition, C described the full set of cards to A before describing the same cards to B. In the *High Load* condition, C alternated between the two Matchers, describing one card to each in turn.

In the Low Load condition, associations between individual conceptualisations are likely to help speakers remember the "right" perspective for the current addressee. For example, remembering that one card was called *the goalie* with the current Matcher should increase the likelihood of remembering that the next card was *the car seat*, because the perspectives *goalie* and *car seat* were established in the same conversation. To put it more formally, Directors in the Low Load condition should find it easier to recall the "right" conceptualisation for the current Matcher, because associations between memory representations encoded in the same context (i.e., during the same conversation) should increase activation for the descriptions associated with that person, relative to the descriptions associated with the other Matcher. In the High

Load condition, constantly switching between partners should make it less likely that associations between the various cards help Directors to remember exactly the right conceptualisation for the current addressee, because both sets of conceptualisations are activated concurrently. If this is the case, we would expect C's descriptions to show less evidence of optimal design in the High Load condition than in the Low Load condition. The High Load condition might also be more difficult for the *Matchers* (players A and B) in round 9, because they have to alternate between being an addressee and being an overhearer, and must keep track of when they should attend to the Director's descriptions.

Greater difficulty in the High Load condition could manifest itself in several different ways. Firstly, global measures of *collaborative* effort (e.g., number of speaking turns and amount of time taken to complete the matching task) should be higher for High Load groups than for Low Load groups. In addition, Directors' initial descriptions for the tangrams (what they say in their first speaking turn for each card, before the Matcher gives feedback) should also show an effect of increased cognitive load. Directors might be more likely to get their two partners' previous descriptions mixed up, in which case their references may include words or phrases from the "wrong" conceptualisation. They might also be less confident that the referring expressions they are producing are the right ones, in which case they might use more hedges – markers of provisionality (Brennan & Ohaeri, 1999; Horton & Gerrig, 2002) – in the High Load condition. Additionally, their initial descriptions might be less focused; they might include more information in their initial descriptions, on the chance that something they say will strike a chord with the addressee. So, High Load Directors may use more words or idea units (cf. Fussell & Krauss, 1992; Horton & Gerrig, 2002) to describe the cards.

3.4 Method

3.4.1 Participants

36 volunteers from the University of Edinburgh student community took part in this experiment for a small payment. Almost all were native speakers of English (one participant was Swedish/English bilingual with native-like competence in English). Volunteers participated in twelve groups of three people. Most of the students already knew at least one of their partners in the experiment. Only one group was made up of participants who had not met either of their partners before.

3.4.2 Materials

20 tangram figures were used in this study, individually photocopied onto 7 by 7 cm white cards and laminated (see Appendix A). 16 of the tangrams were previously used by Schober and Clark (1989), and four additional tangrams (judged to be similar in shape to at least one of the Schober and Clark figures) were added to complete the set. The tangrams were pre-tested for name agreement. 12 volunteers were shown the figures printed on white paper, and were asked to provide names or short descriptions for each one. The four cards with the highest name agreement were excluded from the last round of the experiment (see Procedure). These four tangrams are numbered 17-20 in Appendix A.

3.4.3 Procedure

Three volunteers arrived at the laboratory and drew lots to see who would take which role in the communication game: player A, B, or C. They were first asked to read a set of instructions tailored specifically to their role in the experiment, and there was an opportunity for participants to ask questions about the experimental procedure. The experimenter explained that participants would play a communication game, the aim of which was for the Matcher to arrange their set of cards into the same order as the Director's (the target order). They were told that the Director and Matcher could talk to each other as much as they needed in order to complete the task as quickly and accurately as possible. In the first part of the experiment, players A and B would take turns to be the Director, and would describe the cards to player C four times each, with the Director's and Matcher's cards being shuffled between each round so that both players started out with a different random order each time. In the second part of the experiment, player C would become the Director, and would have to describe two sets of the same cards to the Matchers, one set to player A and one to player B.

Once the experimental procedure had been explained, player B was taken to another room to wait, and players A (the Director for this part of the experiment) and C (the Matcher) sat at tables divided by a vertical screen. On each table was a set of tangram cards placed face down in a matrix of 5 columns by 4 rows, with the cells of the matrix numbered to help participants keep track of the cards more easily. The Director's and Matcher's cards had been arranged in two different random orders before the experiment began. Players turned over their cards, being careful to keep them the right way up. The Director was asked to describe the cards in sequence, working from left to right in the first row before moving on to the second row. Rounds were timed and tape-recorded, and later transcribed verbatim. On each round, timing started when the Director began to describe the first card, and ended when the Director and Matcher agreed that they had matching orders. The Matcher's cards were checked against the target order by the experimenter, and any errors were pointed out to both players. Then both sets of cards were shuffled and randomly reordered, and the procedure was repeated.

After round 4, player A was taken to another room to wait. The experimenter explained that C would play the game four more times with player B as the Director, and C was encouraged to let B find their own way of describing the cards, rather than imposing the descriptions they had established with player A. This was to maximise the likelihood that C would have to remember two different conceptualisations for each tangram. Player B was brought back into the room and described the same set of cards to player C four more times, following the procedure described above. Finally, player A was brought back to the lab for the final part of the experiment.

All three players sat at tables divided by screens such that no player could see another person's cards. Player C became the Director, and players A and B were asked to match their cards to the descriptions they heard, as before. Once again, the aim was for the Matchers to put their cards into the same order as the Director's as quickly and accurately as they could. Player C had two identical sets of 16 cards selected from the 20 cards used in the first part of the experiment. These 16 cards were found to have the lowest name agreement when the tangrams were pre-tested (see Materials). The rationale was that these 16 cards should be the ones most likely to elicit different names or descriptions from players A and B during the first part of the experiment. If A and B settle on different conceptualisations for the same card, C must recall two different conversational precedents when they take over as Director in round 9. This maximises the likelihood of a difference between the Low and High Load conditions showing up. Since Directors only described 16 of the 20 cards in round 9, Matchers A and B each had four cards left over at the end of the experiment, such that they could not identify the last card in the set by a process of elimination.

Player C was asked to carry out the matching task in one of two ways. In the High Load condition, C alternated between players A and B, describing one card from each set in turn. In the Low Load condition, C was asked to describe all 16 cards to one Matcher and then 16 cards to the other. Groups were counterbalanced for the order in which C matched the cards with players A and B; for half of the groups, C talked to player A first, for the other half C interacted first with player B.

3.5 Rounds A1-A4 and B1-B4

3.5.1 Errors

Matchers were very accurate overall, making fewer than one error per round on average (mean = 0.76).

3.5.2 Measures of collaborative effort

Previous research suggests that referential communication gets more efficient across rounds in this kind of task (e.g., Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989; Clark & Schober, 1989; Hupet et al., 1991; Krauss & Weinheimer, 1964; Wilkes-Gibbs & Clark, 1992). Pairs of interlocutors typically spend less collaborative effort on the second round than on the first, and on the third round than on the second, and so on. This finding should be replicated in rounds A1-A4 and B1-B4 here. As crude measures of collaborative effort, the mean number of seconds and the mean number of number of speaking turns to place a card were calculated for each round. Turns were counted whenever there was a change of speaker (cf. Brennan & Ohaeri, 1999; Clark & Schaefer, 1987), although purely nonverbal turns like laughing and coughing were excluded from the count. Under this approach, back-channel responses (Yngve, 1970) are counted as separate turns, even though the current speaker does not give up the floor (Bavelas et al., 2000).

FIGURE 2

Experiment 1: Average number of speaking turns and seconds taken to place each card in rounds A1-A4 (Panel A) and B1-B4 (Panel B)

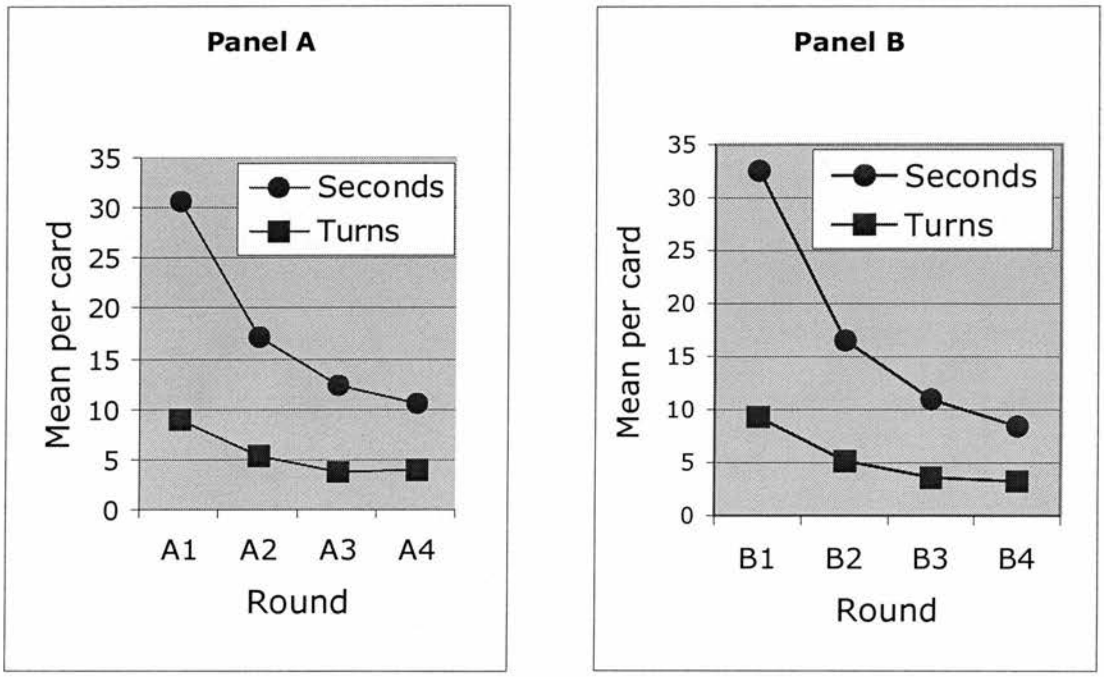


Figure 2 shows that the mean number of seconds taken by Directors and Matchers to place each card got shorter and shorter across rounds (rounds A1-A4: $F(1, 11) = 80.46$, $MSe = 31.84$, $p < .001$; rounds B1-B4: $F(1, 11) = 142.39$, $MSe = 25.56$, $p < .001$). Mean number of speaking turns needed to place each card also decreased over the course of the experiment (rounds A1-A4: $F(1, 11) = 34.20$, $MSe = 4.89$, $p < .001$; rounds B1-B4: $F(1, 11) = 66.93$, $MSe = 3.51$, $p < .001$). In line with previous research (e.g., Clark & Wilkes-Gibbs, 1986; Hupet et al., 1991; Wilkes-Gibbs & Clark, 1992), then, Directors and Matchers needed fewer turns, and less time, to complete the matching task across rounds.

What happens when a new Director takes over in round B1? We would expect them to use long descriptions because they are unfamiliar with the tangrams, and this should make communication less efficient. Indeed, this was the case: Directors and Matchers used more speaking turns ($t(11) = 7.34, p < .001$) and took longer to place each card ($t(11) = 9.13, p < .001$) in round B1 than in round A4. On the other hand, we might predict that round B1 would be quicker and more efficient than round A1 overall. By this point, the *Matcher* (player C) has already played the game four times and is much more experienced with the tangrams than they were at the start of the experiment. However, round B1 is just as long and laborious as round

A1, despite the Matcher's experience. Pairs take just as long to complete the matching task ($t(11) = .589, p = .567$) and use just as many speaking turns ($t(11) = .363, p = .723$) in round B1, when the new *Director* is introduced, as they do in round A1, when both the Director and the Matcher are new to the game. Apparently, having to match familiar cards to new descriptions makes the game difficult for player C when a second partner is introduced (cf. Barr & Keysar, 2002; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992). This underlines the collaborative nature of the referring process (Clark & Wilkes-Gibbs, 1986). Apparently an "expert" Matcher in this kind of task does not guarantee efficient communication (Bortfeld & Brennan, 1997; Isaacs & Clark, 1987).

3.6 Round 9 - What happens when C takes over as Director?

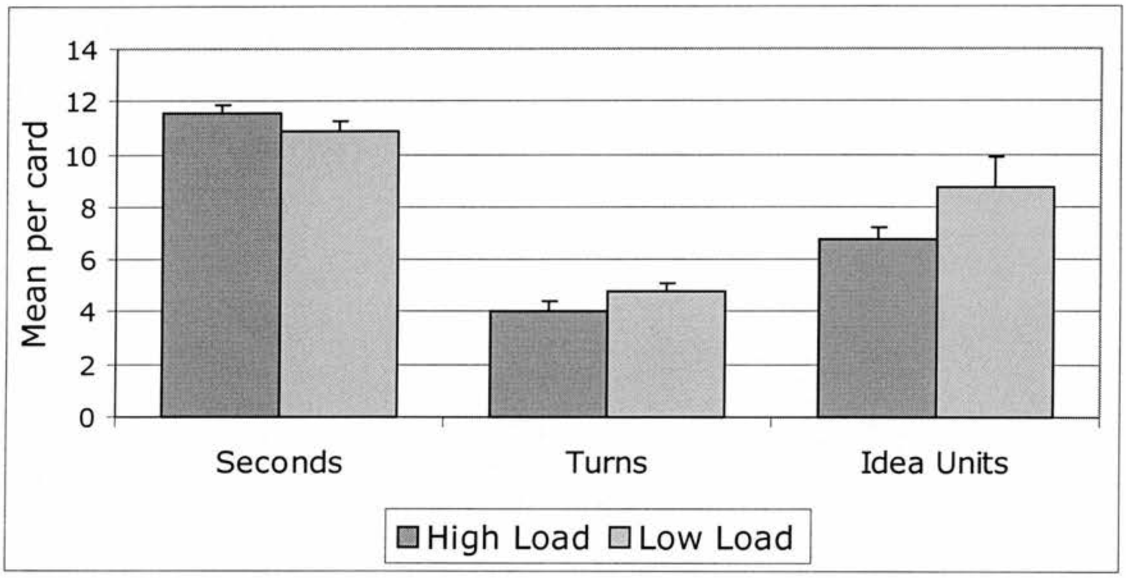
In round 9, player C takes over as Director, having been the Matcher for eight previous rounds. By this point, it was assumed that C would have quite firmly established two different sets of conversational precedents for the tangrams with their two partners. Is taking over the role of Director relatively straightforward for player C? It seems not. Directors and Matchers used more speaking turns (round B4: mean = 3.26, SD = 0.62; round 9: mean = 4.40, SD = 0.92; $t(11) = 4.64, p = .001$) and took longer to place each card (round B4: mean = 8.43, SD = 1.82; round 9: mean = 11.23, SD = 1.88; $t(11) = 3.61, p = .004$) in round 9 than in round B4. However, clearly round 9 was much more efficient than either rounds A1 or B1, in terms of both the mean time taken per card (means: A1 = 30.69, SD = 1.96; B1 = 32.55, SD = 2.30; round 9 = 11.27, SD = 0.54; both ps < .001) and the mean number of speaking turns that Directors and Matchers used to place each card (means: A1 = 8.96, SD = 0.86; B1 = 9.33, SD = 0.83; round 9 = 4.40, SD = 0.26; both ps < .001). But given the cognitive load manipulation, the interesting question is this one: is communication *less* efficient when player C has to hold two conversations at once in round 9, than when they talk to just one Matcher at a time?

3.7 Round 9 - Effects of Cognitive Load

The major prediction for Experiment 1 is that C should find this task more difficult in the High Load condition, because they have to constantly switch between two sets of conversational precedents or conceptual pacts. A general test of this prediction would be to see how much collaborative effort groups spend on completing the matching task. Figures 3 and 4 compare groups in the High and Low Load conditions on various measures of global collaborative effort. Each measure is discussed in turn below.

FIGURE 3

Experiment 1: Mean seconds, turns, and idea units per card, by condition (High versus Low Load)



3.7.1 Errors

We might expect Matchers in the High Load condition to make more errors overall in placing the cards. Why? Well, High Load Directors might be less able to concentrate on making sure that the current Matcher has got the right card, and they might have to attend more to remembering who-said-what, or preparing for their next description to the other Matcher. Only two Matchers (one in the High Load condition and one group in the Low Load condition) made any mistakes at all, so the overall error rate was extremely low (1%), and the *pattern* of errors (comparing the Low and High Load conditions) was uninformative (and is therefore not graphed in Figure 3 above).

3.7.2 Time to complete the task

Greater cognitive load in the High Load condition might be reflected in the overall amount of time needed for Directors and Matchers to complete the task. In fact, High Load groups completed the matching task just as quickly as the Low Load groups (mean seconds per card: High Load condition = 11.55, *SD* = 1.61, Low Load condition = 10.90, *SD* = 2.23; $t(10) = .580$, $p = .576$). Of course, this is a between-participants comparison, and since some people talk more quickly than others, a crude measure such as time to complete the task may not tell us very much about differences between the High and Low Load conditions.

3.7.3 Number of speaking turns

Speaking turns were counted whenever there was a change of speaker in round 9 (cf. Brennan & Ohaeri, 1999; Clark & Schaefer, 1987). There are two reasons why High Load groups might use more turns to complete the task than Low Load groups. Firstly, if High Load Directors are having trouble remembering who-said-what, they might try to present their descriptions in smaller instalments, and wait for the addressee's feedback to help them work out whether they have remembered the right perspective for the current Matcher. Secondly, if a Director initially forgets the conceptualisation that the current addressee previously used for a tangram, and the Matcher cannot identify it on the basis of what they have heard so far, then the pair would need to take more turns to match the card than if the Director produced the right description straight away. Either way, we would expect High Load groups to use more speaking turns overall than Low Load groups. However, Low Load groups actually used numerically *more* turns to complete the task than High Load groups (means: Low Load condition = 152.33, *SD* = 29.18; High Load condition = 129.00, *SD* = 27.11), although number of turns was not significantly different between the two conditions ($t(10) = 1.44, p = .182$).

3.7.4 Idea units

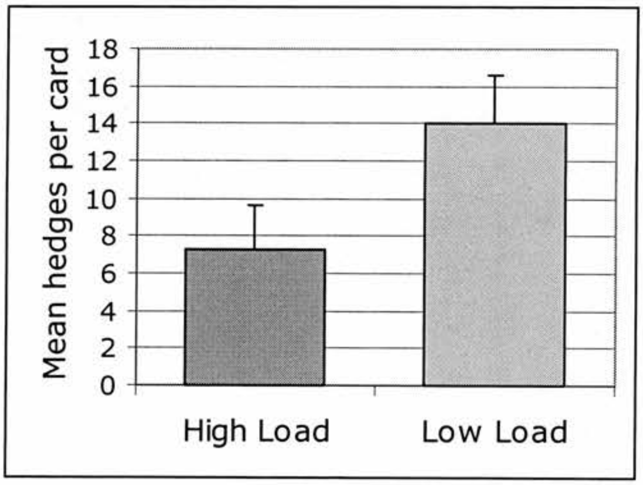
If Directors have trouble recalling particular perspectives, speakers and addressees might have to exchange more information (per card) to carry out the task. To test this hypothesis, the overall number of idea units produced by Directors and Matchers in round 9 was calculated. Following Horton and Gerrig (2002), idea units rather than words were counted to avoid inflating counts with speech that has nothing to do with describing the cards, but instead is used to manage the dialogue task (e.g., helping the Matchers keep track of which card position is being talked about). Presumably there would be more of this kind of talk in the High Load condition, because Directors have to make sure that the two Matchers are "tuning in" to descriptions at the right times. Counting idea units rather than number of words is therefore a more conservative test of the cognitive load manipulation. Each noun and adjective was counted as belonging to a separate idea unit. The following example includes eight idea units, which are underlined:

> *number 6, er, it's like an abstract shape, it's a bit like, in the bottom there's like a little triangle like extra, and on the left there's a square*

There was no significant difference between the Low and High Load conditions in terms of idea units ($t(10) = 1.61$, $p = .153$), although the Low Load groups produced numerically more idea units on average than the High Load groups.

FIGURE 4

Experiment 1: Mean hedges per group, by condition (High versus Low Load)



3.7.5 Hedges

Hedges are expressions that are considered to mark an utterance as provisional or tentative (Brennan & Clark, 1996; Brennan & Ohaeri, 1999; Horton & Gerrig, 2002). As Brennan & Ohaeri (1999, p. 230) put it:

> *hedges… display that the speaker is taking a provisional stance toward the utterance; they may grant the addressee license to reject or modify the utterance… What hedges […] signal is that a speaker's stance toward a proposition is one that invites (but does not demand) a partner's input.*

For the purposes of this analysis, the following expressions were counted as hedges: *kind of (like), sort of (like), a bit (like), maybe, pretty much, almost, I think, I guess, I dunno, I can't remember, or whatever, or something, something like that,* and the morpheme *–y*, as in *with the weird tail-y thing.* Note that the word *like* is often included in counts of hedges (e.g., Brennan & Ohaeri, 1999). *Like* can also be used as a filled pause (similar to *um* or *uh*), as in this example from one of the High Load Directors:

> *Number 9 it's like a diamond shaped like cube erm balancing on a big rectangle which has like a triangle shape cut off at the bottom… and it's extra, like, it's like a big rectangle but there's just stuff cut off*

Moreover, *like* can also be used as part of an analogical expression, as in *it's like a big rectangle* above. To avoid ambiguity, *like* was excluded from the hedge count, unless it was used in the expressions *a bit like, sort of like* or *kind of like.*

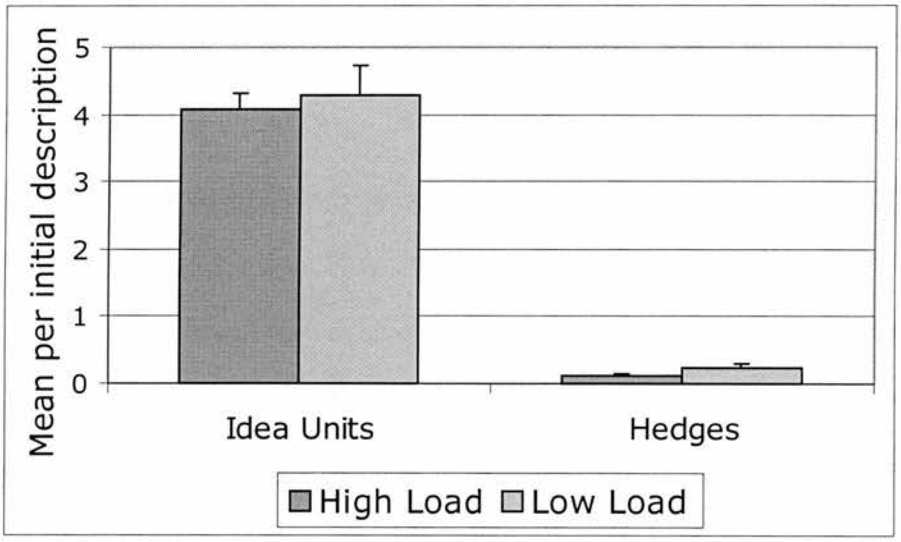
If Directors in the High Load condition have more trouble remembering which of two conceptualisations for the same tangram "belongs to" each of their partners, they may be less confident (than Low load Directors) that the descriptions they are producing are adequate. Hedging allows the speaker to mark a description as tentative, and invites the addressee's feedback. If High Load Directors seem less than confident about their descriptions, Matchers may tentatively request clarification before placing a card. For both of these reasons, we might expect participants in the High Load condition to hedge their contributions more than participants in the Low Load condition. Statistically, there was no significant difference between the High and Low Load conditions in terms of the number of hedges they produced ($t(10) = 1.84$, $p = .096$), although Low Load groups produced about twice as many hedges overall (see Figure 4 above).

3.8 Director's initial descriptions

Cognitive load effects do not seem to show up on global measures of collaborative effort, but do they have an impact on Directors' initial descriptions for each tangram? This might give us a more sensitive measure of how difficult Directors find it to remember Matcher-specific precedents in advance of any feedback from their partners (Horton & Gerrig, 2002). To this end, three measures were investigated: the number of idea units and the number of hedges that Directors produced in their initial descriptions for each card, and the proportion of descriptions that elicited an immediate acceptance from the Matcher. Figures 5 and 6 present the data, and each measure is discussed in turn below.

FIGURE 5

Experiment 1: Player C's initial descriptions – idea units and hedges



3.8.1 Idea Units

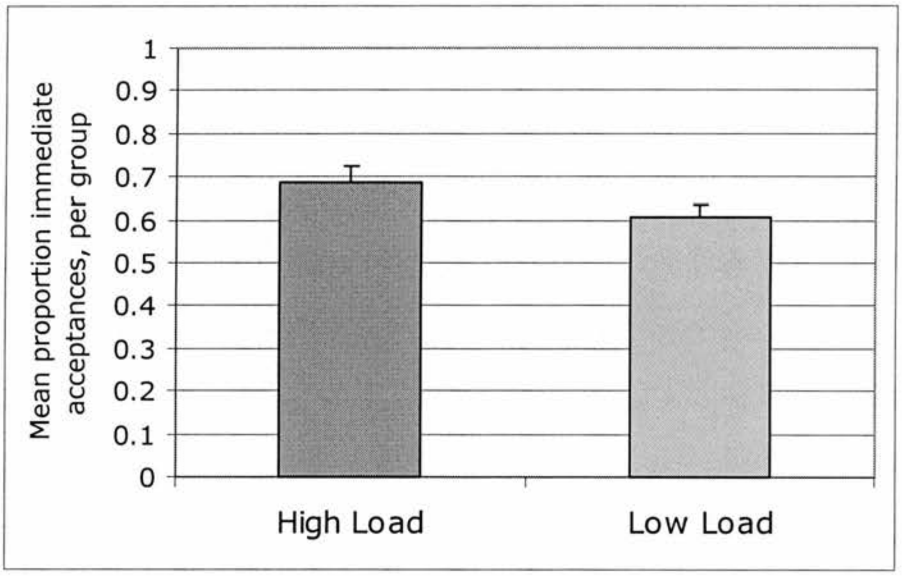
Although Directors and Matchers do not use more idea units overall in the High Load condition, it is possible that High Load Directors do produce more idea units in their initial descriptions than Low Load Directors. If having to switch between two sets of conceptualisations makes it harder to remember who-said-what, High Load Directors might produce more information in their initial descriptions, as they may be more likely to conflate their two partners' perspectives. In fact, as with the global measures of effort reported above, there was no significant difference between the High and Low Load conditions ($t(10) = .462$, $p = .654$).

3.8.2 Hedges

Again, we might expect High Load Directors more than Low Load Directors to hedge their initial descriptions, if they are having trouble recalling a particular perspective, or are not sure that their description is appropriate for the current Matcher. But once again, there was no significant difference between the High and Low Load conditions ($t(10) = 1.89$, $p = .089$).

FIGURE 6

Experiment 1: Round 9, Matchers' immediate acceptances



3.8.3 Matcher's feedback

So far, it seems as though initial descriptions produced in the High Load condition are not very different from initial descriptions produced by Directors in the Low Load condition. Statistically, there is no difference between the High and Low Load conditions in terms of either the mean number of idea units or the mean number of hedges per initial description. But this does not necessarily mean that they are producing the "right" descriptions for their addressees. Matchers' responses that immediately follow Directors' initial descriptions may be informative here (cf. Horton & Gerrig, 2002). If High Load Directors are not saying enough (or, not providing descriptions that strike a chord with their addressees), then their partners should be less likely to accept descriptions as sufficient, and more likely to query them or seek clarification. Following Horton and Gerrig (2002), Matchers' responses to Directors' initial descriptions were therefore coded as *acceptances* or *queries*. A response was coded as an acceptance when the Matcher indicated that they had successfully identified the card by saying *Sure*, *Okay*, *Yes*, *Yeah* or *Got it*. All other responses were coded as queries, including those that indicated successful identification but included additional information as a way to clarify C's description, as in the following example:

C: *'kay, 8 is the dancer, funny dancer*
B: *8's the funny dancer, yep, head thrown back?*
C: *head thrown back, yep*

Acceptances and queries are in complementary distribution, and an analysis was conducted on the proportion of acceptances. Once again, there was no significant difference between the High and Low Load conditions ($t(10) = 1.81$, $p = .101$). Matchers in the High Load condition were no less likely to immediately accept a description than Matchers in the Low Load condition (mean proportion acceptances: High Load condition = 0.69, SD = 0.09; Low Load condition = 0.61, SD = 0.06). So it seems as though Matchers are not having appreciably more trouble identifying the tangrams when the Director is under increased load.

3.9 Qualitative Analysis of Directors' Descriptions

Having considered various quantitative measures of individual and collaborative effort, it may be interesting to consider the nature of the descriptions that Directors were producing in round 9. Broadly speaking, did player C manage to remember the conceptualisations that had previously been established with each partner? To this end, an exploratory qualitative analysis was conducted, which compared C's initial descriptions against A and B's most recent descriptions (from rounds A4 and B4) for the same cards. Six broad types of description were identified. Some seemed to be entrained upon the same or similar conceptualisation that A or B had previously used. Others expressed a completely different conceptualisation. In the middle were descriptions that had elements in common with A or B's previous conceptualisation, but incorporated new elements that A or B had not included in their description for the same card (an "element" here is defined as either an idea unit repeated verbatim from a previous description, or a phrase that expresses the same or similar propositional meaning; e.g., *woman* and *girl*). The six types of description are outlined below, with examples in Table 2 below.

- Type 1: Basically entrained

  Player C produced the same or similar conceptualisation that the current Matcher had used in their most recent description of the same card.

- Type 2: Entrained plus new elements

  Player C produced the same or similar conceptualisation that the current Matcher had most recently produced, but they incorporated one or more new elements – words or phrases added by player C themselves, which had not been used by either A or B in rounds A4 or B4.

- Type 3: Entrained plus one "rogue" element

  Player C produced the same or similar conceptualisation that the current Matcher had most recently produced, but they included exactly one "rogue" element from the *other* Matcher's previous description.

- Type 4: Mixed conceptualisation

  Player C produced a description that mixed elements from the current Matcher's most recent description with more than one element from the *other* Matcher's description.

- Type 5: "Wrong" conceptualisation

  In describing a card to A or B, player C produced the same or similar conceptualisation that *other* partner had most previously used. Note that this is basically a source error – there is nothing inherently "wrong" about the description, but player C has misremembered from whom they originally heard it.

• Type 6: Total reconceptualisation

The final kind of description expressed a complete reconceptualisation of the tangram, bearing no relationship to either A or B's previous conceptualisation.

The distribution of descriptions across these six categories is reported in Table 1, for the High and Low Load conditions. Clearly, this analysis should be regarded as highly exploratory, since there is no measure of the reliability of the coding. However, with this caveat in mind, it is interesting to note that both the High and Low Load Directors produced a high proportion of entrained initial descriptions (i.e., they seemed to express the "right" conceptualisation for the current Matcher). Numerically, High Load Directors produced more descriptions of Type 3, which were basically entrained on the current partner's previous conceptualisation but included one "rogue" element from the *other* partner's description. This might point to an interesting qualitative aspect of descriptions produced under increased cognitive load, but on the basis of this exploratory analysis, we cannot draw any firm conclusions. Whether this classification scheme would generalise to other data sets is a potential avenue for further research.

TABLE 1

Experiment 1: Distribution of description types, by condition

(*Standard Error* in brackets)

	Condition	
	High Load	Low Load
Type 1: Entrained	44% (4%)	51% (2%)
Type 2: Entrained plus new elements	17% (4%)	15% (3%)
Type 3: Entrained plus one "rogue" element	13% (2%)	4% (1%)
Type 4: Mixed conceptualisation	18% (5%)	12% (2%)
Type 5: "Wrong" conceptualisation	4% (2%)	9% (2%)
Type 6: Total reconceptualisation	4% (1%)	9% (2%)

TABLE 2

Experiment 1: Examples illustrating six kinds of description

Type 1: Entrained

A's description to C (round A4): *The upside-down T*

C's description to A (round 9): *The upside-down T*

Type 2: Entrained plus new elements

A's description to C: (round A4): *The swimming man*

C's description to A (round 9): *The swimmer's logo*

Type 3: Entrained plus one "rogue" element

A's description to C (round A4): *Got a left hand sticking up, no feet or anything*

B's description to C (round B4): *The guy with his hand in the air waving*

C's description to A (round 9): *The one with no legs and just one arm waving*

Type 4: Mixed conceptualisation

A's description to C: (round A4): *Somebody facing left, its head slightly tilted to the left, looks like it's reading*

B's description to C (round B4): *Diamond head, could be reading, a straight back until about half-way down where it goes in to the left*

C's description to B (round 9): *The one where the back goes three-quarters of the way down and then the head's slightly tilted to the left*

Type 5: "Wrong" conceptualisation

A's description to C (round A4): *The boxer*

B's description to C (round B4): *Arrow pointing straight up*

C's description to A (round 9): *The arrow sticking upwards*

Type 6: Total reconceptualisation

A's description to C (round A4): *The strange-shaped thing, which has a square at the top and a diamond parallel to it, not the crab the other one*

B's description to C (round B4): *Square again for a head, looks like its leaning back with one small leg and one big leg*

C' description to A (round 9): *Looks like a weight-lifter I think*

3.10 Discussion

Evidence from the first phase of Experiment 1 replicates a robust finding in the referential communication literature: namely that communication in this kind of matching task gets more efficient over time. In rounds A1-A4 and B1-B4, pairs of interlocutors used fewer speaking turns and took less time to match the tangram cards as the experiment went on. In the second phase (round 9), Experiment 1 tested the hypothesis that speakers under increased cognitive load are less likely to show evidence of optimal design. Cognitive load was operationalised here in terms of whether the Director tried to recall all of the conversational precedents they had established with one partner at once, before recalling all of the precedents they had established with the other partner (Low Load condition), or whether they had to switch between partners, recalling one precedent for each partner at a time (High Load condition). The logic behind the cognitive load manipulation was that it should be harder to constantly switch between two sets of conceptual pacts than to remember all of the conceptualisations established with one partner before moving on to those established with the other partner. All theories about the role of optimal design in production (including Clark and colleagues' Collaborative Framework; see Clark & Wilkes-Gibbs, 1986) predict that increased load during planning should make it harder to tailor an utterance to the addressee's perspective.

Surprisingly, this did not seem to be the case. On all of the global measures of collaborative effort (on the part of Directors and Matchers together) and individual effort (on the part of player C), groups in the High Load condition completed the task just as efficiently as participants in the Low Load condition. Directors and Matchers used about the same number of speaking turns to play the game, and spent about the same amount of time placing each card. This is particularly striking given that Directors in the High Load condition probably had to spend more effort managing the task itself, for example in making sure that the two Matchers were listening to the "right" descriptions, keeping track of which card they were up to, and so on (Horton & Gerrig, 2002). An exploratory analysis of the initial descriptions that Player C produced in round 9 suggests that there may be some interesting qualitative differences between speakers under High and Low cognitive load, but the figures presented in Table 1 should be treated with some caution, since there is no measure of the reliability of the coding.

Why was there no effect of the Load manipulation? One possibility is that any differences between the High and Low Load groups were masked by other factors. In this experiment, cognitive load was manipulated between participants. For player C, the task in round 9 is essentially a test of their episodic memory: how well can they remember which partner used what conceptualisation? We often assume that university students will be roughly matched in terms of cognitive variables like memory, but of course even within a relatively homogenous sample there is some variation. Memory span is known to affect performance in other linguistic domains (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1993), and it may be that the Directors in the High Load condition simply came to the task with better memories (for things like *who-said-what*) than Directors in the Low Load condition. Any effect of cognitive load could be hidden by between-participants differences. So, a stronger test of the Load hypothesis would be to compare *the same speakers* under High and Low Load conditions. But in an unpublished study, Horton and Gerrig (submitted) also manipulated load as a between-participants factor, and found robust differences between their "high load" (Overlapping) and "low load" (Orthogonal) conditions in the same-sized sample of university students. They also tested 12 groups of participants, and had Directors play four rounds of the communication game with each Matcher during the first phase of the experiment. So it seems unlikely that baseline memory differences between the High and Low Load groups in Experiment 1 would completely obscure effects of cognitive load.

In a sense, Horton and Gerrig's (submitted) experiment studied increased load during encoding, at the point of tagging conversational precedents in memory as "belonging" to one partner or the other. They made it difficult for speakers in the Overlapping condition to keep track of which partner had what individual cards, versus a condition where it was easy to keep the two sets of conversational precedents separate by remembering that one Matcher had, say, all of the fish cards and the other had all of the birds. Experiment 1 loaded the system at retrieval, or the point at which speakers had to remember the conceptual pacts they had created with different partners. It may be that cognitive load simply has a smaller effect at recall than at encoding, because once strong partner-specific associations have been created between one interlocutor and a particular set of conceptualisations, they are well remembered. The cognitive load manipulation used here, switching between partners, may not stress the system strongly enough to make a difference.

Indeed, memory experiments using dual task paradigms suggest that cognitive load has different effects at encoding and retrieval (e.g., Cinan, 2003; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Naveh-Benjamin, Craik, Guez, & Dori, 1998; Naveh-Benjamin & Guez, 2000). While memory performance is compromised when participants' attention is divided during encoding, performance at retrieval is immune to the load imposed by a secondary task (although performance on the secondary task itself is impaired). These findings have been interpreted as evidence that encoding processes require attentional resources, while retrieval processes do not

(Naveh-Benjamin et al., 1998). The memory experiments used to support this argument are very different from the referential communication task reported here. In the studies cited above, memory is measured by how well participants recall a list of words, and the secondary task is usually non-linguistic in nature (e.g., a visual reaction time task requiring a manual response). In Experiment 1, there was no secondary task as such; the load manipulation basically compared the effect of concurrently carrying out the same linguistic task (describing a set of picture cards) with two different partners, versus carrying out the task consecutively with each partner in turn. So it is difficult to know whether claims from the memory literature can be used to explain the lack of an effect of the cognitive load manipulation in Experiment 1. But it is clear that referential communication paradigms may offer an interesting point of contact between the fields of language and memory research.

3.11 Chapter Summary

Experiment 1 investigated the effects of cognitive load on a speaker's ability to recall conceptual pacts that they had previously established with two interlocutors. Speakers seemed to be remarkably good at tailoring descriptions to their addressee's perspective, even when they had to hold two conversations at once. Directors and Matchers completed the task just as efficiently whether they were in a High Load condition, where Directors alternated between two partners, or a Low Load condition, where they had two separate conversations.

CHAPTER 4

Optimal Design in Noun Phrase Production

4.0 Chapter Overview

In Chapter 3, we saw that speakers tailor the *content* of referring expressions with particular addressees in mind. In this chapter, we turn to the question of whether referring expressions are also optimally designed with respect to grammatical form, and in particular, word order. Experiments 2 and 3 investigated this question by organising a box of picture cards such that descriptions with some word orders (e.g., *dotty orange square*) would be easier for the addressee to incrementally map onto the array than others (*orange dotty square*). The results suggest that speakers do take the addressee's perspective into account in formulating word order, once they know how the addressee's array is organised.

4.1 Optimal Design in Grammatical Encoding?

In principle, optimal design could influence any or all of the stages of utterance production: conceptualisation, grammatical encoding (functional and positional processing), and phonological encoding. Much of the optimal design literature focuses on conceptual stages of utterance planning, and on the lexical choices that speakers make when they decide how to express a message. As we saw in Chapter 3, there is good evidence that lexical and conceptual choices are affected by optimal design; speakers tend to tailor the *content* of what they say to their addressee's perspective. Less is known about whether *grammatical* aspects of production are also directly influenced by a model of the listener. Some studies suggest that adults and older children produce simplified syntax when they talk to younger children (Fernald & Simon, 1990; Shatz & Gelman, 1973), but this is a fairly coarse adaptation to addressee needs (Brown & Dell, 1987). Do speakers also show more fine-grained syntactic adaptations to an adult interlocutor's perspective?

Some evidence has been used to argue against this idea (e.g., Branigan, McLean, & Reeve, 2003; Brown & Dell, 1987; Dell & Brown, 1991; Ferreira & Dell, 2000). Other experimental findings have been used to suggest that syntactic aspects *are* affected by what the speaker knows about their addressee (e.g., Lockridge & Brennan, 2002), although it is not clear that these effects are located at the level of grammatical encoding *per se*, rather than at conceptualisation. These findings are discussed in 4.3.1. First, we consider the functional role of optimal design at different stages of production, and ask why speakers should bother to take their addressee's perspective with respect to syntax. Does it make sense for grammatical encoding to be sensitive to optimal design?

4.2 Are Grammatical Adaptations Important for Communication?

Presumably the primary goal of communicating is for speaker and addressee to end up with aligned mental representations of the speaker's intended message. We can think of the comprehension process as the addressee's journey from their initial mental model (their starting destination) to one which is in all crucial respects the same as the speaker's (Pickering & Garrod, in press; although of course the addressee's final situation model might not exactly match that of the speaker; Branigan, in press; Schober, in press). When speakers make choices about the content of their utterances, they are trying to ensure that their addressee arrives at the correct mental 'destination' – a situation model that is aligned with their own. In contrast, the grammatical form of an utterance can be thought of as the route along which the addressee travels to reach that destination. Different grammatical routes (e.g., active and passive sentence structures) often lead to the same destination. In other words, there may be different ways to say much the same thing. So, speakers have choices to make with respect to the grammatical route along which they direct their addressee.

Lexical and semantic choices may have consequences for the addressee's final mental destination, so it makes sense for speakers to design *what* they say carefully, with the addressee's perspective in mind. However, in most cases it isn't crucial for the speaker to tailor the grammatical *form* of their utterance to fit the addressee. It doesn't much matter what the grammatical route looks like if it leads to the right place. The addressee might run along a short, newly asphalted motorway (i.e., parse an utterance easily and efficiently) or they might hobble slowly down a long and winding country lane (i.e., parse it more slowly and with greater difficulty). So long as they can eventually get to the right destination, communication has been successful. So there are good reasons to expect to see less evidence of optimal design in grammatical aspects of production than in the lexical and conceptual choices that speakers make. As we will see in 4.3, much of the empirical evidence to date suggests that optimal design has little or no influence on grammatical encoding.

4.3 Evidence for Optimal Design in Grammatical Encoding

4.3.1 Syntactic form of referring expressions

There is some evidence that the syntactic form of a referring expression is tailored to the addressee's knowledge state. For example, in a story re-telling task, Lockridge and Brennan (2002) found that speakers were more likely to mark the instrument of an action as indefinite (*an icepick*) when the instrument constituted New information for the addressee than when it was (situationally) Given by virtue of appearing in a picture. Wilkes-Gibbs & Clark (1992) demonstrated a similar finding in a referential communication task; speakers marked referring expressions as definite more often when they expressed old information for the addressee than when they constituted new information. Likewise, Bard and Aylett (in press) showed that the syntactic form of referring expressions produced during a route communication task reflected the discourse status of landmarks for the addressee, rather than the speaker. When speakers mentioned the same landmark twice, they were less likely to produce a syntactically reduced expression (such as a pronoun) second time around when expression introduced a new entity for the addressee than when the landmark was "old news" for that partner.

To the extent that the syntactic form of a referring expression is generated during grammatical encoding, these findings suggest that syntactic aspects of production are sensitive to optimal design. But it is not entirely clear that the effects actually originate at the level of grammatical encoding. Whether an entity constitutes Given or New information (either for oneself or for an addressee) is determined at the level of the discourse model, and discourse-level information presumably influences choices made during message generation and conceptualisation. So, although these findings are suggestive, they do not unambiguously demonstrate a role for optimal design in grammatical encoding.

4.3.2 Ambiguity avoidance

Grammatically ambiguous sentences are harder to understand than their unambiguous counterparts (e.g., Ferreira & Henderson, 1991; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983) so if grammatical encoding were sensitive to optimal design, we might expect speakers to avoid producing ambiguous syntactic structures as far as they can. *Global* syntactic ambiguities, of course, have consequences for the addressee's mental destination – their final interpretation of the speaker's intended meaning. *Temporary* ambiguities, on the other hand, are generally assumed not to have consequences for the addressee's understanding of what the speaker meant (though see Christianson, Hollingworth, Halliwell, & Ferreira, 2001, for an alternative view), though they may make comprehension more laborious. If we are interested in finding evidence for optimal design at the level of grammatical encoding, the avoidance of temporary syntactic ambiguity might be a good place to look for it.

In a series of experiments, Ferreira and Dell (2000) studied speakers' use of optional grammatical function words, in sentences like *The coach knew (that) you missed practice.* Without the sentence complementizer *that*, the structure of this kind of sentence is temporarily ambiguous; *you* could turn out to be either a direct object of the verb *knew*, or the subject of an embedded clause. In fact, the latter interpretation turns out to be the correct one. If the production mechanism systematically avoids ambiguity as a means of helping addressees find easy routes to comprehension, we would expect speakers to include optional function words more often in utterances that would be ambiguous without them, compared to utterances that are inherently unambiguous (e.g., *The coach knew (that) I missed practice*, where *I* is unambiguously case-marked and therefore should not be misparsed as the direct object of *knew*). In six experiments, however, Ferreira and Dell failed to find systematic evidence for an ambiguity-avoidance strategy in production. Speakers were no more likely to include optional function words in utterances that would otherwise contain a temporary ambiguity than in their unambiguous counterparts.

4.4 Incrementality and Optimal Design

4.4.1 Exploiting incrementality in comprehension

Another way in which speakers could make utterances easier to understand is by exploiting incrementality in language processing. Comprehension research suggests that listeners start to interpret incoming linguistic input piecemeal, rather than waiting for a clause boundary or a similar juncture (e.g., Altmann, 1999; Just & Carpenter, 1980; Steedman, 1989; Tanenhaus & Spivey-Knowlton, 1996). Eye-movement data from "visual world" experiments suggest that the interpretation of reference in particular is closely time-locked to linguistic input (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, Tanenhaus, Chambers & Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Listeners seem to map referring expressions onto objects out in the world as soon as enough information is available to identify a unique referent, or potential set of referents (e.g., Sedivy et al., 1999). For example, on hearing *Touch the plain red square* in the context of several decorated squares, listeners look at the intended referent as soon as they hear *plain*. Since there is only one plain object in the context, the listener can infer which referent is intended even before they hear the rest of the referring expression. However, in the context of several plain red objects, the same noun phrase does not uniquely pick out the intended referent until the listener hears the head noun, *square* (Eberhard et al., 1995).

In principle, speakers could exploit incrementality to help addressees choose the most efficient route to comprehension. In fact, some linguistic theories predict that speakers should use word orders that put the most informative piece of information in a referring expression first. In a series of empirical demonstrations, Oller and Sales (1969) showed that normal adjective order preferences can be overridden by contextual factors, such as what distinguishes the intended referent from other objects in an array. For example, in the context of three striped squares, only one of which was shaded, speakers tended to produce descriptions like *Number 3 is a shaded striped square*, putting the most informative (or "most limiting", in Oller and Sales' terminology) adjective first. Likewise, when the display consisted of three large squares, each a different colour (red, green or yellow), speakers tended to reverse the normal "size before colour" preference in order to produce descriptions like *Number 2 is a red large square* (see also Danks & Glucksberg, 1971; Danks & Schwenk, 1972).

Oller and Sales (1969) basically framed their explanation of these effects in terms of optimal design, although of course they did not use that terminology. They proposed that, given the objects in the array, the "most limiting" adjective (or, the property that best distinguishes the target referent from the other objects in the set) appears first, followed by the next most limiting adjective, and so on, right up to the least limiting attribute, which appears right before the noun. As Oller & Sales (1969, p.219) put it:

> *We may regard the addition of modifiers as a sort of zeroing-in process whereby the speaker limits the range of possible interpretations which may be imposed by the hearer.*

For example, in the context of several large squares, only one of which is red, *red large square* is a more informative description than *large red square*, because the very first modifier that the listener hears allows them to narrow down their search for potential referents to only the red objects in the array. Essentially, Oller and Sales suggested that speakers should exploit the incremental nature of comprehension to order adjectives in the most informative way for a listener. However, none of the demonstrations they reported actually involved an addressee. The task was monologic; speakers produced utterances in isolation, and had no real communicative goal. So we cannot really make any claims about optimal design with respect to word order on the basis of these results. What Oller and Sales' data might instead show is that information that is salient or accessible to a speaker (in this case, the property that best distinguishes a target object from the rest of an array) tends to find its way into early word order positions. In the following section, we will see that precisely this explanation has been put forward to explain another robust phenomenon in production: the tendency for "Given" information to precede "New" (Prince, 1981).

4.4.2 Incrementality in Production and Given-New Ordering

The tendency for salient or accessible information to precede less accessible information is a robust phenomenon in spoken language production (e.g., Bock & Warren, 1985; Branigan & Feleki, 1999; Prat-Sala & Branigan, 2000), and has been interpreted as evidence for incrementality in the production system. Current models assume that as soon as a linguistic representation has been processed at one level of the system, it is fed forward to the next stage (e.g., Bock, 1982; Kempen & Hoenkamp, 1987; Levelt, 1989). The different stages of the process may be cascaded, such that grammatical encoding can begin before the whole pre-linguistic message has been generated, and the retrieval of phonological forms can begin before the syntactic structure of an utterance has been completely specified (e.g., Meyer, 1996).

The strong cross-linguistic preference for Given information to precede New information in sentence production (Bock & Irwin, 1980; Clark & Haviland, 1977; Osgood, 1971; Prat-Sala & Branigan, 2000; Prentice, 1967; Sridhar, 1988) is often thought to reflect the fact that Given entities are already activated at the conceptual level, and are therefore available for grammatical encoding sooner than New entities. In other words, Given-New ordering effects are thought to reflect what is easiest or most efficient for a speaker. An alternative interpretation is that Given-New ordering is a device that helps the speaker "optimize the entry of data into the hearer's knowledge store." (Vallduvi, 1992). On hearing the Given information, the addressee knows which part of their "knowledge store" to access, and they update it with the New information. Under this view, speakers mention Given information before New because this makes comprehension easier for the addressee.

To test this hypothesis, Branigan, McLean, & Reeve (2003) investigated Given-New ordering in a dialogue task, in situations where the speakers' and addressees' prior knowledge of various discourse entities differed. They had speakers describe picture cards to an addressee whose task was to decide whether they had a matching card. Sometimes just the speaker (*Privileged Knowledge* condition), or the speaker *and* the addressee (*Shared Knowledge* condition), or neither (*No Knowledge* condition) saw a scenario picture before the to-be-described "target" card. The scenario was introduced as the beginning of a story, and the target as "how the story ends". Scenarios introduced one of the entities in the target picture. For example, if the target showed a vampire being hit on the head by a falling banana, the scenario would show either the vampire or the banana. Agents and patients therefore sometimes became Given by virtue of appearing in an 'Agent scenario' (the banana, in this case) or a 'Patient scenario' (the vampire). For the target pictures, speakers produced more passive descriptions (e.g., *The vampire is hit by a banana*) after seeing Patient scenarios (cf. Prentice, 1967; Osgood, 1971), regardless of whether the addressee had also seen the scenario card. In other words, entities which were Given for the speaker tended to find their way into subject position of picture descriptions, even when the speaker knew that these same entities were New for the addressee.

Branigan et al. (2003) interpreted this finding as evidence against optimal design. Why? When the scenario card was privileged knowledge for the speaker, the addressee did not see it. In other words, the addressee was listening to the speaker's description for the target card in essentially a null context. Assuming that passive constructions are generally harder to interpret than actives (Bever, 1970; Cupples, 2002; F. Ferreira, 2003, Townsend & Bever, 2001), addressees should always prefer to hear actives over passives in null contexts. The only time addressees should prefer to hear a passive is when they themselves have seen a Patient scenario card (i.e., in the Shared Knowledge condition). The fact that speakers tended to mention entities that were Given for themselves first, regardless of whether the scenario card was Shared or Privileged knowledge, suggests that speakers do not exploit Given-New ordering for the addressee's benefit. Instead, Given-New ordering seems to reflect what is optimal (i.e., most efficient) for the speaker.

So far, we have seen that the evidence for optimal design at the level of grammatical encoding is (at best) mixed. Speakers do not systematically avoid producing syntactically ambiguous constructions that could cause comprehension difficulties for an addressee (Ferreira & Dell, 2000), and they do not seem to tailor Given-New ordering to their interlocutor's knowledge (Branigan et al., 2003). However, speakers *do* seem to tailor the syntactic form of referring expressions to mark the status of referents as Given or New for the addressee. For example, they produce definite expressions less often for objects or entities that are New for their interlocutor, compared to referents that are Given (Bard & Aylett, in press; Lockridge & Brennan, 2002; Wilkes-Gibbs & Clark, 1992). As noted in 4.3.1, it is unclear whether this finding unambiguously demonstrates a role for optimal design in grammatical encoding, but if does, then it presents the intriguing possibility that we might be able to find evidence for Oller and Sales' (1969) hypothesis with respect to prenominal adjective order. Perhaps speakers could be shown to order prenominal adjectives to help the addressee "zero in" on an intended referent? Essentially, this is what Experiment 2 was designed to find out.

EXPERIMENT 2
Prenominal Adjective Order

4.5 Rationale and Predictions

Experiment 2 was designed to find out whether speakers exploit word order to produce picture descriptions that are incrementally optimal from the addressee's perspective. In a more general sense, this question also speaks to the issue of whether grammatical aspects of production can be sensitive to information about an addressee's perspective (cf. Bard & Aylett, in press; Brown & Dell, 1987; Ferreira & Dell, 2000; Lockridge & Brennan, 2002; Wilkes-Gibbs & Clark, 1992). Pairs of participants played a referential communication game in which a Director told a Matcher where to place picture cards (showing coloured, patterned geometric shapes) on a 3 x 3 grid. The Matcher selected pictures from a file box with labelled dividers, in which the cards were either divided into groups by pattern and then subdivided by colour (*Pattern Box*), or grouped by colour and then subdivided by pattern (*Colour Box*). Directors produced noun phrases like *orange dotty square* or *dotty orange square*. Given the way the box of cards was organised, the form of the Director's description had consequences for how easily the Matcher could map that description onto the array. With the Colour Box it would be easier for the addressee to map a "colour-first" description like *orange dotty square* onto the box of cards they were choosing from, because colour is the first property of the card to be mentioned, and the Matcher has to find the orange cards before she can locate the subset of orange cards that are stripy. With the Pattern Box, a "pattern-first" description like *dotty orange square* would make the card easier to find, because the array is structured such that the Matcher has to look for the stripy cards before locating the subset of stripy cards that are orange.

Note that hearing an "incrementally unhelpful" word order (like *orange dotty square* when looking for a card in the Pattern Box) does not necessarily make the Matcher's job particularly difficult. The task of selecting picture cards from a box is fairly easy whether the speaker produces word orders that are helpful or unhelpful. But an unhelpful description cannot straightforwardly be incrementally mapped onto the structure of the card box. Instead, the Matcher will need to mentally recode the order of the adjectives, or at least rehearse the description they have heard to maintain it in short-term memory while they look for the card.

Participants played two roles, Director *and* Matcher, at different points during the experiment. In other words, they had the opportunity both to describe picture cards and to listen to descriptions produced by a partner. Presumably, this switching of perspectives would help speakers to be aware of which word orders can and cannot be incrementally mapped onto the array. If word order is chosen on the basis of what would be helpful for an addressee, then speakers should be more likely to produce pattern-first descriptions (*dotty orange square*) once they know that the Matcher's array is organised by pattern. Conversely, they should be more likely to produce colour-first descriptions (*orange dotty square*) once they know that the Matcher could more easily map this word order onto the array. In contrast, if word order is determined only by what is easiest to produce for the speaker, knowing how the card box is organised should make no difference. Speakers should produce roughly the same proportion of pattern-first and colour-first descriptions both before and after they know something about the Matcher's array.

4.6 Method

4.6.1 Participants

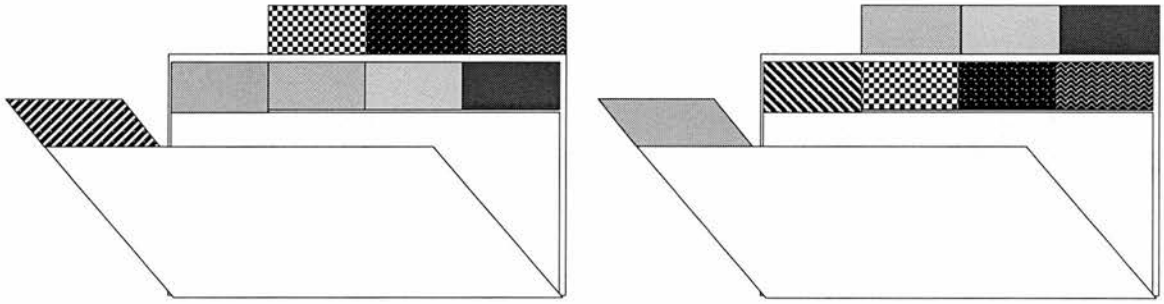
32 volunteers from the University of Edinburgh's student community took part in pairs for payment. Participants were matched on the basis of availability, and only one pair knew each other before the experiment began. All were native speakers of English. There were 5 all-female and 11 male-female dyads.

4.6.2 Design

Pairs of participants played six games of a referential communication task, in two blocks of three rounds. In one block, Matchers selected cards from a box in which picture cards were grouped by pattern at the top level, and then subdivided by colour. In the other block, Matchers selected cards from a box in which the cards were grouped together by colour at the top level, and then subdivided by pattern (see Figure 7). So, the organisation of the Matcher's array was a within-participants manipulation. In the first round in each block, the Director did not know how the Matcher's cards were organised. But by playing the role of Matcher themselves in the second round, they found out about the structure of the Matcher's array. By the third round (when they went back to the being the Director), they knew what would make for a helpful or an unhelpful description, from the Matcher's perspective. The order in which Matchers were given the Pattern and Colour boxes was counterbalanced, such that half of the groups got the Pattern Box for the first block, and the other half started off with the Colour Box.

FIGURE 7

Experiment 2: Patterned and coloured dividers are used to organise the Matcher's box of picture cards



Panel A
Cards organised by pattern at the top level

Panel B
Cards organised by colour at the top level

4.6.3 Materials

The Matcher's task was to select one card at a time from a set of 64 picture cards showing coloured, patterned geometric shapes. The 64 cards represented all the possible combinations of four shapes (square, circle, star and triangle), four colours (orange, turquoise, yellow and purple) and four patterns (stripy, dotty, chequered and wavy; see Appendix B for a list of items). Patterns and colours that could all be named with bisyllabic words were chosen to try to avoid word order biases based on word length (cf. McDonald, Bock, & Kelly, 1993).

Two sets of cards were created and individually pasted onto 47 by 47 mm white cards. The sets were arranged in two separate card file boxes. In the *Pattern Box*, large dividers with four patterned patches on them (stripy, dotty, chequered and wavy) were used to group the cards according to pattern. Within each pattern group, smaller coloured dividers separated the cards according to colour. In the *Colour Box*, large coloured dividers separated the cards into groups by colour, and smaller patterned dividers then subdivided them by pattern (see Figure 7). So, four cards sharing the same pattern and colour ended up behind each small divider. The square card was always placed at the front, followed by the circle, star and triangle (in that order).

Microsoft PowerPoint was used to create slide shows for the Director's instructions. Each slide demonstrated the next card to be described and where the Matcher should place it on a 3 x 3 grid (Figure 8 gives an example of what the Directors saw). A unique set of instructions was created for each pair of participants by randomising the list of 64 possible items and choosing the first 48 (8 cards x 6 rounds; thus the Matcher can never rely on a process of elimination to select the final card, because there are more cards in the array than they use to complete the task). These 48 items were assigned to six lists of eight cards. Each list represented one round during the experiment.

Items were assigned to lists subject to two constraints. Priming studies show that speakers are more likely to use a particular word order or syntactic construction if they have just produced or heard an utterance with a similar structure (e.g., Branigan, Pickering, & Cleland, 2000 – see Chapter 5). This tendency is increased when the two utterances use some of the same words (Cleland & Pickering, 2003; Pickering & Branigan, 1998). The constraints used to construct item lists for Experiment 2 were designed with the priming literature in mind. In each list, every colour and pattern was represented by two cards, but the same pattern-colour combination did not occur more than once. For example, the Director saw two orange cards and two dotty cards per round, but no more than one exemplar from the category *dotty orange*. Within a round, cards with the same colour or pattern did not appear next to each other (a dotty card was never followed by another dotty card, and so on). This should reduce the effect of priming from one utterance to another, which might be important if we want to see evidence of speakers flexibly tailoring word order to fit an addressee's perspective.

Where the Matcher should place each card on the grid was determined by a fixed randomised list used for every pair of participants. For example, the first card in Game 1 might be assigned to the top left corner, the second card to the centre square, and so on. Slide shows were presented on a Sony Vaio laptop computer with a 15" screen (viewing area equivalent to a 17" monitor, resolution 1280 by 1024 pixels per inch).

4.6.4 Procedure

Two participants arrived at the lab, and were told that they would play a collaborative communication game. They drew lots to determine their role in the experiment (player A or player B) and were seated at tables on either side of an opaque screen (see Figure 8). Both participants read a set of instructions explaining the experimental procedure and aim of the game: for one player (the Director) to help their partner (the Matcher) build up a grid of cards as quickly and accurately as possible. Player A took the role of Director in rounds 1 and 3 in each half of the experiment. Player B took the role of Director in round 2, to give player A the chance to be the Matcher and therefore find out how the box of cards was structured (although participants were not explicitly told this). For each round, the Director was seated in front of the computer for presentation of the instructions, and on the other side of the opaque screen, the Matcher was given an empty 3 x 3 cardboard grid and a box of picture cards.

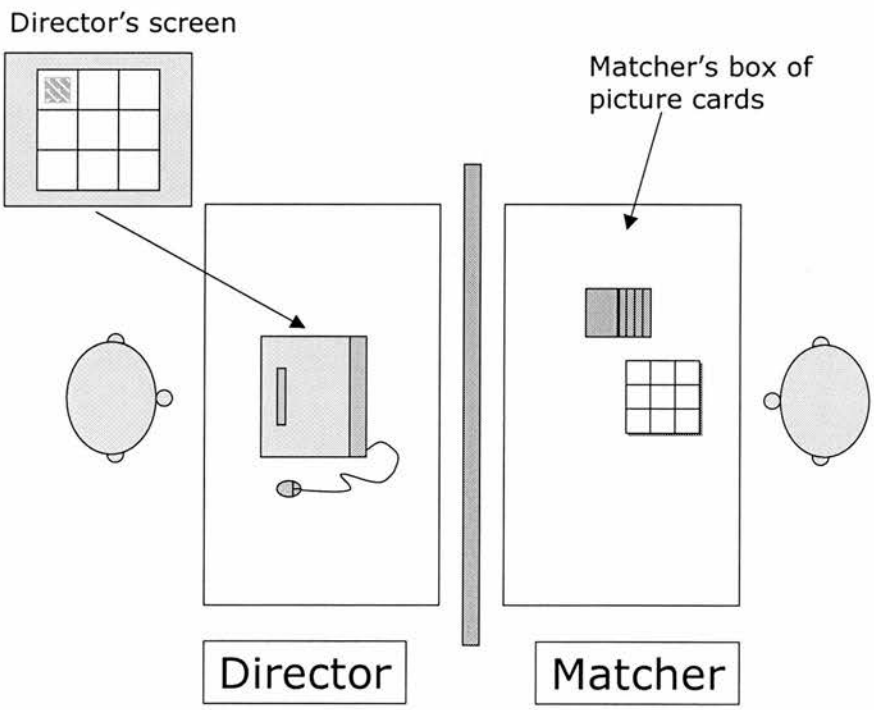
At the beginning of the experimental session, the experimenter introduced the various labels for the patterns, colours and shapes used in the experiment, and both participants had a chance to practise using these labels by describing four warm-up cards (a wavy turquoise circle, a dotty purple square, a stripy orange star and a chequered yellow triangle). Importantly, participants were simply told that some of the cards would be stripy, dotty, chequered and wavy, that some would be orange, turquoise, yellow and purple, and so on. A full noun phrase (*wavy turquoise circle*) was never used to introduce these labels, to avoid biasing participants' choice of word order.

The Director described eight cards per round, leaving one cell of the grid empty each time so that the Matcher could never guess where the last card should be placed. Instructions were presented individually one card at a time, such that the Director never saw the final state of the grid. For each instruction, the Director saw a picture showing (a) the identity of the next card to be described, and (b) where the Matcher should place that card on the grid (see Figure 8) (the grid component was included to disguise the purpose of the experiment, and to give participants an additional task beyond just describing and selecting simple geometric shapes).

Matchers were advised only to give feedback either when they had placed a card, or to request a repeat of an instruction. Occasionally the Matcher was unable to find the card described to them by the Director (usually because the Director had mis-described an earlier item, so that the target card had already been taken from the box); in this case the pair were instructed to move on to the next instruction. Presentation of the stimuli was self-paced: Directors pressed SPACE BAR to begin the round and move on to each new card. After eight cards, a message appeared on screen to signal the end of the round and instruct the Director to change places (and roles) with his or her partner. While the Director and Matcher changed places, the Matcher's filled grid was replaced with an empty one, and the experimenter recorded the Matcher's card placement for that trial (but the participants did not receive any feedback about errors). The cards selected by the Matcher on each round were not replaced in the boxes, but were put to one side.

FIGURE 8

Experiment 2: Overhead view of the experimental set-up



Participants completed two blocks of three rounds. In one block, the Matcher selected cards from the Pattern box. In the other block, they selected from the Colour box. Sessions were counterbalanced for the order in which the two blocks were presented, so half of the pairs saw the Pattern box before the Colour box and the other half saw the Colour box before the Pattern box. After completing one block, participants were asked to fill in a short questionnaire to provide some background information about themselves. This short break allowed the experimenter to unobtrusively switch the card boxes. Although the Pattern and Colour boxes were different colours, Matchers usually did not notice that the boxes had been switched until they began searching for the first card in the next round. All six rounds were timed and tape recorded, and later transcribed verbatim. Timing began when the Director started to describe the first card in each round, and ended when the Matcher indicated that they had placed the final card on the grid. The experimental session lasted approximately 20 minutes in total.

4.6.5 Coding and analysis

The dependent measure of interest was whether player A mentioned pattern before colour in their descriptions, or colour before pattern. Although both players, A and B, took the role of Director at some point during the experiment, only player A produced descriptions both before and after they knew how the Matcher's array was organised, so we focus on player A's descriptions in the results section below. NPs produced by player A were coded as *pattern-first* if the pattern adjective was mentioned before the colour adjective (e.g., *dotty orange square*), or *colour-first* if the colour adjective was mentioned before the pattern adjective (e.g., *orange dotty square*). Crucially, NPs were coded according to the *relative* order of mention for pattern and colour adjectives. For example, a NP like *dotty square that's orange* would be coded as pattern-first, even though it is a complex NP structure in which the head noun, *square*, intervenes between the pattern adjective and the colour adjective. Likewise, *square that's dotty and orange* would be coded as pattern-first, even though pattern is not the first property of the picture card to be mentioned. NPs were coded as *others* if they failed to meet any of the following criteria:

- all three properties of the card must be mentioned (pattern, colour and shape)

- pattern, colour and shape must be mentioned within the same noun phrase
- the description must be a grammatical noun phrase, plus or minus a determiner

Using these criteria, a responses like *it's a square, it's dotty, and it's orange* would be coded as an *other*. A "list-like" description, such as *square, orange, dotty*, would also be coded as an *other*, since the modifiers *orange* and *dotty* follow the head noun but are not used in a grammatical relative clause construction. The proportion of pattern-first NPs produced to describe cards from the Pattern and Colour boxes in rounds 1 and 3 with each box was calculated by dividing the number of pattern-first NPs by the total number of pattern-first and colour-first NPs produced in that round. Analyses were computed on the *proportion* of pattern-first NPs rather than raw scores to allow a fairer comparison of conditions which contained unequal numbers of *other* responses (cf. Pickering, Branigan, & McLean, 2002). Reporting the proportion of pattern-first NPs (rather than the proportion of colour-first NPs) is arbitrary, since the proportions are complementary once *others* are excluded.

4.7 Results

4.7.1 Errors

An error was counted whenever a card placed on the grid was not the card that the Director was supposed to have described for that cell. The overall error rate was 19% (146 errors out of 768 card placements). Almost two thirds of these errors were due to confusions between yellow and orange cards, which looked quite similar on the Director's computer screen. When these errors were excluded, the error rate dropped to 7% (53 errors). Directors, overall, were less accurate than Matchers; 68% of errors (99 errors, out of 146) were due the Director giving inaccurate descriptions, while 23% (34 errors) were due to the Matcher selecting the wrong card or placing it in the wrong position on the grid. The remaining 9% (13 errors) were due to the Matcher being unable to find a particular card in the box. This usually came about because the Director had mis-described a previous card, or because the card they were looking for had accidentally been moved to a different part of the box.

4.7.2 Baseline word order preferences

Averaging across all of the NPs produced in the first game (including those coded as *others* for the main analysis), pattern was mentioned before colour in 39.7% of responses. This figure is not significantly lower than chance would predict ($t(15) = 1.44$, $p = .171$), although it does suggests a small baseline "colour before pattern" bias for speakers in this experiment. Colour may be the most salient property of the cards in perceptual terms (Deutsch & Pechmann, 1982; Pechmann, 1989; Pechmann & Zerbst, 2002), so colour information might be available for linguistic encoding earlier than pattern information (Pechmann, 1989; Schriefers & Pechmann, 1988).

4.7.3 *Other* responses

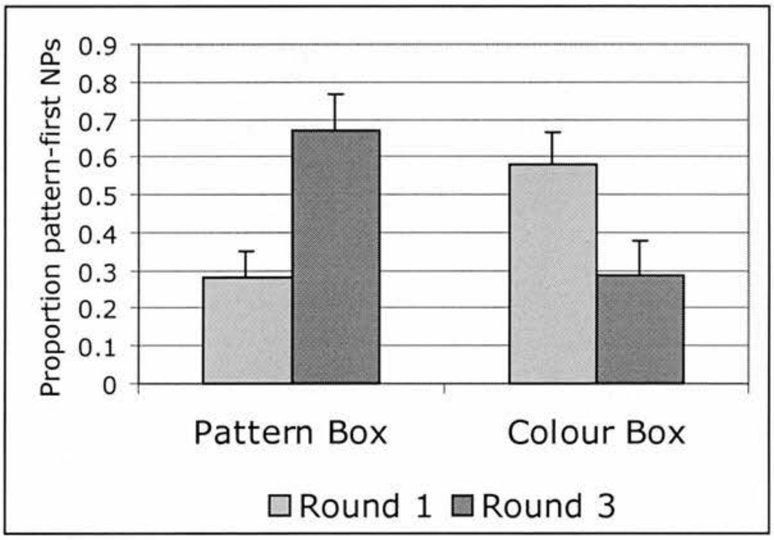
47 NPs out of 512 descriptions in total (or 9% of the data) were coded as *others* and were excluded from the analysis reported below. Of the 47 descriptions coded as *others*, 11 mentioned pattern before colour (23%), and the remaining 36 NPs (77%) mentioned colour before pattern. Once *others* were removed, two subjects had empty cells, which were replaced by the grand mean for the following analyses (although the results are almost identical even when *others* are included).

4.7.4 Is word order designed to be optimal for the addressee?

Figure 9 shows the proportion of pattern-first NPs produced by player A in rounds 1 and 3 with each box of cards. In round 1, player A does not know how the Matcher's array is organised (at the start of the second half of the experiment, player A might have assumed that the box was organised the same way as when they previously played the role of Matcher, but in fact, the box of cards had been switched after game 3, without their knowledge). In round 3, however, A has just played the role of Matcher themselves, so they know how the array is structured.

FIGURE 9

Experiment 2: Proportion of pattern-first NPs produced by player A when describing cards from the Pattern and Colour boxes



These data were submitted to an analysis of variance with two within-participants factors: Array (Pattern Box versus Colour Box) and Round (Round 1 versus Round 3), and one between-participants factor (Presentation Order: Pattern Box first or Colour Box first). The results suggest that word order can be influenced by what the speaker knows about the addressee's perspective. Player A produced more pattern-first NPs in Round 3 than Round 1 when the Matcher was choosing cards from the Pattern Box. However, when the Matcher was choosing cards from the Colour Box, player A produced *fewer* pattern-first NPs in Round 3 than in Round 1 (Array x Round interaction: $F(1, 14) = 14.18$, $MSe = .129$, $p = .002$). In other words, speakers tended to produce descriptions that were optimal for the addressee with respect to word order. Once they had discovered how the cards were organised (by playing the role of Matcher themselves, in round 2 with each box,), they were more likely to produce adjective orders that could be mapped incrementally onto the array.

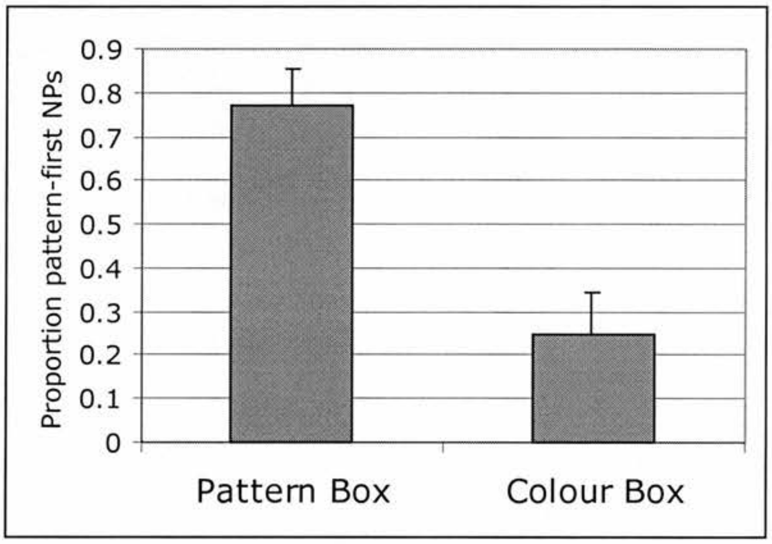
The interaction between Array and Presentation Order approached significance ($F(1, 14) = 4.14$, $MSe = .033$, $p = .061$). This seems to reflect differences between speakers' baseline word order preferences. In the first game, Directors whose partner had been given the Pattern Box showed a stronger preference for the colour-first word order than Directors whose partners had been given the Colour Box (mean proportion pattern-first NPs: 0.27 versus 0.51, respectively). However, this baseline difference between the groups was non-significant ($F(1, 14) = 3.77$, MSe = .07, $p = .072$) and is assumed to be due to chance. Presentation order was randomly assigned, and at this point in the experiment, the Director had no idea how the Matcher's array was organised (they had seen the box of cards that the Matcher was using, but the lid was closed). No other main effects or interactions approached significance (all Fs < 2.8).

4.7.5 What about player B?

In round 2 with each box, player B got the chance to play the role of Director. If word order is optimally designed, then we would expect player B to produce adjective orders that incrementally map onto the card box during these games, given that they already know (from first-hand experience in the previous round) how their partner's array is organised. Of course, unlike player A, we have no baseline data to make "before" and "after" comparisons. But what we can sensibly do is compare the proportion of pattern-first NPs that player B produces when they are describing cards from the two different arrays (the Pattern Box and the Colour Box). Figure 10 shows that player B produced more pattern-first NPs when they were describing cards from the Pattern Box (mean proportion = .77, $SD = .33$) than when they were describing cards from the Colour Box (mean = .25, $SD = .38$). This effect of Array is significant ($t(15) = 4.69$, $p < .001$).

FIGURE 10

Experiment 2: Proportion of pattern-first NPs produced by player B when describing cards from the Pattern and Colour boxes



4.7.6 Do helpful descriptions make Matchers quicker?

So far we have seen that speakers tend to produce noun phrases that "match" the array from which their partner is selecting picture cards, in the sense that the first property to be mentioned (pattern or colour) corresponds to the property that has been used to group the Matcher's cards at the top level of organisation. Word orders that match the array are assumed to be helpful, because they can be incrementally mapped onto the structure of the array, without the need to mentally rehearse or recode the description that has been heard. If mentally recoding or rehearsing a description takes time, then we might expect Matchers to be slower to pick cards out of the boxes after hearing unhelpful descriptions than after helpful descriptions. To find out, all of the noun phrases produced in Experiment 2 were recoded as "helpful" if they matched the way the Matcher's array was structured, or "unhelpful" if they did not. The proportion of helpful descriptions produced per game was correlated against the time (number of seconds) that the participants took to complete that game. This is a fairly crude measure of how easily Matchers selected cards from the array, but if helpful descriptions have an appreciable effect, then there should be a negative correlation: the more helpful descriptions the Director produces, the less time the game should take.

Time to play the game and proportion of helpful descriptions were not significantly correlated ($r = -.112, p = .278$). There was a numerical trend in the predicted direction, but we cannot conclude from these data that helpful descriptions make the task easier for Matchers.

4.8 Discussion

In Experiment 2, speakers exploited word order flexibility to produce utterances that could be mapped incrementally onto their partner's array. These "incrementally helpful" noun phrases are optimally designed from the point of view of comprehension, because they can be interpreted without the need for any mental recoding or maintenance in short term memory. When the speaker knew that the addressee's task was to select picture cards from an array organised by pattern, they were more likely to produce NPs like *a dotty orange square*. When the speaker knew that the addressee was choosing cards from an array organised by colour, they tended to produce NPs in which the prenominal adjectives appeared in the opposite order (*an orange dotty square*). Speakers exploited word order even though this sometimes meant going against a small bias for colour adjectives to precede pattern adjectives. They tailored word order to their addressee's perspective even though this seemed to have no effect on how quickly Matchers were able to select cards from the box.

EXPERIMENT 3
Non-Canonical Noun Phrases

4.9 Rationale and Predictions

In English, adjectives usually appear prenominally, i.e., before the head noun that they modify. However, adjectives can also appear postnominally, if a relative clause construction is used (e.g., *a square that's orange and dotty*). Experiment 3 was designed to find out whether speakers are more likely to produce postnominally modified NPs when this would be helpful for an addressee. Using a similar design to Experiment 2, Matchers in a referential communication task selected picture cards from a box that was either organised by pattern or by shape at the top level. When the cards are grouped by shape, a noun phrase that mentions shape before pattern (e.g., *square that's dotty and* *orange*) is helpful, in the sense that it can be incrementally mapped onto the structure of array. In round 1, player A did not know how the card box was organised, but they got the chance to find out during round 2, when they took over as Matcher. In round 3, then, they could work out what would be a helpful description from their partner's perspective. If word order can be sensitive to optimal design, then we might expect player A to produce word orders that are incrementally helpful with respect to shape and pattern information. Postnominally modified NPs are rarely produced spontaneously in English (Cleland & Pickering, 2003; Kemmerer, 2000). This observation was borne out in Experiment 2, where only 1% of descriptions contained postnominally modified noun phrases (11 descriptions, out of 768). So, if speakers produced NPs like *square that's dotty and orange* more often when it would be helpful for the addressee to hear shape information before pattern information, we would have fairly good evidence for optimal design with respect to word order.

4.10 Method

4.10.1 Participants

32 volunteers from the University of Edinburgh's student community took part in pairs, for a small payment. Pairs were matched on the basis of availability, and were recruited for an unrelated language study in which they participated immediately prior to this experiment. None of the participants had taken part in Experiment 2, nor had they met their partner before the testing session began. All were native speakers of English. One dyad was all-male, 2 were male-female, and 13 were all-female.

4.10.2 Design

Pairs of participants played six games of a referential communication task, in two blocks of three rounds, as in Experiment 2. In one block, Matchers selected cards from a box in which picture cards were grouped by pattern at the top level, and then subdivided by shape. In the other block, Matchers selected cards from a box in which the cards were grouped together by shape at the top level, and then subdivided by pattern. In the first round with each box, the Director did not know how the Matcher's cards were organised. But by playing the role of Matcher themselves in the second round, they found out about the structure of the Matcher's array. So, by the third game (when they went back to the being the Director), they knew what would make for a helpful or an unhelpful description, from the Matcher's perspective. The order in which Matchers were given the Pattern and Shape boxes was counterbalanced, such that half of the groups got the Pattern Box for the first three rounds, and the other half started off with the Shape Box.

4.10.3 Materials

The same 64 picture cards used in Experiment 2 were employed here. Two sets of cards were arranged in separate card file boxes, as in Experiment 2. In the Pattern Box, large patterned dividers were used to file the cards according to pattern at the top level. Within each pattern group, smaller dividers organised the cards by shape. In the Shape Box, large dividers with shapes on them were used to group the cards by shape at the top level, and smaller patterned dividers sub-divided the cards according to pattern. In each group of four cards, the orange card was always placed at the front, followed by the turquoise, yellow and purple cards (in that order).

Randomised lists of the 64 possible items were used to create a unique set of instructions for each pair of participants. The first 48 items (8 cards x 6 rounds) in each list were used to create slide shows of the Director's instructions, using Microsoft PowerPoint. This time, the constraints placed on item lists were as follows: for every round, each shape and pattern was represented by two cards, such that the same pattern-shape combination did not occur more than once. For example, the Director saw two square cards and two dotty cards per round, but no more than one exemplar from the category *dotty square*. Within a round, cards with the same shape or pattern did not appear next to each other (a square card was never followed by another square card, and so on). Again, this constraint was designed to minimise priming effects from one utterance to the next. Slide shows were presented on a Dell Latitude laptop computer with a 14.1" screen (resolution 800 by 600 pixels per inch).

4.10.4 Procedure

The procedure was identical to that used in Experiment 2. At the end of the experiment, participants were asked to fill out a short questionnaire about their experiences as both a Director and a Matcher. The full questionnaire can be found in Appendix C.

4.10.5 Coding and analysis

The dependent measure of interest is whether speakers mention pattern before shape, or shape before pattern in their descriptions. NPs were coded as *pattern-first* if the pattern adjective was mentioned before the head noun denoting shape (e.g., *dotty orange square*), or *shape-first* if the shape noun was mentioned before the pattern adjective (e.g., *square that's dotty and orange*). Crucially, NPs were coded according to the *relative* order of mention for pattern and shape. For example, a NP like *orange dotty square* would be coded as pattern-first, even though pattern is not the first property of the picture card to be mentioned. Likewise, *orange square that's dotty* would be coded as shape-first, even though the head noun, *square*, is not the very first word in the noun phrase. NPs which were not "well-formed" were coded as *others*, following the same criteria used in Experiment 2. These *other* responses are excluded from the initial analysis reported below. As in Experiment 2, the results are reported in terms of proportion of pattern-first descriptions, but this is arbitrary since the proportions of pattern-first and shape-first descriptions are complementary once *others* are removed.

4.11 Results and Discussion

4.11.1 Errors

The overall error rate was 9% (72 errors out of 768 card placements). One third of the errors were due to confusions over yellow and orange cards, which again tended to look similar on the Director's computer screen. When these errors were excluded, the error rate dropped to 6% (48 errors). Directors, overall, were less accurate than Matchers; 47% of errors (34 out of 72) were caused by the Director giving inaccurate descriptions, while 32% (23) were due to the Matcher selecting the wrong card or placing it in the wrong position on the grid. The remaining 21% (15)

were due to the Matcher being unable to find a particular card in the box.

4.11.2 Baseline word order preferences

Averaging across all of the descriptions produced in the first game (including those coded as *others*), a card's pattern was mentioned before its shape in 93% of responses (119 out of 128 card descriptions). This is significantly more often than chance would predict ($t(15) = 6.695$, $p < .001$), which supports the intuition that speakers have a strong preference for prenominal adjective orders in English. Of the nine descriptions that included postnominal modification in round 1, eight were produced by just one participant. It seems that NPs constructions involving postnominal modification really are dispreferred relative to prenominal adjective orders, for most native speakers of English (Cleland & Pickering, 2003; Kemmerer, 2000).

4.11.3 *Other* responses

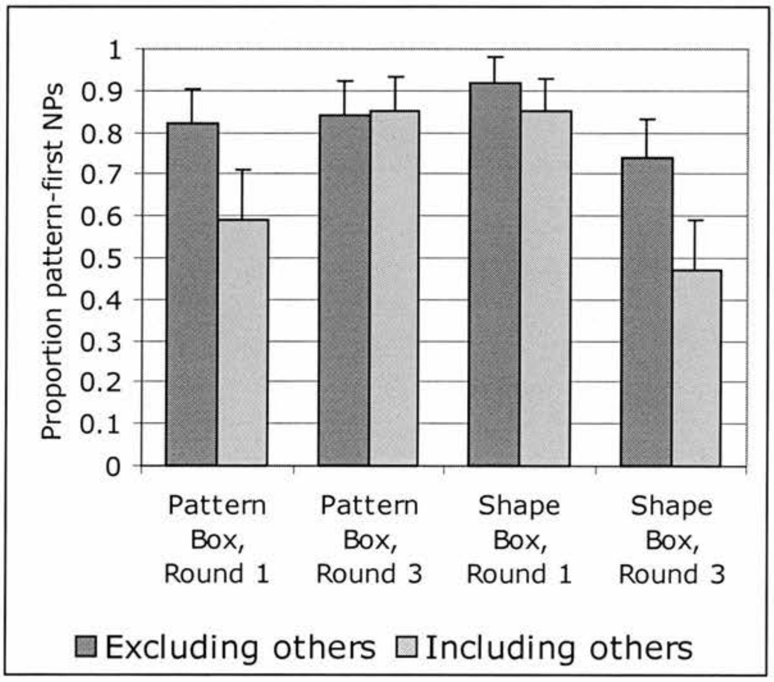
Descriptions coded as *others* accounted for 24% of the data overall (125 descriptions out of 512 produced in total) and are excluded from the initial analysis below. Six speakers were left with empty cells, which were replaced by the grand mean. Of the *other* descriptions, just 26% (33 out of 125) mentioned pattern before shape. To pre-empt the discussion in 4.11.4, this is clearly quite different from the baseline preference for pattern-first descriptions reported above. It looks as though descriptions coded as *others* are more likely to mention shape before pattern than descriptions that are included in the main analysis.

4.11.4 Is word order optimal for the addressee?

Figure 11 shows the proportion of pattern-first NPs produced by player A in rounds 1 and 3 with each box. For the initial analysis, responses coded as *others* were excluded. The remaining data (represented by the darker bars in Figure 11) were submitted to an analysis of variance with Array (Pattern Box versus Shape Box) and Round (Round 1 versus Round 3) as within-participants factors, and Presentation order (Pattern Box first or Shape Box first) as a between-participants factor. Numerically, the data followed the predictions of the optimal design hypothesis. That is, speakers tended to produce more pattern-first descriptions in round 3 (mean proportion = 0.84) than round 1 (0.82) when the addressee's array was organised by pattern, and *fewer* pattern-first descriptions in round 3 (0.74) than round 1 (0.92) when the addressee's box was organised by shape. But statistically, the interaction between Array and Round was not significant ($F(1, 14) = 2.32$, $MSe = .068$, $p = .15$). This lack of an interaction appears to be explained by the fact that participants whose partners selected cards from the Pattern Box for the first half of the experiment actually produced *fewer* pattern-first descriptions in round 3 (mean proportion = 0.84, $SD = 0.06$) than round 1 (mean = 0.98, $SD = 0.35$), contrary to the predictions of optimal design.

FIGURE 11

Experiment 3: Proportion of pattern-first NPs produced by player A when describing cards from the Pattern and Shape boxes



In fact, this pattern appears to be related to the high proportion of descriptions coded as *others* here. As noted above, speakers showed an overwhelming bias for "pattern before shape" descriptions in the main analysis, but when we look at the proportion of *other* responses, quite a different pattern emerges. Player A produced pattern-first *other* responses more often in round 3 (mean proportion = 0.72) than in round 1 (0.0) when the Matcher's box was organised by pattern. However, they produced pattern-first *other* responses *less* often in round 3 (0.02) than in round 1 (0.55) when the Matcher's box was organised by shape. This is interesting, because it suggests that player A was trying to be helpful with respect to word order, but that helpful descriptions tended to be realised as ungrammatical expressions, according to the coding criteria. So, perhaps the crucial interaction between Array and Round in the main analysis was obscured because so many responses had been excluded? To find out, a second analysis was conducted on the entire data set.

The paler bars in Figure 11 represent all of the descriptions produced by player A in rounds 1 and 3 with each box, including those coded as *others* for the initial analysis. These data were submitted to an analysis of variance, with the same factors described above. This time, the crucial interaction between Array (Pattern Box versus Shape Box) and Round (1 versus 3) was reliable ($F(1, 14) = 15.58$, $MSe = .105$, $p = .001$). No other main effects or interactions were significant (all $Fs < 1.9$). So when responses coded as *others* are included in the analysis, the predictions of optimal design are borne out. Directors produced more pattern-first descriptions in round 3 (mean proportion = .85, $SD = .34$) than in round 1 (mean = .59, $SD = .48$) when the Matcher's box was organised by pattern, but they produced *fewer* pattern-first descriptions in round 3 (mean = .47, $SD = .46$) than in round 1 (mean = .85, $SD = .31$) when the array was organised by shape.

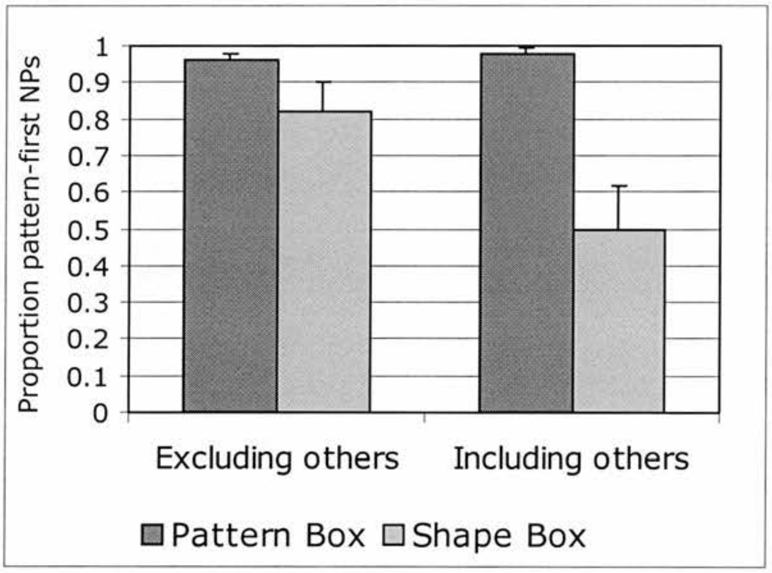
It looks as though speakers do design utterances to be cooperative with respect to word order under some circumstances. Once the Director had learned how the Matcher's cards were organised (by playing the role of Matcher themselves in round 2 with each box), they produced word orders that could be incrementally mapped onto the array. This effect is particularly striking in Experiment 3, because speakers sometimes produced non-canonical phrases like *square that's dotty and orange* in order to be helpful. Often they produced list-like expressions (e.g., *square, orange, dotty*) that would be considered non-standard or only marginally acceptable in most linguistic theories. It is interesting to speculate as to why speakers produced so many *others* in Experiment 3 (24% of the data were coded this way). Perhaps it simply did not occur to speakers to use a relative clause construction, like *square that's dotty and orange*, when the Matcher's array was organised by shape. Or maybe forfeiting grammaticality was the net result of trading off between two competing goals: being helpful with respect to word order, and producing timely contributions, so as not to keep the Matcher waiting

(Clark, 2002; Clark & Wilkes-Gibbs, 1986; Schober & Brennan, 2003).

4.11.5 What about player B?

FIGURE 12

Experiment 3: Proportion of pattern-first NPs produced by player B when describing cards from the Pattern and Shape boxes



As in Experiment 2, player B was the Director for the second round with each box of cards, to give player A a chance to find out how the Matcher's array was structured. Figure 12 shows the proportion of pattern-first NPs produced by player B when they described cards from the Pattern and Shape boxes. When *others* are excluded, player B produced about 14% more pattern-first NPs when the array was organised by pattern than when the array was organised by shape, but this difference was not significant ($t(15) = 1.620$, $p = .126$). When *others* are included in the analysis, player B mentioned pattern before shape 48% more often, on average, when the array was organised by pattern (Pattern Box: mean proportion = .98, SD = .07; Shape Box: mean = .50, SD = .46). This difference was very reliable ($t(15) = 4.03$, $p = .001$). These results show the same pattern as those for player A. Speakers did produce helpful word orders, but often they sacrificed grammaticality in order to do so.

4.11.6 Questionnaire data

At the end of the communication task, participants were asked to fill out a questionnaire that was designed to find out how aware they were of the experimental manipulation, whether they used word order strategically to help the addressee, and whether they noticed their partner using word order in a helpful way. The full questionnaire can be found in Appendix C. Out of the 32 participants, 23 (72%) mentioned word order, or the order in which information was mentioned in a description, when they described what they thought the experiment "might be about". 23 participants (72%) wrote that they had made the task easier for the Matcher by using helpful word orders once they knew how the box of cards was organised. A further 7 participants (22%) wrote that they could have used word order helpfully but did not, or that they could have been more consistent in producing incrementally helpful descriptions. 14 participants (44%) wrote that when they were the *Matcher*, their partners had made the game easier by using word order helpfully, and a further 10 participants thought that their partner could have made the game easier by adapting their descriptions to match the structure of the array.

Only one pair of participants did not mention word order (or the order in which pattern and shape information was conveyed) at all in their responses. These participants had had problems understanding each other's way of describing locations on the 3 x 3 grid. One of them had consistently used *row* to refer to what her partner (along with most other participants) called a column, and *column* to refer to what her partner called a row. For example, she described the bottom left-hand corner of the 3 x 3 grid as *the first row, third column*. So, their questionnaire responses focused on their different "description schemes" for the grid (Garrod & Anderson, 1987).

The fact that so many participants could verbalise how they tried to make the game easier for an addressee suggests that optimal design in Experiment 3 (and perhaps Experiment 2) may be under conscious control, at least for some speakers. Many of them claimed to be deliberately choosing word orders based on what they knew about the Matcher's array. In other words, they were optimally designing their descriptions in an entirely conscious, strategic way.

4.12 General Discussion

4.12.1 Word order seems to be sensitive to optimal design

Experiments 2 and 3 suggest that a speaker's choice of word order is sensitive to optimal design. In Experiment 2, speakers tended to produce pattern and colour adjectives in the order that would be most helpful for their partner to hear. When they were describing picture cards that the Matcher would have to select from an array organised by pattern, they were more likely to mention pattern before colour (*e.g., dotty orange square*). Conversely, they were more likely to mention colour before pattern (*orange dotty square*) when they knew that the Matcher's array was organised by colour. In Experiment 3, speakers tended to mention a card's shape before its pattern when the addressee was selecting cards from an array organised by shape. This finding is particularly striking, because mentioning shape before pattern required speakers to use non-canonical word orders which are typically dispreferred in English. For example, they sometimes produced postnominally-modified NPs (e.g., *square that's dotty and orange*) but more often, they adapted word order by forfeiting grammaticality altogether and producing list-like expressions (*square, dotty, orange*).

Crucially, speakers were more likely to produce "optimal" word orders once they knew from first-hand experience (being the addressee themselves) how the Matcher's array was organised. Taken together, the two experiments suggest that speakers do put the most "incrementally informative" word first (cf. Oller & Sales, 1969; Danks & Glucksberg, 1971; Danks & Schwenk, 1972), but only once they know what would be most helpful, from the addressee's point of view. Participants exploited word order flexibility to produce descriptions that could be easily mapped onto the array that their partner was working with. So, speakers in these experiments did seem to be behaving cooperatively with respect to choice of word order (cf. Branigan et al., 2003). To put the results in more general terms, it looks as though grammatical aspects of utterance production can be sensitive to pragmatic considerations, such as what the speaker knows about the addressee's perspective.

This is surprising because, going back to the journey metaphor introduced at the start of the chapter, optimal design with respect to syntax and grammatical choice is not normally necessary for successful communication. Word order in particular might have consequences for the route to comprehension that an addressee takes (i.e., how easily they understand an utterance), but it is unlikely to affect their ultimate interpretation. So why were speakers going out of their way to be so helpful in these experiments? Perhaps word order was more relevant to successful communication than it is in everyday conversation. The way the cards were grouped together in the two boxes might have drawn speakers' attention to the fact that word order was likely to have an impact on the ease with which their descriptions could be understood. Indeed, the questionnaire data from Experiment 3 backs this up; participants were very aware of the experimental manipulation, and of what would make for a "helpful" description. But even unhelpful word orders did not cause major comprehension difficulties for Matchers, who selected cards just as quickly after hearing unhelpful descriptions as after helpful ones. Hearing an unhelpful word order might make the task of selecting cards a little more laborious, but it doesn't seem to have major consequences for how efficiently Matchers get the job done.

Several other studies investigating optimal design at the level of grammatical encoding have found little evidence that addressee needs influence speakers' syntactic choices (e.g., Branigan et al. 2003; Brown & Dell, 1987; Dell & Brown, 1991; Ferreira & Dell, 2000). With respect to word order in particular, Branigan et al. (2003) found no evidence that speakers order discourse entities according to what is Given and New for an addressee. Why were speakers in their experiments so uncooperative?

It may be that speakers do not bother to adapt to the addressee's perspective until they have been on the receiving end of cooperative or uncooperative language behaviour themselves. Participants in Branigan, McLean and Reeve's studies were randomly assigned either to the role of Describer or Matcher, and they played that role throughout the experiment. So speakers never had the chance to find out what it was like to listen to and act upon someone else's descriptions. Matchers may have found passive descriptions more difficult or more laborious to understand than actives when both of the entities in the sentence were New (F. Ferreira, 2003), but Describers had no way of knowing this, at least not from their own experience in the experimental task. Perhaps speakers need some kind of direct experience as an addressee in a particular language situation to work out what would be optimal, from the point of view of comprehension (cf. Horton & Gerrig, 2002). In Experiments 2 and 3, participants had the opportunity to be both the Director and the Matcher, and this might explain why speakers were able to use word order in an addressee-optimal way.

Participants in Experiments 2 and 3 might also have been more aware (than Branigan et al.'s speakers) of how they could exploit word order for the addressee's benefit. Presumably, the card box manipulation makes it easy for speakers to see (quite literally) what would be most helpful from the Matcher's perspective, and the questionnaire data from Experiment 3 bear this out. Almost all of the participants mentioned word order (directly or indirectly) in their answers. For participants in Branigan et al.'s study, it was almost certainly more difficult for speakers to work out the connection between word order and what would be helpful for an addressee, particularly since they had no experience as Matchers themselves, as noted above. Indeed, when Branigan et al.'s participants were asked (in a post-experiment questionnaire) whether they had done anything in particular to make the Matcher's job easier, they focused on lexical aspects of the task, such as using the same words as their partner to refer to the entities in the pictures (see Experiment 6 for a similar finding). Awareness (whether explicit or implicit) of what makes for a "helpful" or an "unhelpful" description may be crucial for optimal design in this kind of goal-directed task.

4.12.2 Are speakers really tailoring word order "for" the addressee?

In terms of interpreting the word order effects found in Experiments 2 and 3, there is an alternative explanation that has nothing to with speakers taking an addressee's perspective into account. The alternative explanation goes something like this: maybe speakers mentioned pattern or colour or shape early in their descriptions because that particular property of the cards had been made salient. After all, Directors discovered how the array was organised by playing the role of Matcher themselves in an immediately preceding game. They selected eight cards from the array during their stint as Matcher, during which time either the pattern dividers, or the colour dividers, or the shape dividers (depending on which box they were working with) were visible throughout. Perhaps this made pattern, or colour, or shape conceptually more accessible than other properties of the cards, when they went back to playing the role of Director. Under this account, word order choice is simply driven by what is most accessible to the speaker. Suddenly we're back at an egocentric view of production that has nothing to do with what is optimal for the addressee. Speakers generate word orders that happen to be optimal from the point of view of comprehension, but this is just serendipity – a convenient by-product of accessibility-based production (cf. Brown & Dell, 1987; Dell & Brown, 1991; Garrod & Anderson, 1987; Horton & Keysar, 1996; Pickering & Garrod, in press).

Moreover, if word order is driven by the *conceptual* accessibility of some aspect of the picture cards, then the data do not speak to whether optimal design can influence *grammatical* aspects of production directly. Activation of some property (be it pattern, or shape, or colour) at the conceptual level simply makes that information available for grammatical encoding earlier than the other properties of the card. We cannot ascribe effects to the level of grammatical encoding *per se*.

A second issue that deserves some attention here is the extent to which *priming* contributed to the word order effects in Experiments 2 and 3. We know that speakers have a tendency to perseverate in producing expressions and linguistic structures (e.g., Bock, 1986; Branigan, Pickering, & Cleland, 2000; Cleland & Pickering, 2003; Pickering & Branigan, 1998; see Chapter 5 for a review of priming phenomena in language production). Once a speaker has heard or produced a particular word order once (say, pattern before colour), they are probably more likely to repeat that word order in a subsequent utterance. The usual explanation for such priming effects is that the mental representation of one structure (say, the mental representation of the "pattern before colour" adjective order) is more activated than a possible alternative (the "colour before pattern" representation). In fact, current models would have difficulty accounting for priming effects on prenominal adjective order (see Chapter 5 for discussion).

In Experiments 2 and 3, priming could have contributed to the helpfulness effect both directly (Player A was primed to produce pattern-before-colour descriptions in round 3 after hearing Player B produce similar descriptions in round 2) and *indirectly*, via the mental recoding of unhelpful descriptions. Imagine that Player B produces mostly unhelpful descriptions during round 2; for example, she uses colour-before-pattern adjective order in six out of eight of her descriptions, but the Matcher's box is organised by pattern. Player A may mentally transform those unhelpful descriptions in order to find the cards he is looking for. For example, if player B produces a description like *purple wavy star* to describe a card from the pattern box, player A may recode that word order as *wavy purple star* while locating the card in the box. This recoding activates the mental representation of the pattern-before-colour word order, which makes it somewhat more likely that player A will use pattern-first word order when he goes back to being the Director in round 3. In effect, player A has *primed himself* to produce helpful descriptions in the next round. This issue of priming is clearly important; if player A is primed to produce helpful word orders by hearing both helpful *and* unhelpful descriptions from player B in the previous round, then the optimal design effect is likely to be inflated. What look like adaptations to the addressee's perspective could come about without reference to the addressee, via automatic processes of alingment (Pickering & Garrod, in press).

However, the questionnaire data from Experiment 3 strongly suggest that the optimal design effects in these experiments are unlikely to be due simply to either the salience of some property of the cards, or to priming. Responses to the questionnaire suggested that many of the Directors had deliberately tried to use word order in a cooperative way, once they knew (from their own experience as a Matcher) what would make for a helpful description for the addressee. In fact, in describing what they had done to make the game easier for the Matcher, one participant wrote: *"[I] tried to say them in the order you would look them up in, but it didn't really sound/come naturally and so I sometimes forgot."*. If speakers are consciously using word order strategically, then it is unlikely that word order effects can be explained simply by priming, or by what is salient to the speaker. Instead, speakers seemed to be deliberately tailoring descriptions based on what they knew about the addressee's perspective. Of course, speakers' post-hoc reflections on their own language behaviour do not necessarily reflect the *causes* of that behaviour. Put another way, just because participants *said* they used word order deliberately, this doesn't necessarily mean that conceptual-level priming did not have some part to play in determining word order. So far, then, we cannot rule out the alternative, accessibility-based explanation for the data.

Very occasionally, speakers produced false starts, as in *dotty, I mean, orange dotty square* in which they overtly "repaired" a description that had started off unhelpfully (cf. Eikmeyer & Ahlsen, 1998). These false starts were clearly reflecting conscious processing. Directors began to speak, realised that they had begun to say something unhelpful, and then corrected themselves (Levelt, 1983). Like the questionnaire data, these overt repairs suggest that speakers were strategically tailoring word order to the addressee's perspective. This supports the idea that an alternative account based on either priming or salience at the level of conceptualisation is unlikely to fully account for the word order effects reported here.

4.13 Chapter Summary

Experiments 2 and 3 suggest that grammatical aspects of utterance production (specifically, word order) can be sensitive to optimal design, at least under some circumstances. This supports Lockridge and Brennan's (2002) claim that knowledge of a partner's perspective shapes syntactic choices. Alternative explanations based on priming or conceptual salience seem to be challenged by off-line questionnaire data suggesting that speakers strategically engaged in tailoring word order to the addressee's perspective.

CHAPTER 5

Priming, Optimal Design, and Unhelpful Interlocutors

5.0 Chapter Overview

In Chapter 4 we saw that speakers spontaneously tailor the grammatical form of noun phrases according to the addressee's needs. Untrained speakers in a simple referential communication task chose word orders and syntactic structures that could be incrementally mapped onto an array. Optimal design, then, seems to have an impact at the level of grammatical formulation, at least under some circumstances. In this chapter we turn to the interaction between optimal design and another process known to influence grammatical choices: syntactic persistence, or priming (Bock, 1986). Experimental studies show that speakers can be primed to produce different syntactic structures during dialogue tasks; in other words, hearing a partner produce a particular grammatical form (such as a passive, e.g., *the boat carried five people*) increases the likelihood that a speaker will produce the same kind of structure (in this case, another passive, such as *the boy was woken by the alarm clock*) in their next utterance (e.g., Branigan, Pickering, & Cleland, 2000). So, syntactic priming refers to the fact that production of a particular syntactic structure is facilitated if the same or similar syntactic structure has recently been processed. Syntactic priming has received a good deal of attention in the production literature (e.g., Bock, 1986; Bock & Griffin, 2000; Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992; Branigan, Pickering, & Cleland, 1999; 2000; Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995; Chang, Dell, Bock, & Griffin, 2000; Cleland & Pickering, 2003; Hartsuiker & Kolk, 1998; Hartsuiker, Kolk, & Huiskamp, 1999; Hartsuiker & Westenberg, 2000; Pickering & Branigan, 1998; Potter & Lombardi, 1998), and seems to be a robust phenomenon. The aim of the two experiments reported in this chapter was to investigate the relationship between optimal design and priming in speech production.

5.1 Repetition and Priming Effects in Production

Languages are flexible; there is no one-to-one mapping between meaning and form. So, for every utterance they produce, speakers are faced with choices: choices about how to conceptualise an object or idea, which words they will use to describe it, the syntactic structures those words will enter into, how to actually articulate the message they have formulated (quickly, slowly, quietly, loudly), and so on. Given all of these choices, it is hardly surprising that language shows immense variability. Nonetheless, it has been noted that there is much less variability *within* conversational exchanges than *between* conversations (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987; Garrod & Doherty, 1994). When two or more people start talking to each other, they begin to repeat things (e.g., Schenkein, 1980; Tannen, 1987, 1989). Much of this repetition is lexical (interlocutors often end up using the same words and expressions to refer to things, as we saw in Chapter 3), but corpus analyses show that syntactic structure is also repeated. For example, one corpus study showed that around 70% of utterances with a passive structure were preceded by another passive in the previous five utterances (Weiner & Labov, 1983). Likewise, Levelt and Kelter (1982) demonstrated that answers to simple questions often show repetition of surface form. When Dutch shopkeepers were asked *What time do you close?* they tended to answer with (something like the Dutch equivalent of) *Five o'clock*. But when they were asked *At what time do you close?*, they tended to give "syntactically congruent" answers like At *five o'clock* (Pickering & Branigan, 1999).

So, it looks as though the form of one utterance can 'prime' the form of a subsequent utterance. But on the basis of the corpus evidence, we cannot be sure what causes these repetition effects. The utterances that shared syntactic structure in Weiner and Labov's (1983) analysis were also similar in other ways; for example, they tended to use the same words, and they shared the same metrical structure. Similarly, the questions and answers studied by Levelt and Kelter (1982) either both contained the preposition *at*, or neither contained *at*. It is possible that the shop keepers were simply primed to produce (or not produce) a preposition. However, experimental evidence collected by Bock and colleagues suggests that syntactic priming effects obtain even when other explanations (such as lexical, conceptual and metrical overlap between one utterance and another) are ruled out.

The first experimental demonstration of syntactic priming came from a picture description task disguised as a memory experiment (Bock, 1986). Participants alternately repeated prime sentences aloud (which were auditorily presented) and described "target" pictures. Bock found that participants were more likely to produce a passive target description (*the boy was woken by the alarm clock*) after repeating a passive prime (e.g., *the referee was punched by one of the fans*) than after repeating an active prime (*one of the fans punched the referee*). Likewise, they were more likely to produce a "double object" dative construction after repeating a prime that also contained a double object (*a rock star sold an undercover agent some cocaine*), than after a prime that contained a "prepositional object" (*a rock* *star sold some cocaine to an undercover agent*).

Given that the events depicted in the target pictures were semantically unrelated to the prime sentences, Bock argued that the priming effect could not only be due to semantic factors. Other studies have ruled out other counter-explanations based on prime and target sharing the same words (Bock, 1989), metrical properties and thematic structures (Bock & Loebell, 1990; Bock, Loebell & Morey, 1992). So, it seems as though the syntactic structure of one sentence or utterance primes the syntactic structure of a subsequent sentence. This is true not only in picture description, but also in sentence completion (e.g., Hartsuiker & Westenberg, 2000; Pickering & Branigan, 1998), and sentence recall tasks (e.g., V. Ferreira, 2003; Potter & Lombardi, 1998).

The experimental evidence cited above is based on *monologue* tasks: participants repeat, recall, or read aloud a prime sentence, or they complete a prime fragment in their own way (e.g., Pickering & Branigan, 1998). Then they describe a target picture, or complete a target fragment, and the syntactic form of their target production is what is analysed. Crucially, the participant has produced both the prime and target. But other evidence demonstrates that priming also occurs in *dialogue* tasks, where one interlocutor produces the prime and a second interlocutor produces the target. Branigan, Pickering and Cleland (2000) had pairs of participants play a communication game in which they took turns to describe picture cards to their partner, and select picture cards from an array on the table in front of them. On critical trials, the pictures showed characters like pirates, doctors and ballerinas involved in various ditransitive events. These critical pictures could be described using either a prepositional object (PO) construction (*the pirate showing the book to the monk*) or a direct object (DO) construction (*the pirate showing the monk the book*). One player was a confederate who read their prime descriptions from a script. Half of the primes contained PO and half contained DO structures. Naïve participants tended to use the same syntactic structure (a PO or a DO) for their target description as they had just heard the confederate produce as a prime. For example, participants produced more DO descriptions after hearing a DO prime than after a PO prime. The priming effect was observed even when prime and target involved different verbs, but it was bigger when they both used the same verb.

5.2 Priming Noun Phrase Structure

All of the priming studies discussed so far have focussed on priming for verb phrase structure, but recent evidence suggests that priming also occurs within noun phrases. Cleland and Pickering (2003, Experiment 1) exploited the same dialogue technique as Branigan et al. (2000) to show that speakers could be primed to produce noun phrases like *the triangle that's green* in a picture description task. Naïve participants interacted with a confederate who was scripted to produce prime descriptions like *the red square* or *the square that's red*. The relationship between the lexical content of the confederate's prime description and the content of the participant's target description was manipulated. Sometimes both noun and adjective were shared between prime and target (e.g., the confederate said *the red square* and the participant described a picture of a red square on the following turn). Alternatively, only the noun was shared, or only the adjective was shared, or there was no lexical overlap at all.

Participants produced about 19% more target descriptions like *the triangle that's green* after hearing primes like *the square that's red* than after primes like *the red square*. Put another way, participants were more likely to produce NPs containing relative clauses after they heard the confederate produce a noun phrase with the same structure. The priming effect was largest when both prime and target contained the same adjective and head noun. When there was *no* lexical overlap between prime and target, participants produced about 10% more post-nominally modified noun phrases (*the triangle that's green*) following a prime with the same structure than after a prime with the alternative form. So, the priming effect was larger when primes and targets used the same words. This is consistent with Branigan et al.'s (2000) finding that shared lexical content leads to bigger priming effects in dialogue. However, even when primes and targets had no lexical content in common, speakers were more likely to repeat the syntactic structure of the utterance they had just heard from an interlocutor than to produce the alternative form.

Given that noun phrase structure appear to be sensitive to both priming (Cleland & Pickering, 2003) and optimal design (see Chapter 4), then this seems to be a good place to investigate the relative influence of these two processes on spoken language production, and in particular, on the grammatical form of referring expressions. Using the same kind of paradigm as Experiments 2 and 3, we should be able to set up a situation in which a speaker has been primed to produce a "colour-first" NP (*the green triangle*) but knows that their addressee would find a "shape-first" NP (*the triangle* *that's green*) easier to interpret, relative to the task at hand. What kind of noun phrase will the speaker produce? Experiments 4 and 5 were designed to find out.

EXPERIMENT 4
Priming, Optimal Design and Syntactic Form

5.3 Rationale and Predictions

In effect, Experiment 4 pitted priming and audience design against each other in order to investigate the relative influence of each process on the form of referring expressions produced in a goal-directed task. A referential communication paradigm similar to the one used in Experiments 2 and 3 was again employed. This time, one of the players was a confederate, who was scripted to produce particular kinds of noun phrases (cf. Branigan et al., 2000; Cleland & Pickering, 2003). The players took turns to describe and select coloured wooden blocks (squares, triangles, circles, etc.) from an array. The kind of noun phrase produced by the confederate on each turn was manipulated; under some circumstances she produced noun phrases with prenominal modification (e.g., *the red square*) and under other circumstances she produced postnominally modified NPs (e.g., *the square that's red*). In addition, the organisation of the array was also manipulated. Players selected wooden blocks from two small chests of drawers. In one chest (the *Colour Box*), the blocks were grouped according to colour. In the other (the *Shape Box*), the blocks were grouped according to shape.

Following the same logic that motivated Experiments 2 and 3, finding a block from the "Shape" box should be easiest following a description like *the square that's red*, because the order in which information about the block is conveyed (shape before colour) matches the way the array is organised. The drawer containing the square blocks has to be opened before the *red* square block can be found. In other words, a noun phrase like *the square that's red* is "incrementally helpful" in a way that *the red square* is not. *The square that's red* is therefore optimal for comprehension in this case because the addressee can find its referent without holding the description in working memory, or mentally recoding it such that it can be incrementally mapped onto the structure of the array.

If speakers design noun phrases "optimally" with respect to grammatical form, then they should be more likely to say *the triangle that's green* when they are describing blocks from the Shape Box than when they are describing blocks from the Colour Box, regardless of the kind of prime they have just heard from the confederate. Conversely, if noun phrase structure is influenced more by priming than by optimal design, then naïve participants in the game should be more likely to produce a noun phrase like *the triangle that's green* following a prime like *the square that's red*, regardless of the box from which they are describing a block.

5.4 Method

5.4.1 Participants

16 volunteers (4 males, 12 females) from Edinburgh University's student community took part for payment. All were native speakers of English, and none had taken part in Experiments 2 or 3.

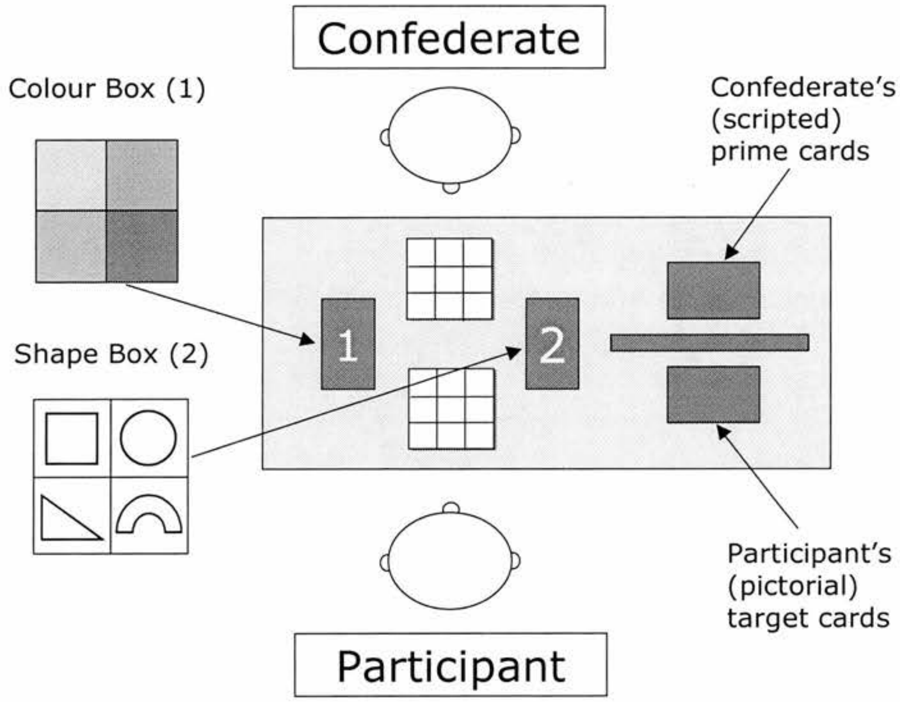
5.4.2 Materials and Design

In this experiment, players took turns to give and follow instructions about where to place small wooden blocks on a 5 x 4 cardboard grid. Both players (the confederate and the naïve participant) were given an empty grid at the start of the game. The columns of the grid were labelled with numbers (1-5) and the rows were labelled with letters (A-D). 2 identical sets of 16 wooden blocks were used in the experiment. In each set, there were four square blocks, four circles, four triangles and four bridge shapes. One of each shape was red, one was blue, one was green and one was yellow. The two sets of blocks were arranged in separate cardboard chests of drawers, which were labelled '1' and '2' on the top (see Figure 13 below). Each chest contained 4 pull-out drawers. In the *Colour Box* (chest '1'), the drawers were labelled with coloured patches to indicate that one drawer contained all the red blocks (one square, one circle, one triangle, one bridge), one contained all the blue blocks, and so on. The drawers in the *Shape Box* (chest '2') were labelled with pictures to indicate that one contained all the square blocks (one red, one blue, one green, one yellow), one contained all the circle blocks, etc. Individual blocks were labelled with numbered stickers to show whether they came from (box) 1 or (box) 2.

Two sets of instruction cards were created for the players to describe. The confederate's instruction cards (the prime set) had written instructions printed on them, so that she could simply read each one aloud. The confederate gave her scripted instructions in three prosodic instalments, specifying the box, the target block, and a location on the grid. For example, the confederate might read aloud a description like *It's Box 1, it's the red triangle, and it goes in 3B*. The confederate's descriptions were 'helpful' in the sense that she always used a *colour-first* prime NP (e.g., *the red square*) to describe a block from the Colour Box and a *shape-first* prime NP (e.g., *the square that's red*) to describe a block from the Shape Box. The participants' instructions (the target set) showed a pictorial representation of the 5 x 4 grid with a wooden block in one square. The picture also indicated which box the block should come from (1 or 2). Half of the participants' instructions showed a block from Box 1, and half showed a block from Box 2.

FIGURE 13

Experiment 4: Overhead view of the experimental set-up



32 pairs of instruction cards were created by pairing a prime card from the confederate's set with a target card from the participant's set. Pairs were constructed so that there was no lexical overlap between the prime NP and the target NP. In other words, the colour and shape of the target

block were always different to the colour and shape of the block in the immediately preceding prime. For example, a red prime block would never be followed by a red target, a square prime would never be followed by a square target, and so on. As in Experiments 2 and 3, this constraint was designed to reduce priming effects driven by the accessibility of repeated lexical items (cf. Branigan, Pickering, & Cleland, 2000; Cleland & Pickering, 2003; Pickering & Branigan, 1998).

An experimental item was defined as the confederate's scripted prime NP description plus the participant's paired target card. A mistake in the experimental design meant that items were not fully rotated through conditions. So, in this experiment there were 32 unique items in total (see Appendix B), eight in each of four conditions:

1) Colour-first prime NP, participant describes target block from the Colour Box
2) Colour-first prime NP, participant describes target block from the Shape Box
3) Shape-first prime NP, participant describes target block from the Colour Box
4) Shape-first prime NP, participant describes target block from the Shape Box

Prime-target pairs were assigned to one of two lists, which were presented as two separate blocks of 16 trials (or two 'games') during the experimental session. The two lists were constructed such that every wooden block from the array was described only once in each list, either as part of a prime or as part of a target. Within each list, the order of presentation of prime-target pairs was randomised for each participant.

5.4.3 Procedure

The confederate and participant were seated on opposite sides of a table, with the two chests of drawers containing the wooden blocks on the table between them (see Figure 13 for an overhead view of the experimental set-up). Instruction cards were placed to the side of the players in 8" x 6" card file boxes. Both players read instructions explaining the aim of the communication game: to fill up their empty grids with wooden blocks as quickly and as accurately as possible. They alternated between two tasks: describing instruction cards to their partner, and selecting wooden blocks to put on their own grid. The players were invited to look inside the chests of drawers at the start of the experiment, to find out what was inside. They were reminded that it was very important when giving instructions to provide enough information for the other player to find the right block from the correct box. The confederate always gave the first description. Both the confederate and the participant had four empty spaces on their grid at the end of a game. Between games, the experimenter put the wooden blocks back into the chests of drawers while the players filled out a short background questionnaire to collect some demographic information (age, courses studied at university, and so on). After the second game, participants filled out a second questionnaire, which was designed to assess how aware they were of the experimental manipulations, and whether they had used syntactic structure or word order in order to be helpful towards their partner. The experimental session lasted about 25 minutes in total. Each session was recorded on audio tape and later transcribed verbatim. Target NPs produced by participants were coded the as *colour-first* (e.g. *the green triangle*), or *shape-first* (e.g. *the triangle that's green*). Participants always mentioned both the colour and shape of the target block, so no responses were coded as *others* in this experiment, and every response was included in the following analysis.

5.5 Results

5.5.1 Effects of priming and optimal design

As noted above, a mistake in the experimental design meant that items were not fully rotated through conditions. The *F2* analyses reported below are therefore between-items comparisons. As a consequence, the data should be treated cautiously, since statistically significant results could in principle be due to the way items were randomly assigned to the different conditions. With this caveat in mind, Table 3 reports the proportion of colour-first NPs (*the green triangle*) produced in each condition. Reporting colour-first NPs rather than shape-first NPs is arbitrary, since the proportions are complementary (no responses were coded as *others* in this experiment).

These data were submitted to 2 x 2 analyses of variance, with Prime (Colour-first vs. Shape-first) and Array (Colour Box vs. Shape Box) as within-participants factors in the *F1* analysis, and between-items factors in the *F2* analysis. These analyses revealed a small but very reliable main effect of Prime ($F1(1, 15) = 7.94$, $MSe = .004$, $p = .013$; $F2(1, 28) = 12.29$, $MSe = .001$, $p = .002$). Participants produced 5% more colour-first target NPs (e.g. *the green triangle*) following a colour-first prime (*the red square*) than following a shape-first prime (*the square that's red*). In other words, speakers tended to repeat the same syntactic structure that their interlocutor had used in an immediately prior description. Descriptions also showed some evidence of optimal design. Speakers produced 13% more colour-first NPs when describing a block that their partner would have to select from the Colour Box. This main effect of Array was significant ($F1(1, 15) = 4.41$, $MSe = .064$, $p = .053$; $F2(1, 28) = 98.68$, $MSe = .001$, $p < .001$). The interaction between Prime and Array was significant by items and marginal by participants ($F1(1, 15) = 3.33$, $MSe = .005$, $p = .088$; $F2(1, 28) = 5.46$, $MSe = .001$, $p = .027$). This interaction seems to be driven by the fact that participants almost always produced colour-first descriptions when they described blocks from the Colour Box, such that priming had little effect.

TABLE 3

Experiment 4: Proportion of colour-first NPs (*the green triangle*) produced in each experimental condition (*Standard deviations* in brackets)

	Participant describes target block from:	
	Colour Box	Shape Box
Confederate produces:		
colour-first prime (*the red square*)	.99 (.03)	.89 (.03)
shape-first prime (*the square that's red*)	.98 (.07)	.81 (.03)

5.5.2 Questionnaire data

At the end of the referential communication task, participants filled in a questionnaire designed to find out how aware they were of the experimental manipulation(s), and of the confederate's language behaviour. Firstly they were asked to write down what they thought the experiment "might be about". Then they answered a series of more specific questions which aimed to gauge whether they were aware of their partner using word order in a helpful way, and whether they had used word order strategically in order to be helpful themselves. The full questionnaire can be found in Appendix D.

Six of the 16 participants (38%) explicitly mentioned word order (or the order in which colour and shape information was conveyed) when they described what they thought the experiment was about. Three participants (19%) wrote that they tried to use word order to make the game easier for their partner, and a further 5 (31%) wrote that they *could* have made the game easier for their partner by tailoring word order to the box from which they were describing a card. Ten out of the 16 participants (63%) mentioned that the confederate used different word orders for the two boxes. Table 4 provides examples of responses which explicitly mention (or do not mention) word order, or the information in which colour and shape information is conveyed.

Potentially, the questionnaire data give us an interesting way to look at whether explicit awareness is important for optimal design, and whether participants who *said* they behaved helpfully with respect to word order were actually doing so. We might expect participants who explicitly mentioned word order to produce more "helpful" descriptions than participants who did not. It seems problematic to test this statistically using these data, because on the basis of the questionnaire, only four people can be described as "unaware" of the optimal design manipulation. These four people did not mention word order at all in their responses (although of course we cannot assume that using word order in order to be helpful had not occurred to them during the game). Comparing participants who seemed to be "aware" of how word order might be used helpfully versus "unaware" participants would therefore involve looking for differences between two unequally-sized samples. However, it is interesting to take an exploratory look at patterns in the data.

Generally speaking, participants who produced fewer colour-first descriptions when they described blocks from the Shape Box (i.e., those who showed evidence of optimal design) also tended to explicitly mention word order in their responses to the questionnaire. Participants who hardly ever (or never) produced shape-first descriptions (or at least, not when they described blocks from the Shape Box) also tended to be the people who did not explicitly mention word order on the questionnaire. Participants' questionnaire responses did not always accurately reflect their linguistic behaviour during the communication game. For example, six participants who wrote that they had not done anything special to make the game easier for their partners had actually produced (numerically) fewer colour-first descriptions when they described blocks from the Shape Box than when they described blocks from the Colour Box, just as optimal design would predict. This suggests that there might be an imperfect relationship between participants' reflections on their own linguistic behaviour during the task, and how they actually performed with respect to optimally designing word order.

TABLE 4

Experiment 4: Sample responses to the post-experiment questionnaire

---

What was the experiment about?

Mentions word order:

I think it was maybe about the way we gave our instructions i.e. which order we said them in. For example: square that's yellow (for box 2) *or* *yellow square*

Doesn't mention word order:

To determine whether there is a difference between the sexes in communication.

Did you do anything special to make it easy for your partner to build up the grids?

Mentions word order:

I started saying the colour first then the shape if it was box one and the shape first then the colour if it was box two.

Doesn't mention word order:

Just spoke clearly.

Do you think you could have been more 'helpful' when you described your cards?

Mentions word order:

Said shape first on box 2.

Doesn't mention word order:

Spoke clearer?

Did your partner do anything special that made it easy for you to build up the grids?

Mentions word order:

Put shape or colour first for appropriate box.

Doesn't mention word order:

Not really

5.6 Discussion

The results of Experiment 4 replicate Cleland and Pickering's (2003) finding that noun phrase structure can be primed. Speakers were significantly more likely to produce shape-first NPs (*the triangle that's green*) following a shape-first prime from the confederate. Cleland and Pickering reported a small priming effect when there was no lexical overlap between prime and target (just 10% priming in their "different adjective/different noun" conditions in Experiment 1 – an effect that was only significant in the by-items analysis) So, although the 5% priming effect found here is small, it is roughly comparable with Cleland and Pickering's results. Since the confederate always produced "helpful" descriptions in Experiment 4, there is no issue about the naïve participant having to mentally recode unhelpful utterances (as they might have done in Experiments 2 and 3) in order to map them onto the array. Nonetheless, even though no recoding was necessary, the priming effect was numerically rather small here. This argues against a strongly automatic alignment process and in favour of a priming mechanism that is subject to other influences.

Experiment 4 also suggests that speakers designed noun phrase structure with their addressee's perspective in mind. They tended to produce more NPs like *the green triangle* when describing blocks from the Colour Box, and more NPs like *the triangle that's green* when describing blocks from the Shape Box. This is optimal for comprehension, because the addressee can incrementally map the noun phrase onto the array without having to do any extra mental "work" (such as recoding or rehearsing the description). So the results support those of Experiments 2 and 3. Speakers seem to tailor the grammatical form of referring expressions to their addressee's "needs", relative to the task at hand. In this sense, speakers can be said to be behaving co-operatively, particularly when producing a helpful word order means overcoming a usual tendency towards repeating the word order that their partner has just used (as is the case when prime and target are from different boxes). Speakers sometimes even produced NPs that are strongly dispreferred in spontaneous speech, like *the triangle that's green*, although they did so relatively rarely (cf. Experiment 3).

The effect of optimal design in Experiment 4 was rather small; speakers only produced about 13% more NPs like *the triangle that's green* when they described blocks from the Shape Box, even though the experimental manipulation was fairly salient (for example, 63% of participants were aware that the confederate was producing different kinds of noun phrase for the two boxes). Why might this be? One interesting possibility is that speakers attended to optimal design in other ways in this experiment. It's possible that rather than tailoring syntactic form to the addressee's task, speakers were using contrastive stress to emphasise the most informative part of the noun phrase (e.g., *the green TRIANGLE*). This conveys the fact that the head noun is the most "useful" part of the description when the addressee is trying to find blocks from the Shape Box, but without the speaker having to produce a strongly dispreferred syntactic form (*the triangle that's green*). This is an interesting idea, because there is some evidence that speakers do use stress contrastively, to emphasise the property of a referent that best distinguishes it from other objects in an array (Krahmer & Swerts, 2001; Oller & Sales, 1969; Pechmann, 1984). However, simply stressing the head noun is not optimal for the addressee, from the point of view of incremental comprehension. If the addressee has to find a block from the box that is organised by shape, then hearing a noun phrase like *the green TRIANGLE* is not that helpful; colour information will have to be maintained in memory while the triangle blocks are located. In this sense, contrastive stress is not good evidence for the kind of optimal design that we are concerned with in these experiments. For this reason, the issue of contrastive stress will not be considered further here.

In Experiment 4, the priming effect (5%) was numerically smaller than the effect of optimal design (13%). There was an interaction between the two which was reliable by items and marginal by participants. The interaction should be interpreted cautiously, for at least two reasons. Firstly, as noted above, a mistake in the experimental design meant that items were not rotated through conditions, such that *F2* analyses were computed between (rather than within) items. As a consequence, any significant effects could in principle be due to the (random) assignment of different items to the four conditions.

Secondly, the interaction seems to be driven by a ceiling effect; participants produced shape-first NPs (*the triangle that's green*) so rarely when they described blocks from the Colour Box that there was no effect of the confederate's prime. Presumably the ceiling effect reflects the fact that post-nominal modification is heavily dispreferred relative to prenominal modification in English, at least in short noun phrases like the ones investigated here (Cleland & Pickering, 2003; Kemmerer, 2000). To find out whether priming and optimal design interact in the absence of such a ceiling effect, we may need to look at their influence on constructions that are less heavily dispreferred. Prenominal adjective order in noun phrases like *dotty orange square* seems to be relatively flexible in English; Experiment 2 showed a very small bias for colour adjectives to precede pattern adjectives, but speakers were willing to override this bias when it would be more helpful for an addressee to hear pattern information before colour. Prenominal adjective order, then, seems to be a suitable place to further investigate the relative influence of priming and optimal design on grammatical encoding, and Experiment 5 returned to speakers' use of colour-first (*orange dotty square*) and pattern-first (*dotty orange square*) NPs in describing the picture card stimuli introduced in Chapter 3.

EXPERIMENT 5
Priming, Optimal Design, and Prenominal Adjective Order

5.7 Rationale and Predictions

In Experiment 5, participants again played a referential communication game with a confederate who was scripted to produce particular kinds of referring expression. Confederate and participant took turns to describe picture cards from two arrays. In one box, the cards were grouped by pattern at the top level and subdivided by colour at the lower level. In the other box, the cards were grouped by colour and then subdivided by pattern. Essentially, this was the same manipulation as in Experiment 2, but both players could see both of the boxes throughout the experiment, and they alternated between the roles of Director and Matcher on a card-by-card basis. If speakers are tailoring the form of their referring expressions in order to be incrementally helpful with respect to word order, then they should be more likely to say *orange dotty square* when describing cards from the Colour Box, and *dotty orange square* when describing cards from the Pattern Box. From Experiment 4 (and from Cleland and Pickering's results, 2003), we would also expect speakers to be primed by the word order of the confederate's preceding description. So, we would expect participants to produce more "pattern-first" descriptions like *dotty orange square* after a pattern-first prime (*wavy purple star*) than after a colour-first prime (*purple wavy star*). Since colour-first and pattern-first adjective orders are spontaneously produced with roughly equivalent frequency in English (see Experiment 2), this experiment gives us a way to further explore the relative influence of priming and optimal design on grammatical aspects of production, and to find out whether the interaction reported for Experiment 4 is reliable.

One final manipulation was added to Experiment 5. Rather than the confederate producing "helpful" word orders all of the time (as in Experiment 4), for half of the participants she consistently produced *unhelpful* word orders. For example, in describing cards from the pattern box, she always produced colour-first primes (*purple wavy star*). Social psychologists assume that interpersonal interactions are governed by a 'norm of reciprocity' (e.g., Gouldner, 1960; Perugini, Galluchi, Presaghi, & Ercolani, in press;); we tend to behave towards others as they behave towards us. Moreover, social cognition research suggests that people can be non-consciously primed to behave in certain ways: aggressively, intelligently, politely, and so on (e.g., Bargh, Chen, & Burrows, 1996; Chen & Bargh, 1997). For both of these reasons, we might expect participants in Experiment 5 to be less likely to engage in optimal design if their interlocutor is behaving in an "unhelpful" way towards them. In addition, there is some evidence that interlocutors who are not getting along are less likely to linguistically align with each other (e.g., Bly, 1993; Danet, 1980; Giles, Coupland, & Coupland, 1991; see Schober & Brennan, 2003). If participants in Experiment 5 actually *like* an unhelpful confederate less than a helpful one, then we might also expect priming effects to be reduced in the Unhelpful Confederate condition. Manipulating the confederate's use of word order, then, gives us an interesting way to investigate the effects of an interlocutor's language behaviour on optimal design with respect to reference.

5.8 Method

5.8.1 Participants

32 volunteers from Edinburgh University's student community were paid to participate. All were native speakers of English, and none had taken part in Experiments 2, 3 or 4.

5.8.2 Materials and Design

This experiment recycled the two sets of picture cards from Experiments 2 and 3. As before, the two sets were identical; each contained 64 cards, which represented every possible combination of four colours (orange, turquoise, yellow and purple), four patterns (dotty, chequered, dotty and wavy) and four shapes (square, circle, triangle and star). Thus each set contained one dotty orange square, one dotty orange circle, one dotty orange triangle, and so on. The sets of cards were arranged in two card file boxes. In the *Pattern Box*, labelled dividers were used to group the cards by pattern and then subdivide them by colour. In the *Colour Box*, similar dividers were used to group the cards by colour and then subdivide them by pattern. In the actual experiment, the Pattern and Colour boxes were marked '1' and '2' respectively, and numbered stickers on the individual cards showed which box they came from.

Two sets of instruction cards were created for this experiment. The confederate's instructions (the prime set) had written instructions printed on them, such that she simply had to read each one aloud from the card, e.g., *Box 2, wavy purple star*. Half of the confederate's descriptions contained *pattern-first* prime NPs (e.g., *wavy purple star*), and half contained *colour-first* prime NPs (e.g., *purple wavy star*). Two versions of the confederate's prime cards were made, a *helpful* set and an *unhelpful* set (see below). Instructions in the participant's set (the target set) showed which picture card they should describe to their partner on each turn. Half of the time they had to describe cards from the Pattern Box, and half of the time they had to describe cards from the Colour Box.

The participant and the confederate each had 64 instruction cards to describe, made up of 16 pairs of cards in each of four conditions. The four conditions were defined by crossing Prime Type (pattern-first vs. colour-first) and Array (whether the participant had to describe a target card from the Pattern Box or the Colour Box), yielding a 2 x 2 within-subjects design. An experimental item was defined as any combination of pattern and colour in the participants' target cards. For example, the dotty orange square, circle, triangle and star cards constituted one single 'dotty orange' item. Equal numbers of square, circle, triangle and star cards were randomly allocated to each condition. In total, then, there were 16 items, each of which appeared in four conditions:

1) Pattern-first prime, target card to be selected from the Pattern Box
2) Pattern-first prime, target card to be selected from the Colour Box
3) Colour-first prime, target card to be selected from the Pattern Box
4) Colour-first prime, target card to be selected from the Colour Box

An additional between-participants manipulation was added to the design; 16 participants interacted with a 'Helpful' confederate, and 16 interacted with an 'Unhelpful' confederate. Note that the same person played the confederate throughout the experiment; the only difference was in the script she was given to read in the Helpful and Unhelpful conditions. In the Helpful condition, the confederate was scripted to produce NPs in which the relative order of colour and pattern adjectives matched the organisation of the box from which she was describing a card. She always produced *pattern-first* prime NPs (*wavy purple star*) when describing cards from the Pattern Box and *colour-first* prime NPs (*purple wavy star*) when describing cards from the Colour Box. This is helpful behaviour in the sense that the naive participant can incrementally map the confederate's description onto the array from which they must select a card. In the Unhelpful condition, the confederate always produced colour-first primes when describing cards from the Pattern Box, and pattern-first primes when describing cards from the Colour Box. This is unhelpful in the sense that the naive participant cannot incrementally map the confederate's description onto the array. The description will need to be mentally recoded, or at least rehearsed in memory (see Chapter 4).

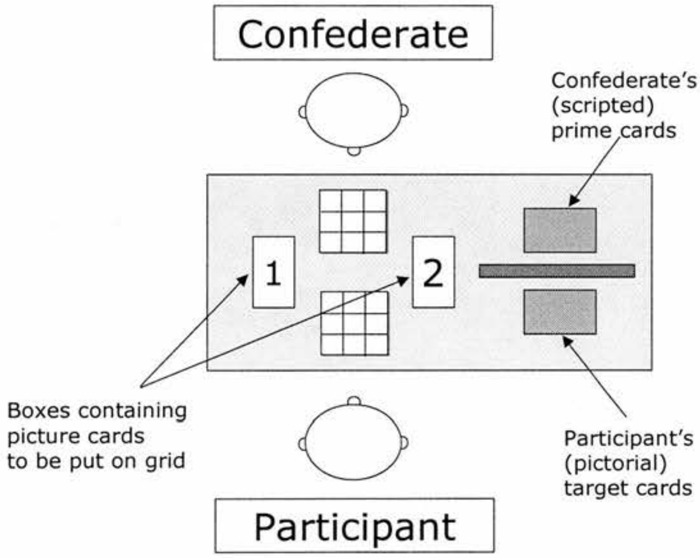
64 pairs of instructions were created by pairing a prime card from the confederate's set with a target card from the participant's set. There was no lexical overlap between the confederate's prime NP and the participant's immediately subsequent target NP. For example, a wavy prime card would never be followed by a wavy target, a purple prime would never be followed by a purple target, and so on. Pairs of instruction cards were randomly allocated to one of four lists, which were presented as four separate blocks of 16 trials (or 4 'games') during the experiment. Within each list, the order of prime-target pairs was randomised for each participant. At the beginning of every game, both players were given an empty 4 x 4 cardboard grid, and were instructed to fill this up with picture cards working left-to-right, top-to-bottom, starting in the top-left corner.

5.8.3 Procedure

Figure 14 shows an overhead view of the experimental set-up. The confederate and participant were seated on opposite sides of a table, with the two boxes of picture cards between them. Instruction cards were placed to the side of the players in card file boxes. Both players read instructions explaining the aim of the communication game: to fill up their empty grids with picture cards as quickly and as accurately as possible. They alternated between two tasks: describing instruction cards to their partner, and selecting picture cards to put on their own grid. The players were invited to look inside the boxes at the start of the session, to find out what was inside. The experimenter explained that it was very important when giving instructions to provide enough information for the other player to find the right card from the correct box. In every game, the confederate always gave the first description. The referential communication task was recorded on audio tape and later transcribed verbatim. As in Experiment 4, participants filled out a questionnaire at the end of the referential communication task. Firstly they were asked to write down what they thought the experiment might be "about". Then they answered a series of more specific questions which aimed to gauge whether they were aware of their partner using word order in a helpful (or unhelpful) way, and whether they had used word order strategically in order to be helpful themselves. The experimental session took about 40 minutes in total.

FIGURE 14

Experiment 5: Overhead view of the experimental set-up



5.8.4 Coding

Participants' target NPs were coded as pattern-first if pattern was mentioned before colour (e.g., *dotty orange square*), or colour-first if colour was mentioned before pattern (e.g., *orange dotty square*). Noun phrases that did not mention all three properties of the card within the same (grammatical) noun phrase (see Experiment 2), or which were not preceded by the right prime from the confederate, were coded as *others*. *Others* accounted for 3% of the data (71 responses out of 2048), which were excluded from the analyses reported below.

5.9 Results

5.9.1 Effects of priming and optimal design

Table 5 reports the proportion of colour-first NPs produced in each condition. Reporting colour-first NPs is arbitrary, since the proportion of pattern-first and colour-first NPs are complementary when *other* responses are excluded. Analyses of variance were conducted on these data, with Prime (Pattern-first vs. Colour-first) and Array (Pattern Box vs. Colour Box) as within-participants and within-items factors. Confederate (Helpful vs. Unhelpful) was entered as a between-participants factor in the *F1* analysis, and as a within-items factor in the *F2* analysis.

TABLE 5

Experiment 5: Proportion of colour-first NPs produced in each condition

(*Standard deviations* in brackets)

	Participant describes target card from:	
	Pattern Box (Box 1)	Colour Box (Box 2)
Helpful confederate produces:		
pattern-first prime (*wavy purple star*)	.51 (.30)	.52 (.28)
colour-first prime (*purple wavy star*)	.72 (.29)	.78 (.26)
Unhelpful confederate produces:		
pattern-first prime (*wavy purple star*)	.63 (.32)	.56 (.30)
colour-first prime (*purple wavy star*)	.44 (.29)	.42 (.28)

Unlike Experiment 4, there were no effects of either Prime or Array overall. Averaging over the Helpful and Unhelpful conditions, participants produced only about 3% more colour-first NPs following a colour-first prime from the confederate, a difference which was only significant in the by-items analysis ($F1(1, 30) = .39$, $MSe = .121$, $p = .54$; $F2(1, 15) = 16.09$, $MSe = .011$, $p = .001$). Participants were no more likely to produce a colour-first NP when they were describing a card from the Colour Box than when they were describing from the Pattern Box ($F1(1, 30) = .131$, $MSe = .008$, $p = .72$; $F2(1, 15) = .228$, $MSe = .021$, $p = .64$). There was an interaction between Prime and Array by participants, but not by items ($F1(1, 30) = 4.15$, $MSe = .006$, $p = .051$; $F2(1, 15) = .729$, $MSe = .694$, $p = .41$).

So, overall, there was no priming effect, and speakers did not seem to tailor the grammatical form of their noun phrases to fit the box from which they were describing a card. However, Table 5 suggests that participants who played the game with a helpful confederate behaved quite differently from participants who interacted with an unhelpful confederate. There were significant interactions between Prime and Confederate ($F1(1, 30) = 11.07$, $MSe = .121$, $p = .002$; $F2(1, 15) = 107.71$, $MSe = .012$, $p < .001$), and between Array and Confederate ($F1(1, 30) = 7.33$, $MSe = .008$, $p = .011$; $F2(1, 15) = 5.55$, $MSe = .001$, $p = .033$), and planned comparisons revealed some interesting differences between the Helpful and Unhelpful conditions.

In the Helpful condition, participants produced about 24% more colour-first descriptions following a colour-first prime from the confederate; i.e., they showed a significant priming effect ($F1(1, 30) = 7.81$, $p = .009$; $F2(1, 15) = 86.91$, $p < .001$). But in the Unhelpful condition, participants produced about 17% *fewer* colour-first descriptions after a colour-first prime than after a pattern-first prime. This "anti-priming" effect was significant by items and marginal by participants ($F1(1, 30) = 3.65$, $p = .066$; $F2(1, 15) = 31.19$, $p < .001$). In other words, participants showed the usual priming effect when they talked to a helpful confederate, but were significantly primed *in the opposite direction* when they talked to an unhelpful confederate.

Participants who talked to a helpful confederate produced about 4% more colour-first descriptions when they described cards from the Colour Box than when they described cards from the Pattern Box. This difference is in the direction predicted by optimal design, but is statistically non-significant ($F1(1, 30) = 2.75$, $p = .108$; $F2(1, 15) = .568$, $p = .463$). In the Unhelpful condition, however, participants produced about 5% *fewer* colour-first descriptions when they were describing cards from the Colour Box (compared to the Pattern Box), a significant difference in the direction opposite to that predicted by optimal design ($F1(1, 30) = 4.71$, $p = .038$; $F2(1, 15) = 7.33$, $p = .016$). So, with an unhelpful partner, speakers were more likely to produce *unhelpful* word orders themselves, but there was no effect either way with a Helpful partner.

5.9.2 Questionnaire data

As in Experiment 4, participants filled out a questionnaire at the end of the referential communication task. The aim was to find out how aware they were of the experimental manipulation(s), and of the confederate's language behaviour (Helpful or Unhelpful). The full questionnaire can be found in Appendix E, and Table 6 gives some examples of participants' responses.

Overall, 14 out of the 32 participants (44%) explicitly stated that the experiment might primarily be about word order, or the order in which pattern and colour information was conveyed. Eight of these participants were in the Helpful condition, and six were in the Unhelpful condition. 19 participants (59%) wrote that they tried to use word order in a helpful way, to make the game easier for their partner. Of these 19, 8 were in the Helpful condition (42%) and 11 were in the Unhelpful condition (58%). A further 10 participants (3 in the Helpful condition, and 7 in the Unhelpful condition) explicitly wrote that they *could* have made the game easier for their partner by using different word orders to describe cards from the Pattern and Colour boxes.

Nine participants (28%) mentioned that their partner (the confederate) made the game easier by producing word orders that were helpful. One of these 9 participants was in the condition where the confederate had been scripted to produce *unhelpful* word orders throughout the experiment (this person claimed to have used word order helpfully themselves, and said that they thought their partner had done the same). A further 10 participants (2 in the Helpful condition, and 8 in the Unhelpful condition) said that the confederate *could* have made the game easier by tailoring word order to fit the box from which they were describing a card. So, from the questionnaire data, it looks as though participants in the Unhelpful condition were just as aware (as participants in the Helpful condition) of what *could* make a description more helpful, but they did not necessarily go to the effort of producing helpful word orders themselves. Some of the issues raised by the questionnaire data are taken up in the General Discussion.

TABLE 6

Experiment 5: Sample responses to the post-experiment questionnaire

What was the experiment about?
Mentions word order:
Maybe the order in which one describes things – I would normally say e.g. a dotted blue star like a clear blue sky, but in box two it is easier for the partner to find shapes if the colour is said first
Doesn't mention word order:
To see who can describe shapes and different colours and patterns better, males or females
Did you do anything special to make it easy for your partner to build up the grids?
Mentions word order:
After the first grid I began giving descriptions for box 1 starting with the pattern and for box 2 starting with the colour
Doesn't mention word order:
In the end I found myself describing cards in the same way as my partner, e.g. my code for "diagonal stripes" changed to "stripy"
Do you think you could have been more 'helpful' when you described your cards?
Mentions word order:
If it was box 1 I could have said pattern, colour then shape when describing and if it was box 2 I could have said colour, pattern and then shape
Doesn't mention word order:
No, apart from being faster, or repeating myself
Did your partner do anything special that made it easy for you to build up the grids?
Mentions word order:
yes said the pattern first for box 1 and colours for box 2
Doesn't mention word order:
Not that I noticed

5.10 Discussion

Taken overall, Experiment 5 failed to find effects of priming and optimal design on speaker's choice of prenominal adjective order within noun phrases. But when the results of the Helpful and Unhelpful conditions are analysed separately, clear effects of both processes emerge. The Helpful Confederate condition extends the noun phrase priming effect found in Experiment 4 to prenominal adjective order. Speakers were more likely to mention colour before pattern (e.g., *orange dotty square*) when they had just heard a prime description with the same word order from the confederate (*purple wavy star*). There was no effect of optimal design in the Helpful condition (although there was a numerical trend in the predicted direction), but speakers in the *Unhelpful* condition behaved significantly unhelpfully themselves with respect to word order. The effects of Helpful and Unhelpful interlocutors are taken up in the General Discussion below.

The optimal design effect in Experiment 5 was much smaller than in Experiment 2, which used exactly the same picture cards as stimuli and found that speakers adapted word order to their addressee's perspective (see Chapter 4). Why were speakers in that experiment more likely to produce helpful word orders? The dynamics of the task may have something to do with it. In Experiment 2, participants described eight cards in a row, and before that they matched cards to eight of their partner's descriptions. This gave them time to find out what would be helpful in terms of word order, and plenty of opportunity to practise using word order in a "helpful" way (it also probably contributed to a priming mechanism that inflated the production of helpful word orders; see Chapter 4). In Experiment 5, on the other hand, players took turns to be the Describer and the Matcher one card at a time. To show evidence of optimal design, the speaker has to constantly re-assess what would be helpful for the addressee, because two cards in a row may require different word orders. This is a more onerous task for the speaker, and may explain why participants in Experiment 5 were less likely to show evidence of attending to the addressee's needs (cf. Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Bard & Aylett, in press; Keysar, Barr, Balin, & Paek, 1998).

5.11 General Discussion

5.11.1 Speakers tailor word order to the addressee's needs?

Experiments 4 and 5 support the idea that speakers tailor fairly low-level aspects of their speech (specifically, word order in referring expressions) to their addressee's needs or perspective, at least under some circumstances. As in Experiments 2 and 3, participants tended to produce noun phrases that addressees could incrementally map onto an array, even when this meant choosing word orders that are normally dispreferred, relative to the alternative. In Experiment 4, speakers mentioned shape before colour if the addressee's array was organised by shape, even though "colour before shape" is the preferred order in English (Kemmerer, 2000). Likewise, speakers were somewhat more likely to mention pattern before colour in Experiment 5 when this would be easier to map onto an array, despite a small "colour before pattern" bias in English (see Experiment 2).

In Experiment 5, the evidence for optimal design was relatively weak. Speakers only produced helpful word orders when their partner (the confederate) was doing the same, and even then, the effect was small (just 4% more colour-first NPs when speakers were describing cards from the Colour Box, compared to the Pattern Box). When the confederate behaved unhelpfully, participants tended to produce unhelpful word orders too. The fact that optimal design interacted with the confederate's behaviour is interesting, for at least two reasons. Firstly, it suggests that optimal design may be under strategic control, in the sense that speakers only attend to it when there seems to be good reason to do so (cf. Hanna, Tanenhaus, & Trueswell, 2003; see 5.11.2). Secondly, it may give us some indirect evidence that speakers are actually adapting word order "for" the addressee.

Throughout the thesis, I have assumed that participants in the experiments reported here were making linguistic choices based on their addressee's perspective. Of course, there are alternative explanations that have nothing to do with the addressee (see 4.12.2 in Chapter 4), and up until now there has been little evidence to unequivocally support a notion of optimal design "for" the addressee. For example, speakers may mention some property of a picture card first because it is perceptually (or conceptually) salient to themselves. But the results of Experiment 5 suggest that there is something more interesting going on. If word order simply reflects what is seen by, known to, or easiest for the speaker, then the confederate's behaviour should have no effect. But the confederate's behaviour *did* have an effect in Experiment 5: participants who talked to a helpful confederate were more likely to show evidence of helpful language behaviour themselves than participants who played the game with an unhelpful confederate. In fact, participants in the Unhelpful condition showed a small but significant tendency to produce unhelpful word orders themselves. Possible explanations for this "anti-helpfulness" effect are discussed in 5.11.2 below. The important point here is that the word order effects seen in Experiments 2 and 3 were unlikely to be solely due to what the speaker was looking at or focusing on at the time of speaking.

5.11.2 Effects of an unhelpful interlocutor

In Experiment 5, participants who played the communication game with an interlocutor who was consistently unhelpful (in terms of their choice of word order) showed "anti-priming" and "anti-helpfulness" effects in their own descriptions. They produced noun phrases that had the opposite word order from the confederate's prime, and they tended to produce descriptions that could not be straightforwardly mapped onto their partner's array. The anti-priming effect in particular is highly unusual; priming experiments generally show positive effects (a significantly increased tendency to repeat the primed structure, relative to the alternative) or null effects, but not usually *negative* effects (a significantly increased tendency to produce the alternative structure, rather than the primed structure).

One way to account for the anti-priming effect is to consider what might happen when participants hear an unhelpful prime description from the confederate. Their partner says something like *Box 2, it's the wavy purple star.* "Box 2" is the Colour Box, so to find the card, the participant has to mentally recode the description they have just heard, or at least ignore the first adjective (*wavy*) and focus first on the second adjective while they locate the purple cards in the Colour Box. Either way, some kind of mental "work" is required in order to carry out the matching task. So, by the time the participant comes to give his own target description, his mental representation of the prime has been mentally manipulated. In fact, the participant's representation of the prime may have been completely recoded to make it incrementally "appropriate", e.g. *purple wavy star.* If this is the case, then it is hardly surprising that we see an "anti-priming" effect for participants in the Unhelpful condition. In effect, they have primed themselves with the opposite word order from the one they heard in the confederate's prime. This is the same 'recoding' explanation that was discussed in Chapter 4 with respect to the contribution of priming to the word order effects reported for

Experiments 2 and 3 (see section 4.12.2, p. 109).

An alternative explanation for the anti-priming effect could be that participants were strategically "deciding" not to align (not necessarily consciously) with the unhelpful confederate (Pickering & Garrod, in press) perhaps because her uncooperative behaviour lead to them liking her less, or feeling less motivated to fall in with her behaviour (Schober & Brennan, 2003). As noted in 5.7, there is certainly some evidence that speakers are less likely to linguistically align with an interlocutor (e.g., in terms of word choice and accent) with whom they are not getting along (Bly, 1993; Danet, 1980; Giles, Coupland, & Coupland, 1991; see Schober & Brennan, 2003). On the basis of Experiment 5, it is difficult to choose between this explanation and the "recoding" explanation for the anti-priming effect, but we return to this issue in Chapter 6.

What about the "anti-helpfulness" effect? Well, the unhelpful confederate was 100% consistent in her use of pattern-first or colour-first word order. When she described cards from the Pattern Box she produced descriptions like *purple wavy star*, and when she described cards from the Colour box she produced descriptions like *wavy purple star*. One way to account for the anti-helpfulness effect is to say that participants had picked up on her (consistent) use of word order, and tried to accommodate to her by using the same word order that she was using. For example, they noticed that the confederate always used a pattern-first word order when she described cards from the Colour Box, and tried to be helpful by doing the same thing back. The questionnaire data suggest that this is not the case; no participant explicitly said that they used "unhelpful" word orders for their partner's benefit. Of course, what the participants wrote on their questionnaires, and what they actually *did* and thought during the game, are not necessarily the same thing. But on the assumption that participants were being honest (and there is no reason to think that they were not; after all, accommodating to a partner is surely a desirable thing to be seen doing), then adapting to the confederate's use of different word orders does not seem to provide an explanation.

What alternative explanations could there be? One way of looking at the anti-helpfulness effect is to think of it as strategic "tit-for-tat" in language behaviour, a sort of "if you don't scratch my back (by producing helpful descriptions), then I won't scratch yours". Tit-for-tat effects are certainly seen in other (non-linguistic) domains where participants can either choose to behave co-operatively or selfishly. For example, in the decision-making game known as the Prisoner's Dilemma (e.g., Pruitt & Kimmel, 1977), people tend to choose co-operative strategies only if they have been treated co-operatively by other players in the past, or if they hope to receive co-operative treatment in the future (e.g., Axelrod, 1984). There is even some evidence that there are individual differences in terms of how likely people are to reciprocate co-operative or selfish strategies that they encounter in interactions with other players (Perugini, Galluchi, Presaghi, & Ercolani, in press).

Alternatively, participants in the Unhelpful condition might have been primed by the confederate to behave unhelpfully themselves, in an entirely unconscious way. Social psychological experiments suggest that people can be non-consciously primed to behaving according to some stereotype (e.g., Chen & Bargh, 1997; Epley & Gilovich, 1999). For example, participants primed with words relating to rudeness in a scrambled sentences task are more likely to subsequently interrupt a conversation, compared to participants in a control condition (Bargh, Chen, & Burrows, 1996). A similar kind of effect might be operating in Experiment 5; the confederate provides a model of "unhelpfulness" or "unco-operativeness", which primes the participant to be unhelpful themselves.

On the basis of the Experiment 5, it seems difficult to choose between the (non-conscious) priming explanation and the (conscious) tit-for-tat explanation for the anti-helpfulness effect observed here. The post-experiment questionnaire didn't probe whether participants did anything to make life more *difficult* for their partner, only whether they did anything helpful. However, an informal look at the relationship between participants' responses to the questionnaire and their behaviour during the game is suggestive. Intriguingly, the three participants who behaved *least* helpfully with respect to word order (i.e., those who showed the biggest "anti-helpfulness" effects, in numerical terms) were all in the Unhelpful condition. All three explicitly stated that they knew they *could* have used different word orders to describe cards from the two boxes, but said that they had not done anything special to make the game easier for the Matcher. Moreover, they all said that their partner had not done anything to make the game easier for them, but were able to write down how she *could* have used word order in a helpful way. This is interesting, because it suggests that these participants may well have deliberately decided *not* to be helpful, based on the confederate's behaviour towards them. In other words, they might have engaged in a conscious tit-for-tat strategy.

5.11.3 Individual differences in priming and optimal design

It is interesting to note that the by-participants (*F1*) and by-items (*F2*) statistics suggest that there is more varibility between participants than between items in some analyses. For example, in Experiment 4, the small effect of Array was highly robust in the *F2* analysis ($p < .001$) but marginal by participants ($p = .053$). This suggests that optimal design effects vary between speakers; some show a good deal of evidence of adapting to the addressee's perspective, while others never seem to change their utterances based on what the addressee is doing. This is exactly what we might expect, given the variety of factors that can affect the likelihood of a speaker attending to optimal design (Schober & Brennan, 2003). They must be able to work out what would be easiest to understand or process given the addressee's perspective, have sufficient cognitive resources to work this out, and also be motivated to behave in a linguistically helpful way.

Priming effects in these experiments also tend to be more homogenous in the item sample than the participant sample. For example, the overall effect of Prime was more robust by-items in both Experiments 4 and 5, and separate planned comparisons for both the Helpful and Unhelpful conditions in Experiment 5 also showed that there was more varibility between participants than between items. This is an interesting observation, because it suggests that priming is not a pervasive, wholly automatic influence on language processing. Instead, it seems that priming may be under some degree of strategic control. Alternatively, perhaps there are several sub-processes involved in alignment, one or more of which is non-automatic.

The fact that there are individual (or pair-wise) differences is particularly important for Experiment 5, which used a between-participants manipulation to study the effects of a Helpful or Unhelpful confederate on priming and optimal design. The interesting interactions observed between Confederate and both Prime and Array could potentially reflect "baseline" differences between speakers, if participants who were fundamentally more addressee-oriented or prone to being primed ended up in the Helpful condition. Given that participants were randomly assigned to the Helpful and Unhelpful conditions, it seems unlikely that between-speaker differences could entirely account for the pattern of results, but it is worth bearing this caveat in mind in interpreting the data.

5.11.4 The locus of noun phrase priming effects

Experiment 4 showed that speakers can be primed to produce dispreferred noun phrase structures, like *the triangle that's green*. The effect is small (about 5%), perhaps because there is no lexical overlap between prime and target, but nonetheless it is robust. This replicates Cleland and Pickering's (2003) findings, using a similar kind of dialogue task. The Helpful confederate condition in Experiment 5 showed that the linear order of adjectives appearing before a head noun can also be primed. Speakers were more likely to mention colour before pattern (*orange dotty square*) when they had just heard a noun phrase with the same adjective order.

Cleland and Pickering (2003) suggested that noun phrase priming occurs at the lemma level, via the activation of combinatorial nodes which encode different phrase structures (cf. Pickering & Branigan, 1998). Some theoretical accounts claim that combinatorial information is shared between comprehension and production (e.g., Branigan et al., 2000), such that hearing a post-nominally modified NP like *the square that's red* activates the combinatorial node encoding the phrase structure "noun + relative clause" and makes it more likely that the speaker will use the same structure themselves in a subsequent utterance. However, if priming were driven only by activation of a combinatorial node, then the relative order of adjectives within a canonical, prenominally-modified noun phrase (e.g., *orange dotty square*) ought not to be prime-able; *orange dotty square* and *dotty orange square* are syntactically identical, and would therefore be represented by the same combinatorial node. So, Experiment 5 suggests that priming effects (at least within noun phrases) do not necessarily arise at the lemma level, or can also arise elsewhere.

5.12 Chapter Summary

Experiment 4 replicated Cleland and Pickering's (2003) noun phrase priming effect. Speakers tended to repeat the syntactic structure or word order of a noun phrase that they had just heard their interlocutor use, even though there was no lexical overlap between the two expressions. Since "pattern-first" (*dotty orange square*) and "colour-first" (*orange dotty square*) noun phrases have exactly the same syntactic structure, Experiment 5 suggests that this phenomenon might best be described as word order priming (cf. Hartsuiker, Kolk, & Huiskamp, 1999; Hartsuiker & Westenberg, 2000), rather than syntactic priming. Experiments 4 and 5 also provide some evidence that speakers choose noun phrase structures (or word order within noun phrases) that are "incrementally helpful" for an addressee, although effects of optimal design were weaker here than in Experiments 2 and 3. Participants tended to mention shape before pattern (*the triangle that's green*) when their partner would have to locate the subset of triangles in an array before finding the *green* triangle. Likewise, they tended to mention pattern before colour (*dotty orange square*) when their partner was choosing picture cards from an array organised by pattern at the highest level. Interestingly, this tendency to use addressee-optimal word order was influenced by the interlocutor's own language behaviour.

CHAPTER 6

Optional Word Mention and Ambiguity Avoidance

6.0 Chapter Overview

This chapter investigates whether the language production system is sensitive to, and avoidant of, ambiguity. We know from comprehension studies that some kinds of (syntactic) ambiguity can disrupt reading (e.g., Boland & Bohem-Jernigan, 1998; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). If speakers design utterances with addressees in mind, we might expect them to avoid producing syntactically ambiguous utterances, because these could be difficult to understand. Speakers have (at least) three devices that they can exploit to avoid syntactic ambiguity in production. Firstly, they can use prosodic cues to signal intended meaning. Secondly, they can exploit syntactic flexibility, and use constituent order to explicitly disambiguate the syntactic structure of an utterance. Thirdly, they can use optional grammatical words to bias comprehension processes towards the correct interpretation of an utterance. A number of studies have investigated whether speakers use these devices, and the major findings are reviewed below. In 6.3 we turn to the question of whether the use of optional function words can be primed. 6.4 outlines the motivation for Experiment 6, which examines speakers' use of the optional complementizer *that* to avoid PP-attachment ambiguities.

6.1 Ambiguity Avoidance and Optimal Design

Ambiguity is rife in language. For example, a spoken word like *candle* is initially compatible with cohort competitors (words that share similar phonological onsets) like *candy* (e.g., Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Individual words like *bat* may be lexically ambiguous, noun phrases can be referentially ambiguous (as in, *Bill saw Tom and then <u>he</u> hit <u>him</u> in the mouth*, Ferreira & Dell, 2000, p.330), and syntactic structures often have more than one possible interpretation, as we will see below.

Some sentence-level ambiguities are global, in the sense that they are never linguistically resolved. For example, a sentence like *Angela shot the man with the gun* has (at least) two meanings:

either Angela used a gun to shoot someone, or she shot a man who happened to be carrying a gun (Keysar & Henly, 2002). There is nothing in the sentence itself to help resolve the ambiguity, although a supporting context (either visual or linguistic) might make it clear which meaning was intended. Other ambiguities are only temporary, in the sense that they are resolved by subsequent linguistic material. The experiment reported in this chapter exploits one kind of temporary ambiguity in particular: the prepositional phrase (PP) attachment ambiguity (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

In a sentence like *Put the penguin (that's) in the cup on the saucer*, the function word *that* is optional. A speaker or writer can choose whether or not to include it, because the sentence is equally grammatical whether a *that* is used or not (Bever, 1970). Importantly, the sentence basically means the same thing either way. In other words, the *that* does not carry any propositional meaning in and of itself (although including a *that* might change the semantics of the sentence in a subtle way; e.g., Yaguchi, 2001). Imagine that a speaker, Angela, utters the instruction *Put the penguin (that's) in the cup on the saucer* in the following context: she and her addressee, Nick, are sitting in front of a table with several objects on it. Included in the set of objects are two toy penguins, one inside a small cup, and one standing alone. There is a second empty cup on the table, as well as a saucer. Angela wants to get Nick to take the penguin out of the cup and put it onto the saucer. If she doesn't include a *that* in her instruction, the PP *in the cup* is structurally ambiguous, at least temporarily. It could either be interpreted as a destination or goal (the place where a penguin should be put), or as a modifier of the noun phrase *the penguin* (i.e., it's the penguin in the cup that should be put on the saucer, rather than the other one). When Nick hears the temporarily ambiguous PP *in the cup*, he might initially think that Angela wants him to move one of the penguins into the empty cup. However, if Angela includes a *that*, Nick is likely to come to a modifier interpretation for *in the cup*, which turns out to be the syntactic analysis that Angela originally intended. Note that a goal interpretation for *in the cup* is not inherently 'wrong'. It is simply not what Angela intended, and Nick could probably work that out, given the context.

It is worth noting that temporary structural ambiguities like this one do not necessarily cause comprehension difficulties in and of themselves (Ferreira & Dell, 2000). Problems only arise if the comprehension system parses the ambiguous phrase using a syntactic analysis that turns out to be wrong, given the rest of the sentence (Tanenhaus et al., 1995; Trueswell, Tanenhaus, & Kello, 1993). Difficulties occur when the parser initially commits to the goal interpretation for an ambiguous PP but later discovers (at the second PP, *on the saucer*, in the "penguin" example) that this is incorrect (Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). This kind of syntactic misanalysis is known as a "garden path" effect (e.g., Fodor & Frazier, 1978). Garden paths are most often studied in reading and can be seen:

> ... *when sentences (a) contain temporary syntactic ambiguities and (b) are biased at the point of temporary ambiguity toward a syntactic analysis that is eventually inappropriate.* (Ferreira & Dell, 2000, p.300)

Native speakers of English consistently show a strong bias towards a goal (or "high attachment") interpretation when they read ambiguous PPs like *in the cup* in neutral contexts (e.g., Britt, 1994; Ferreira & Clifton, 1986; Rayner et al., 1983). But as we have seen, the goal interpretation can be avoided if an optional function word, *that*, is included. What does all this have to do with optimal design? If utterances are designed to be optimal for comprehension, we might expect the production system to follow a strategy of "ambiguity avoidance". Speakers could exploit optional *that*-mention (amongst other devices) to avoid producing ambiguous utterances. This would be desirable because it ought to help addressees avoid syntactic misanalyses that could cause comprehension difficulties (Ferreira & Dell, 2000). However, if utterance planning is set up to maximise fluency or ease of processing for the *speaker*, then perhaps *that*-mention wouldn't be used for ambiguity avoidance. If producing extra words means extra effort in encoding an utterance, then production might be easier and more efficient without an optional *that*.

Optional *that*-mention is not the only strategy available to speakers for disambiguating potentially ambiguous utterances. In 6.2.1 and 6.2.2 I introduce two other means available to the production system for ambiguity avoidance: prosody, and constituent order. As we will see, experimental evidence suggests that speakers do not consistently use such strategies, and this observation has been used to support speaker-centred or accessibility-based theories of production.

6.2 Strategies for Ambiguity-Avoidance

6.2.1 Prosody

Prosody can reflect syntactic structure (e.g., Lehiste, 1972), and listeners can use prosodic information to guide their interpretation of an utterance (e.g., Beach, 1991; Carlson, Clifton, &

Frazier, 2001; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Pynte & Prieur, 1996; Warren, Grabe, & Nolan, 1995). For example, a well-placed prosodic boundary can be used to signal either the modifier (low attached) or goal (high attached) interpretation of an ambiguous PP. However, the jury is still out as to whether untrained speakers reliably and spontaneously produce consistent prosodic cues to syntactic structure, and crucially whether they do this "for" the addressee.

Allbritton, McKoon, and Ratcliff (1996) asked trained and untrained speakers (radio announcers and university students) to read aloud paragraphs containing globally ambiguous sentences like *They rose early in May*. These critical sentences were disambiguated by the discourse context, but speakers were not cued to the potential ambiguity. Neither the students nor the radio announcers reliably produced prosodic cues to disambiguate the meaning of the ambiguous sentences under these conditions. Similarly, Keysar and Henly (2002) showed that speakers are not very good at using prosody to disambiguate sentences like *Angela shot the man with the gun* for addressees, even when they think they are producing sufficient prosodic cues to the intended meaning.

Other evidence suggests that speakers *do* use prosody successfully for disambiguation (e.g., Schafer, Speer, Warren, & White, 2000). For example, Snedeker and Trueswell (2003) had speakers memorise written instructions like *Tap the frog with the flower* and then say them aloud so that a listener (who was on the other side of a screen) could correctly carry out an action. Listeners had a set of objects in front of them which supported both a modifier and an instrument interpretation for the ambiguous PP *with the flower*. For example, the listener's array contained a frog holding a small flower, a second frog not holding anything, a large flower, and a couple of distractor objects. The experimenter demonstrated an action to the speaker, which the listener was supposed to reproduce. In 'instrument' demonstrations, the experimenter used the large flower to tap the second frog. In 'modifier' demonstrations, the experimenter tapped the frog holding the flower with her finger.

When speakers had the same set of objects in front of them as the listener, they were very aware of the potential PP ambiguity (over 90% of speakers mentioned it in a post-experiment questionnaire), and they used prosodic cues to signal which meaning of the instruction was intended. However, if the speaker's array did not support both interpretations of the ambiguous PP, then they were less likely to prosodically disambiguate the intended meaning of the instruction. For example, if the speaker's array didn't include the large flower (which rules out the instrument interpretation for *with the flower*), they were less likely to give strong cues to the modifier interpretation, and vice versa. This suggests that speaker awareness is crucial for using prosody to avoid global syntactic ambiguity in spoken production. When speakers appreciate that a potential for misinterpretation exists, they exaggerate prosody to help convey intended meaning to the listener. When they are unaware of a potential problem, they put less prosodic 'effort' into helping the listener reach the correct interpretation. Snedeker and Trueswell's findings have important implications for optimal design. To have even a hope of exploiting prosody "for" addressees, is seems, the production system needs to be aware of potential ambiguities. Moreover, speakers need to be aware (not necessarily explicitly, but at some level) that ambiguity can be difficult for addressees to deal with.

6.2.2 Constituent order

Constituent order within a sentence is another device that speakers could (in principle) use to avoid temporary ambiguities. In a (spoken) sentence like *The judge sent the letter to the president to the members of the congressional subcommittee*, the first prepositional phrase, *to the president*, is structurally ambiguous. It could either be interpreted as a goal (the president was the person to whom a letter was sent), or as a modifier (i.e., the thing being sent is *the letter to the president*). The latter interpretation (the modifier reading) turns out to be correct; at the second PP, it becomes clear that the actual recipient of the letter is not *the president*, but *the members of the congressional subcommittee*. This ambiguity could be avoided by ordering the constituents of the sentence in a different way. For example, a double object construction (*The judge sent the members of the congressional subcommittee the letter to the president*) or shifted-PP structure (*The judge sent to the members of the congressional subcommittee the letter to the president*) would remove the ambiguity. These constructions are found less commonly in English speech corpora than prepositional objects (Pickering, Branigan, & McLean, 2002), but both avoid the PP-attachment ambiguity. If the language production process is optimally ambiguity-avoidant, we might expect speakers to use these kinds of unusual constructions more often when the equivalent prepositional object construction would otherwise contain an ambiguity than when it would not. In Experiments 2 (Chapter 4) and 5 (Chapter 5), we saw some evidence that speakers can and do use word order *within noun phrases* in an addressee-optimal way. By extension, we might expect speakers to use *constituent order* within a sentence in a similarly "helpful" way. In fact, the experimental evidence suggests that this is not the case.

Arnold, Wasow, Asudeh, and Alrenga (2004) had speakers read and remember sentences like *The judge sent the letter to the president to the members of the congressional subcommittee.* Once they had memorised the sentence, their partner – the addressee – read aloud a question (e.g., *What did the judge do?*). The speaker's task was to answer this question so that the addressee could choose one of two alternative responses on an answer sheet. The results suggest that speakers do not use constituent ordering to avoid ambiguities. Participants were no more likely to use double object and shifted-PP constructions when the stimulus sentence they had read was temporarily ambiguous than when it was inherently unambiguous (e.g., *The judge sent the letter for the president to the members of the congressional subcommittee*; cf. Ferreira & Dell, 2000).

To summarise, the experimental evidence demonstrates that speakers do not exploit constituent order to avoid ambiguities, and that they do not systematically use prosodic cues for disambiguation either. Prosody and constituent order will not be discussed further here as they are not investigated in Experiment 6. However, the studies discussed so far provide a back-drop to the issue of interest here: speakers' use of optional function words.

6.2.3 Optional function words

Beyond prosody and constituent order, speakers have another way to disambiguate syntactic structure. As described at the start of this chapter, they can use optional function words, such as the complementizer *that*, to help addressees avoid garden paths. The use of optional function words is an intriguing issue, because speakers are unlikely to be explicitly aware of their linguistic behaviour in this respect. This gives us an interesting comparison with Experiments 2, 3, 4 and 5, where speakers did seem to be aware of how they could flexibly use word order and syntax "for" an addressee. Do speakers implicitly use optional function words to avoid ambiguities? Experimental evidence to date suggests not.

Ferreira and Dell (2000) studied speakers' use of optional function words across six different experiments and in two kinds of construction: sentence complements (e.g., *The coach knew (that) you missed practice*), and passive relative clauses (e.g., *The astronauts (who were) selected for the Apollo missions made history*). The rationale behind all of Ferreira and Dell's experiments is this: if production is fundamentally ambiguity-sensitive, speakers ought to include optional function words more often in sentences that would be ambiguous without them, compared with sentences that are inherently unambiguous (cf. Arnold et al., 2004). A recall-based sentence production paradigm was used in this set of experiments. Speakers read and memorised critical sentences, and later were given cue words as a prompt to recall each one aloud.

Experiments 1 and 2 investigated optional *that*-mention in sentence complement constructions like *The coach knew (that) you missed practice*. Without a *that*, the pronoun *you* is ambiguous between a direct object interpretation (as in, *The coach knew you very well*) and an embedded subject interpretation (which turns out to be the correct analysis for the sentence). Ferreira and Dell compared *that*-mention in potentially ambiguous and inherently unambiguous sentences. In *The coach knew (that) I missed practice*, the pronoun *I* is unambiguously case-marked as a grammatical subject, so the direct object reading is blocked. If production is sensitive to ambiguities, we would expect speakers to include *that*s more often in the ambiguous "*you*" sentences, compared to the unambiguous "*I*" sentences. This was not the case. Although speakers included *that*s a little more often in the ambiguous sentences (77% of trials) compared to the unambiguous *I* sentences (74%), the difference was small and not statistically significant (in the by-subjects analysis, although it *was* significant in the by-items analysis). The same (null) effect was found when *she* and *he* were used as unambiguous pronouns, instead of *I* (Experiment 2). Again, speakers included *that*s a little more often in ambiguous sentences (70% of trials) than in unambiguous sentences (68%), but the difference was not significant. When the results of Experiments 1 and 2 were combined, the ambiguity effect was marginally significant by speakers ($p < .06$) but not by items ($p < .13$). This might suggest that there is a weak tendency towards ambiguity-avoidance after all. Experiment 3 was designed to explore this possibility.

Ferreira and Dell's (2000) Experiment 3 investigated speakers' use of optional function words in passive relative clauses (*The astronauts (who were) selected for the Apollo missions made history*). The ambiguity avoidance hypothesis predicts that speakers should include optional words more often when the main verb of the sentence could cause a garden path (e.g., *selected*) than when it is unambiguously marked as a past participle (as in, *The astronauts chosen for the Apollo missions made history*). The "reduced relative" ambiguity is known to cause strong garden path effects (Gorrell, 1995; Pritchett, 1992). If there *is* a weak ambiguity-avoidance strategy in production, then it ought to show up better using this manipulation. In fact, the results were just the same as in Experiments 1 and 2. If anything, there was a slight trend in the "wrong" direction. Speakers were slightly *less* likely to include *who were* when the main verb was ambiguous (*selected*) than when it was unambiguous (*chosen*).

So, "Across three experiments, investigating two kinds of ambiguity, speakers did not consistently avoid producing sentences with temporary structural ambiguities." (Ferreira & Dell, 2000, p. 313). This was true even when other linguistic variables (such as how often a particular verb appears in direct object and sentence complement constructions, or the plausibility of a direct object interpretation for the ambiguous fragment) were factored out. That is, *that*-mention did not seem to be used to avoid ambiguities even in those sentences that were most likely to cause garden-path effects. Instead, the variable that best predicted *that*-mention was how often a specific verb normally co-occurred with *that* in a sentence complement structure, as indexed by production norms (Garnsey, Perlmutter, Myers, & Lotocky, 1997).

Three further experiments showed that speakers' use of optional *that*-mention could be better explained in terms of the accessibility of material in the critical sentence. If the subject of the embedded clause was the same as the subject of the main verb (Ferreira and Dell's Experiments 4 and 6), speakers were more likely *not* to include a *that*. For example, speakers were more likely to omit a *that* in a sentence like *You knew you had booked a flight for tomorrow* than in *You knew I had booked a flight for tomorrow*. Ferreira and Dell explained this effect in terms of a 'principle of immediate mention'. If the embedded subject has already appeared as the subject of the main verb, an optional *that* is likely to be omitted, to allow the repeated pronoun to appear sooner rather than later.

Ferreira and Dell's first five experiments seem to show that production is not inherently ambiguity-avoidant. But perhaps this is because speakers in these studies did not have any real communicative goals or intentions. The cued-recall paradigm is essentially a monologue task; speakers' main intention is to successfully recall the original sentence, and there is no particular consequence of being ambiguous. Maybe *that*-mention is sensitive to ambiguity in tasks more akin to real-world language situations? When being understood is important, "Speakers may detect which of their produced sentences are ambiguous (or at least, "funny sounding") and spontaneously clarify such sentences." (Ferreira & Dell, 2000, p.322).

In a sixth and final experiment, participants carried out the same sentence recall task, but this time half of them spoke to an addressee who rated their utterances for "clarity". The results replicated those of the previous five experiments: accessibility affected *that*-mention, but ambiguity had no effect. There was however one interesting finding. Participants who spoke to addressees produced about 7% more *that*s overall than participants who did not. This difference was reliable in the by-items analysis, but not in the by-participants analysis. Ferreira and Dell suggest that this could be an effect of register (i.e., speakers talk more formally when the need for clarity is emphasised), but it might also show a coarse-grained sensitivity to ambiguity (cf. Brown & Dell, 1987). Interestingly, addressees' clarity ratings were high whether speakers included a *that* or not (5.30 and 5.43 respectively, on a 7-point scale), suggesting that listeners are often just as unaware of potential pitfalls for comprehension as speakers seem to be (see Snedeker & Trueswell, 2003 for discussion). Overall, Ferreira and Dell (2000, p. 326) concluded that:

> *Speakers are thus capable of making gross changes to their sentences in response to communicative demands, which are likely to be the kinds of changes that speakers make when addressing communicatively impaired individuals... However, speakers do not seem to adjust their production based on subtle properties of sentences that have measurable impact on comprehension, such as garden paths that are contingent on the case properties of an embedded subject pronoun.*

An unpublished study by Kraljic and Brennan (2003) came to a similar conclusion, using a different methodology (one which is arguably more akin to real-world language use), and studying a different use of the optional function word *that*. Rather than having speakers memorise and reproduce sentences, participants gave spoken instructions for a partner to move objects about in a vertical display (cf. Barr & Keysar, 2002; Metzing & Brennan, 2003). Speakers were given pictorial representations of the display, which were marked to indicate which object the speaker should tell their partner to move, and where it should go. So in this experiment, speakers were generating utterances that were spontaneous but controlled for content. Critical instructions potentially contained an ambiguously-attached prepositional phrase. For example, the PP *in the basket* in the instruction *Put the dog in the basket on the star* could initially either be interpreted as a potential goal (where to put the dog) or as a modifier of the NP *the dog* (i.e., giving additional information about which object should be moved – a dog that's in a basket). Half of the critical instructions were produced in so-called ambiguous contexts. For example, the array would include a star, a dog in a basket, and an empty basket on a star, supporting both of the possible interpretations (modifier or goal) for the PP *in the basket*. In unambiguous contexts, only one interpretation was intended to be viable, given the objects in the array. As in Ferreira and Dell's (2000) experiments, ambiguity had no effect on *that*-mention. Speakers were no more likely to include a *that* in ambiguous contexts than in unambiguous contexts (T. Kraljic, personal communication).

Like Snedeker and Trueswell (2003), Kraljic and Brennan conclude that their speakers simply didn't seem to be aware of the potential ambiguity "on-line". In a follow-up study (Kraljic & Brennan, 2003, Experiment 2), they showed participants photographs of the same arrays used in the experiment described above, along with written versions of the instructions. For each photograph, the participant was asked to decide whether the instruction was ambiguous, given the context. Even though the arrays contained a relatively small number of objects, it typically took participants around four seconds to make their decision. Given the speed at which spoken language proceeds, it is hardly surprising that speakers didn't manage to plan ambiguity-free instructions in Kraljic and Brennan's production experiment; utterances would be planned and executed well before the potential ambiguity could be detected.

To summarise, the evidence to date seems fairly convincing with regard to the use of optional function words as an ambiguity-avoidance strategy. Ferreira and Dell (2000) and Kraljic and Brennan's (2003) studies suggest that speakers don't use optional function words to avoid ambiguities that are known to cause comprehension difficulties (albeit mild difficulties; Sturt, Pickering, & Crocker, 1999) in reading.

6.3 Priming Optional Complementizers

In 6.2 we saw that speakers don't exploit the *that*-complementizer to avoid (two different kinds of) syntactic ambiguities. But recent experimental evidence shows that *that*-mention is not completely insensitive to context. More specifically, *that*-mention can be primed. In a series of 3 experiments, V. Ferreira (2003) showed that speakers are more likely to include a sentence-complementizer-*that* if they have just produced a *that* in their previous utterance. Ferreira's experiments employed a recall-based sentence production task similar to the one used by Ferreira and Dell (2000). Speakers memorised pairs of sentences, and later reproduced them as faithfully as they could. The degree of *that*-priming was measured by comparing the proportion of *that*s produced in target sentences following different kinds of primes.

Three experiments showed robust priming effects for sentence complement *that*s. Speakers were more likely to include a *that* in a target sentence (e.g., *The mechanic mentioned that the car could use a* *tune up*) when the preceding prime sentence contained a that (*The company insured that the farm was insured for two million dollars*) than when it did not. However, not just any old *that* will do; for targets with sentence complement structures, primes containing other kinds of *that*s (determiners and noun complements) were ineffective. Ferreira used this finding to argue that *that*-priming is a syntactic effect and cannot simply be explained in terms of lexical repetition. In other words, it is the abstract syntax of a sentence which is primed, not just whether the sentence contains a particular lexical item (for similar conclusions see Bock, 1987, Bock & Loebell, 1990, and Bock, Loebell and Morey, 1992).

EXPERIMENT 6
Optimal Design in *That*-Mention

6.4 Rationale and Predictions

In 6.3 we saw that *that*-mention can be primed. In V. Ferreira's (2003) experiments, speakers produced about 25% more sentence-complement *that*s after a prime containing a *that* than after a prime omitting a *that*. This effect is within-speaker; participants produce both prime and target. Although *that*-mention seems to be primable, it does not appear to be motivated by addressee needs. Studies by Ferreira and Dell (2000) and Kraljic and Brennan (2003) suggest that speakers do not include optional function words more often if an utterance could potentially cause a listener to garden-path. This finding seems to extend to ambiguity avoidance more generally; the production system appears to be geared up to favour accessibility over ambiguity-avoidance (Arnold et al., 2004; Ferreira & Dell, 2000). In other words, utterances tend to be optimally designed from the speaker's point of view, but not necessarily from the addressee's.

There were two aims to Experiment 6. The first was to look at optimal design with respect to *that*-mention. Experiments 2, 3, 4 and 5 all suggest that speakers can and do produce word orders and syntactic structures which are "optimal" (from the addressee's perspective) given the context. On the basis of previous findings (Ferreira & Dell, 2000; Kraljic & Brennan, 2003), however, it would be sensible to predict that speakers will not use a rather subtle ambiguity-avoidance strategy, *that*-mention, in such an optimal way. Whereas Experiments 2-5 investigated speakers' ordering of *content* words (nouns and adjectives), Experiment 6 looks at use of a function word, *that*, which carries little meaning and is probably less likely to be under a speaker's strategic control. So *that*-mention gives us an interesting way to look at optimal design "for" the addressee in a domain of linguistic behaviour that is probably outside the speaker's explicit awareness.

The second aim was to replicate Ferreira's *that*-priming effect in a dialogue task. Previous work on syntactic priming has focused on alternating syntactic structures, such as the double object/prepositional object alternation (e.g., Pickering & Branigan, 1998), prenominal versus postnominal modification in NP structure (Cleland & Pickering, 2003), or actives versus passives (e.g., Bock, 1986). Experiments show that these abstract syntactic structures can be primed within-speaker, in "monologue" tasks like picture description (e.g., Bock, 1986) or sentence completion (Hartsuiker & Westenberg, 2000). Priming effects also obtain between-speakers, in dialogue tasks (e.g., Branigan, Pickering & Cleland, 2000). V. Ferreira's (2003) findings extend priming effects to a new syntactic domain, the use of optional function words, but only in monologue (his participants produced both primes and targets). We cannot assume *a priori* that priming effects for grammatical function words will also occur between speakers, in a dialogue. So Experiment 6 investigates *that*-priming in an interactive task, similar to the one reported in Chapters 4 and 5.

In Experiment 6, participants played a communication game with a confederate who was scripted to produce utterances (potentially) containing a PP-attachment ambiguity. The confederate explicitly disambiguated the PP in half of her utterances by using the optional complementizer *that* (e.g., *Put the penguin <u>that's</u> in the cup on the square*). On the following turn, participants either gave an instruction for an array containing one referent (one pig) or two referents (two pigs, one on a block and one not on a block). Utterances were analysed for evidence of priming (using a *that* more often after a prime containing a *that*) and an effect of context (using *that* more often in two-referent contexts, where the listener could potentially be garden-pathed by a syntactically ambiguous instruction). If the production system is set up to produce utterances that are easy to understand, then regardless of priming, naïve speakers ought to use optional *that*s more often in two-referent contexts than in one-referent contexts, in order to avoid potential ambiguities. But if production is designed to be easy for the speaker, then we might expect to see a priming effect but no effect of referential context.

6.5 Method

6.5.1 Participants

30 volunteers from Edinburgh University's student community were paid to participate. They were all native speakers of British or American English.

6.5.2 Design

Two factors were manipulated in Experiment 6: whether or not the confederate's prime instruction contained a *that*, and whether the participant's subsequent target instruction was to be produced in a two-referent context (e.g., two penguins, one stood in a cup and one alone, plus an empty cup) or a one-referent context (just one penguin in a cup). These two factors were crossed to yield a 2 x 2 within-participants design with four conditions:

1) no-*that* prime, one-referent target context
2) no-*that* prime, two-referent target context
3) plus-*that* prime, one-referent target context
4) plus-*that* prime, two referent target context

There was an additional between-participants manipulation; whether the confederate was "helpful" in terms of when she produced *that*s in her instructions. A helpful confederate included *that* whenever she produced an instruction that referred to a two-referent context. An unhelpful confederate produced *that* only in one-referent contexts. In other words, the confederate always used *that* to avoid a potential garden path in the helpful condition, but never in the unhelpful condition. This manipulation was included to provide a comparison with Experiment 5, which found effects of the confederate's behaviour on speakers' use of helpful and unhelpful word orders within noun phrases. Half of the participants interacted with a helpful confederate, and half interacted with an unhelpful confederate. Note that the same person played the confederate throughout the experiment, the only difference was in the script she was given to read in the Helpful and Unhelpful conditions.

6.5.3 Apparatus and Materials

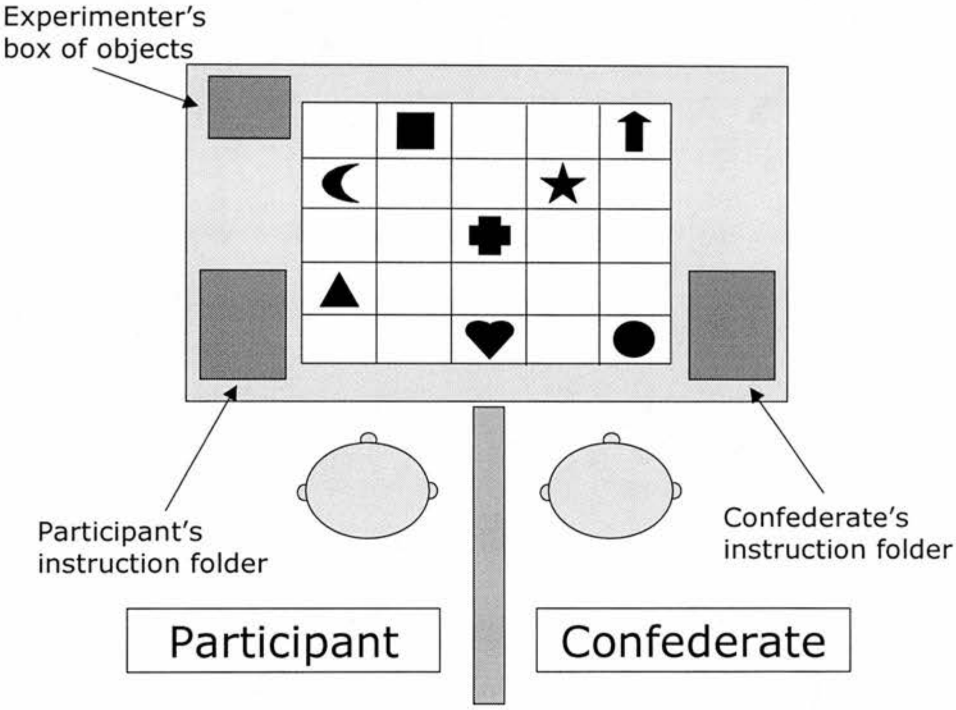
The playing area for the game was a 5 x 5 laminated cardboard grid with geometric shapes (square, arrow, crescent, star, cross, triangle, heart and circle) pasted into eight of the cells (see Figure 15). During the communication game, a player's task was to help their partner move a set of objects around on the playing area as quickly and as accurately as possible. Eight kinds of object were moved about during the game; four of those objects (a plastic penguin, a plastic pig, a metal key and a round tea-bag) appeared only in experimental items, and four (a plastic sheep, a plastic rabbit, a pencil sharpener, and a large paperclip) appeared only in fillers. In addition, six kinds of container were on the board during the game. Four containers appeared only in experimental items (a glass jar, a cup, a wooden block, and a plastic slice of bread from a children's tea-set) and two further containers (a green plastic plate and a small plastic tumbler from a children's tea-set) appeared only in fillers. I will refer to the objects and containers used in experimental items as *experimental objects*, and to entities and containers used only in fillers as *filler objects*. The complete set of objects used for the experiment contained two of each of the experimental objects (two penguins, two pigs, two jars, two cups, etc.) and one each of the filler objects (one sheep, one rabbit, one plate, etc.).

Two sets of instruction cards were created. The confederate's instructions (the prime cards) had sentences printed on them, such that she simply had to read each instruction aloud from the page. There were 24 experimental primes, all of the form *Put the OBJECT (that's) in/on the CONTAINER on the SHAPE* (e.g., *Put the pig (that's) on the block on the heart*). Half of the confederate's experimental instructions contained *that* and half did not; these were equally distributed across the four experimental conditions (see 6.5.2). 24 filler instructions were also created. Half of the fillers included a container object (e.g., *Put the sheep on the plate on the star*) and half did not (*Put the paperclip on the circle*). None of the confederate's filler primes contained a *that*.

Digital colour photographs were taken of all the objects standing alone and inside and on top of the appropriate containers. For example, the penguin (an experimental object) was photographed alone, standing inside the glass jar, inside the cup, on top of the wooden block, and on top of the plastic slice of bread. The sheep (a filler object) was photographed alone, inside the tumbler, and on top of the plate. These photographs were used to make the participants' instruction cards (the target set). Each target card showed a miniature version of the 5 x 5 playing area, with photographs in the various cells to represent each object's position on the board.

FIGURE 15

Experiment 6: Overhead view of the experimental set-up



In each of the participant's target cards, one cell on the board was highlighted with a yellow box; this was to show which object the participant should tell their partner (the confederate) to move on that turn. One of the geometric shapes was also highlighted with a yellow box; this was to show the object's target location. For example, yellow boxes around a pig on a block and around a star would indicate that the participant should tell their partner to *Put the pig (that's) on the block on the star*. The prime instructions and target instructions were printed on A4 paper, one instruction per page, and filed in separate ring binders for the confederate and participant to use during the game.

24 pairs of instruction cards were created by pairing a prime card from the confederate's set with a target card from the participant's set. Each experimental object appeared six times in a prime and

six times in a target. Across primes and targets, experimental objects were counterbalanced such that each entity (penguin, pig, key and tea-bag) appeared with each container (jar, cup, block, bread) an equal number of times. The objects mentioned in a prime and to-be-described objects on the corresponding target card were always different. For example, if the confederate's prime instructed the participant to *Put the penguin in the jar on the square*, then the participant's target card did *not* show either a penguin or a jar. As in Experiments 2, 3, 4 and 5, this constraint was designed to rule out a simple lexical repetition explanation for any priming effects.

An experimental item in this experiment was defined as the confederate's prime instruction plus the participant's paired target card. So, there were 24 unique items in total, six in each of the four experimental conditions (see Appendix B). A mistake in the experimental design meant that items were not fully rotated through conditions, so the *F2* analyses reported in 6.6 are between-items comparisons. The 24 prime-target pairs and 24 fillers were assigned to 12 experimental blocks (two experimental items and two fillers per block), which were presented as 12 'games' during the experimental session. Across the experiment, the order in which fillers and experimental items were presented within each block was rotated.

6.5.4 Procedure

The confederate and the participant sat on the same side of a table, facing the board, with a vertical screen between them so they could not see either the other player or their instruction cards (see Figure 15 for an overview of the experimental set-up). Both players read instructions explaining the aim of the communication game: to help their partner move objects onto the geometric shapes on the board, by giving verbal instructions. They alternated between two tasks: describing instruction cards to their partner, and moving objects around on the board. The players were told that when it was their turn to give an instruction, their job was to tell their partner which object to move, and where to put it. They were told that the to-be-moved object would sometimes be inside or on top of something else, but that they should only move the topmost object; the container object itself was to stay behind.

The confederate and the participant played this game for a total of twelve rounds. In each round, the confederate and the participant described four instruction cards (two experimental items and two fillers) each. The confederate always gave the first instruction. After the confederate carried out the participant's final instruction, the board was set up ready for the next round. Objects not used during a particular round (duplicate experimental objects) were hidden out of the players' view until they were needed again. Participants' responses were transcribed verbatim during the experimental session. Sessions were also recorded on audio or DAT tape as back-up. Unfortunately recordings for the first four participants were lost due to equipment failure, but the other recordings were checked against the transcripts made during the experiment and were found to be accurate.

6.5.5 Post-experiment questionnaire

At the end of communication task, participants filled out a questionnaire which was designed to tap into their intuitions about their own language behaviour, and the nature of the communication task. On the first page of the questionnaire participants were asked to write down what they thought the experiment might be about. The following pages asked a series of questions about the communication task, and about how they and their partner played the game. Some of the questions were open-ended (such as, *Was the game easier on some turns than others? If so, why?*) and required a short written response. For example one of the questions asked whether the participant "did anything special to make the game easy" for their partner when they gave instructions (the full questionnaire is included in Appendix F). Open-ended questions were intended to find out how aware participants were of the priming and context manipulations, whether they noticed the potential garden path, and whether they had consciously used *that*-mention or other ambiguity-avoidance strategies in order to disambiguate their instructions. Other questions provided a 10-point rating scale and participants were asked to circle the point on the scale that reflected their opinion about some aspect of the task, or their partner's (the confederate's) behaviour. Rating scale questions were primarily included in order to find out how the confederate's behaviour (helpful versus unhelpful) affected participants' feelings and opinions about the task. The experimental session lasted about 35 minutes in total, including time to fill out the questionnaire.

6.6 Results

6.6.1 Items analyses

Due to a mistake in the experimental design, items were not fully rotated between conditions, so the *F2* analyses reported below are between-items comparisons. Items in the Helpful and Unhelpful conditions are counted separately, because a particular prime produced by the helpful confederate belongs to a different condition from the same prime produced by the unhelpful confederate. Thus in the *F2* analyses, there are 12 items per condition, 6 from the helpful confederate and 6 from the unhelpful confederate.

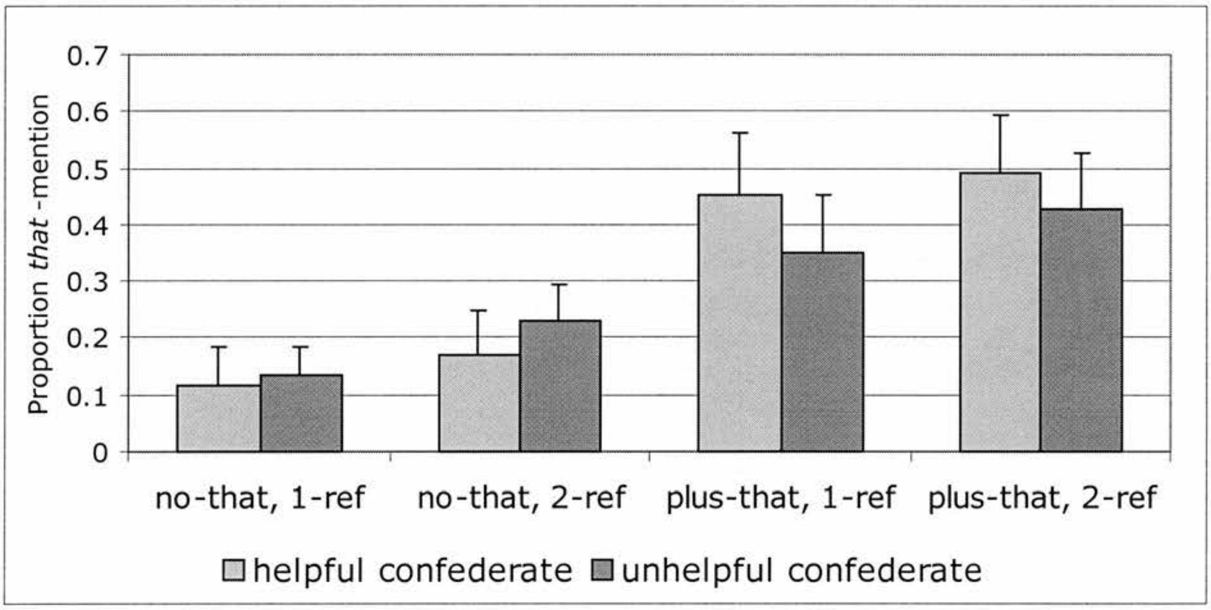
6.6.2 Basic analysis: Speakers' use of *that*-mention

The dependent measure of interest is how often speakers produce *that*s in their target instructions: do they produce more *that*s following plus-*that* primes from the confederate (*Put the pig that's on the block on the heart*), and are they more likely to insert a *that* when they describe two-referent contexts, which could potentially cause a garden path? Speakers' responses were therefore initially coded according to a strict coding scheme; any instructions *not* of the form *Put the OBJECT (that's) in/on the CONTAINER on the SHAPE* were coded as *others*. Responses classified as *others* according to this coding scheme included instructions that contained grammatical function words other than *that*, such as *Put the penguin from the cup on the square* or *Put the key which is in the cup on the square.* Target responses were sometimes not preceded by the intended prime (usually because the participant had turned over more than one page at a time in their instruction folder), and these responses were coded as *others*. Responses where the participant did not mention the container at all (e.g., *Put the penguin on the star*) were also classified as *others*. 117 responses (out of 720, or 16.25% of the data) were coded as *others* using these criteria. The remaining 603 responses were coded using two categories: plus-*that* (response includes a *that*) and no-*that* (response does not include a *that*). *Other* responses are excluded from the basic analysis, such that the proportion of plus-*that* and no-*that* responses is complementary.

Speakers included a *that* in about 30% of their target instructions overall. Figure 16 shows the proportion of *that*s produced in each of the four experimental conditions: no-*that* prime, one-referent target context; no-*that* prime, two-referent target context; plus-*that* prime, one-referent target context; plus-*that* prime, two-referent target context. The data were submitted to 2 (Prime Type: No-*that* versus Plus-*that*) x 2 (Context Type: One-referent versus Two-referent) x 2 (Confederate Type: Helpful versus Unhelpful) analyses of variance with Prime Type and Context Type as within-participants factors, and Confederate Type as a between-participants factor. *F1*s are within-participants comparisons, while *F2*s are between-items comparisons.

FIGURE 16

Experiment 6: Proportion *that*-mention, by condition (basic analysis)



The results replicate V. Ferreira's (2003) finding that the use of optional function words can be primed. Speakers included about 27% more *that*s in their target instructions after hearing primes which contained *that*s from the confederate ($F1(1, 28) = 23.84$, $MSe = .091$, $p < .001$; $F2(1, 40) = 87.03$, $MSe = .011$, $p < .001$). Somewhat surprisingly, there was also a main effect of Context in this experiment. Speakers produced about 7% more *that*s in their instructions for two-referent target contexts than for one-referent target contexts. This effect is robust in the by-participants analysis ($F1(1, 28) = 6.89$, $MSe = .019$, $p = .014$) but not in the (weaker) by-items analysis ($F2(1, 40) = 3.18$, $MSe = .011$, $p = .082$). There was no interaction between Prime and Context ($F1(1, 28) = .066$, $MSe = .027$, $p = .799$; $F2(1, 40) = .043$, $MSe = .011$, $p = .837$).

Speakers who interacted with a Helpful confederate produced about 33% more *that*s following primes containing *that*s, and they included a *that* about 5% more often in two-referent contexts than in one-referent contexts. Speakers who interacted with an Unhelpful confederate showed a smaller priming effect (21%) and a slightly larger effect of context (about 9% more *that*s in two-referent contexts than in one-referent contexts). The interaction between Confederate and Prime was non-significant by participants but significant by items ($F1(1, 28) = 1.25$, $MSe = .091$, $p = .274$; $F2(1, 40) = 6.68$, $MSe = .011$, $p = .014$). No other interactions were significant (Confederate x Context: $F1(1, 28) = .560$, $MSe = .019$, $p = .461$; $F2(1, 40) = .963$, $MSe = .011$, $p = .332$; Prime x Context x Confederate: $F1(1, 28) = 0.0$, $MSe = .027$, $p = 1.0$; $F2(1, 40) = .084$, $MSe = .011$, $p = .773$).

6.6.3 Speakers' use of other grammatical function words

The previous analyses were restricted to responses of the form *Put the X (that's) in/on the X on the X*; all other responses were coded as *others* and excluded. But speakers sometimes produced responses that were syntactically disambiguated using other kinds of grammatical function word, such as *from* (*Put the penguin from the cup on the star*) and *which* (*Put the penguin which is in the cup on the star*). Although these grammatical words almost certainly convey subtly different meanings (i.e., different from *that* and different from each other), they *function* in much the same way as *that*, in the sense that they block a goal interpretation for the following prepositional phrase, and therefore avoid a garden path.

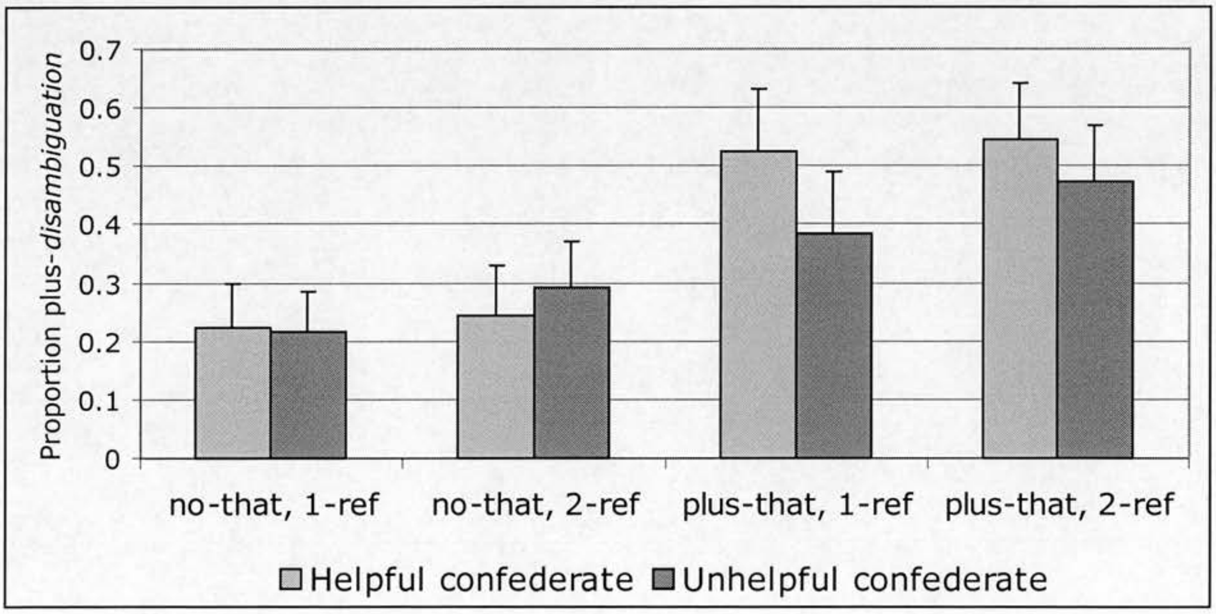
Secondary analyses were conducted in which any response that was disambiguated by a grammatical function word (*that*, *from* or *which*) was coded as plus-*disambiguation*. Responses in which the PP following the head noun was syntactically ambiguous (e.g., *Put the penguin in the cup on the square*) were coded as no-*disambiguation*. Responses in which either the container was not mentioned (e.g., *Put the penguin on the square*), or in which the target was not preceded by the intended prime, were the only targets coded as *others*. *Other* responses account for 6.81% of the data (49 out of 720 trials) using these criteria, and again they are excluded from the following analyses. Proportions of disambiguated and ambiguous responses are therefore complementary, and only the proportion of disambiguated responses is reported below. Averaging across conditions, speakers disambiguated about 36% of their instructions using a grammatical function word (*that*, *from*, or *which*). Figure 17 shows the proportion of disambiguated responses produced in each experimental condition.

These data were submitted to analyses of variance with the same factors described above. Again, there was a very robust effect of Prime. Speakers produced about 24% more disambiguated target instructions following a prime instruction containing a *that* ($F1(1, 28) = 25.11$, $MSe = .068$, $p < .001$; $F2(1, 40) = 58.05$, $MSe = .012$, $p < .001$). In this secondary analysis, there was no effect of context. Speakers produced disambiguated target instructions in about 5% more two-referent contexts than in one-referent contexts. However, this difference was not significant ($F1(1, 28) = 2.83$, $MSe = .026$, $p = .104$; $F2(1, 40) = 1.93$, $MSe = .012$, $p = .173$). As in the previous analysis, there was no interaction between Prime and Context ($F1(1, 28) = .012$, $MSe = .018$, $p = .913$; $F2(1, 40) = .002$, $MSe = .012$, $p = .969$). No other interactions reached significance (Prime x Confederate: $F1(1, 28) = 1.76$, $MSe = .068$, $p = .195$; $F2(1, 40) = 3.54$. $MSe = .012$, $p = .067$; Context x Confederate: $F1(1, 28) = 1.08$, $MSe = .026$, $p = .308$; $F2(1, 40) = 1.93$, $MSe = .012$, $p = .173$; Prime x Context x Confederate: $F1(1, 28) = .032$, $MSe = .018$, $p = .859$; $F2(1, 40) = .049$, $MSe = .012$, $p = .826$).

FIGURE 17

Experiment 6: Proportion of responses coded as plus-*disambiguation*

(secondary analysis)



6.6.4 Questionnaire data

At the end of the communication task, participants filled out a questionnaire about their experiences in the experiment. Some of the questions asked participants to rate the task, the confederate's behaviour, or their own behaviour on a 10-point scale. Other questions were open-ended and invited the participant to write a free response. The full questionnaire can be found in Appendix F.

The open-ended questions were phrased in quite a general way, and it was sometimes to difficult to decide whether participants' responses showed explicit awareness of some aspect of language behaviour or not. For example, answering a question about whether the game was easier on some turns than others (when following the confederate's instructions), one participant wrote: *If the instructions were simple.* From this short answer, it is impossible to know what the participant meant by "simple". They may have been referring to instructions from the confederate that were simple to understand by virtue of a disambiguating *that.* Alternatively, they may just have been referring to filler trials where there was no container, such as *Put the paperclip on the star.* Structured questions would need to be asked to find out for sure whether participants were aware of the syntactic ambiguity or not. However, an exploratory analysis of participants' responses suggests that they were not aware of the *that*-priming manipulation. Several participants said that the experiment might be about how one person's language can influence another's, but these seemed to be references to *lexical* alignment (saying 'moon' instead of 'crescent', for example), or about how they copied the general form of the confederate's instruction (*Put the X in the X on the X*). Certainly none of the responses made explicit reference to the confederate's use of *that.*

One participant (in the Unhelpful condition) referred to the potential garden path. In answer to a question about things which made it difficult to follow the confederate's instructions, he wrote: "*Confusion between the object's current placing and the intended destination – both expressed with the word 'on'.*" In response to a question about whether they had done "… anything special to make the game easier for your partner" another participant (this time in the Helpful condition) reported saying: "*…the pig <u>that's in</u> the jar, instead of <u>in</u> the jar sometimes*" (underlining in the participant's answer). But although other participants mentioned describing the container objects to "help" their partner find the thing they were looking for, or the need to give more information in the two-referent contexts, none of the other responses made explicit reference to syntactic ambiguity, or using *that*-mention to avoid it. So participants seemed to be generally aware that they should avoid referential ambiguities (by specifying which penguin they were talking about in a two-penguin context, for example), but for the most part they were not explicitly aware of using *that*-mention to avoid garden paths.

The same confederate interacted with all of the participants, and the only difference between the Helpful and Unhelpful conditions was in her use of *that*. In the helpful condition, the confederate was scripted to include a *that* only in two-referent contexts. For example, if there were two penguins (one in a cup, one not in a cup) she produced *Put the penguin that's in the cup on the star*. In a context with only one penguin she produced a prime without a *that*, such as *Put the penguin in the cup on the star*. In the unhelpful condition, this pattern was reversed; the confederate only included a *that* in one-referent contexts, never in two-referent contexts. So in the example above, she produced *Put the penguin that's in the cup on the star* in a context containing only one penguin, and *Put the penguin in the cup on the star* in a context containing two penguins. This is rather a subtle behavioural difference, and one which none of the participants seemed to pick up on (or at least, they didn't explicitly mention it in their responses to the questionnaire). Nonetheless, a subtle difference in language behaviour could in principle affect how participants responded to the confederate, and to the task. Perhaps the participants would find the task more difficult or less enjoyable if the confederate behaved "unhelpfully" with respect to *that*-mention. This is the kind of issue that the ratings scales were intended to address.

Table 7 compares participants' mean ratings in the Helpful and Unhelpful conditions. The confederate's behaviour had no effect on ratings of how enjoyable or simple the task was, but participants in the Helpful condition rated the task as marginally easier than participants in the Unhelpful condition ($t(28) = 1.993$, $p = .056$). For questions about "how [the confederate] played the game", ratings on most of the scales were similar across the Helpful and Unhelpful conditions, but two pairs of adjectives did show interesting differences. Participants in the Helpful condition rated the confederate as more friendly ($t(28) = 2.035$, $p = .051$) and more flexible ($t(28) = 3.044$, $p = .005$). None of the other comparisons approached significance (all *t*s < 1.5). So it seems as though manipulating a speaker's linguistic behaviour can have consequences for how their behaviour is perceived more generally.

TABLE 7

Experiment 6: Participants' ratings of the task and how the confederate played the game, by condition (Helpful versus Unhelpful)

	Helpful condition		*Unhelpful condition*	
Rating scale	Mean	(SD)	Mean	(SD)
Task				
Not Enjoyable – Enjoyable	6.67	(1.35)	7.00	(1.46)
Easy – Difficult	2.13	(0.83)	3.20	(1.90)
Simple – Complicated	2.33	(1.18)	3.00	(2.00)
How partner played the game				
Friendly – Unfriendly	2.93	(1.39)	4.20	(1.97)
Unintelligent – Intelligent	8.73	(1.28)	7.80	(2.18)
Inflexible – Flexible	7.67	(1.59)	5.67	(1.99)
Fast – Slow	2.87	(1.19)	3.33	(1.45)
Part of a team – Not part of a team	3.00	(1.41)	3.00	(1.60)
Not Rigid – Rigid	5.80	(2.98)	4.93	(1.91)
Cooperative – Uncooperative	8.80	(1.15)	8.13	(1.60)
Skilled – Unskilled	2.73	(1.75)	3.60	(1.88)

6.7 Discussion

Experiment 6 investigated the conditions under which speakers include optional function words to disambiguate syntactic structure. Specifically, the aim was to find out whether optional *that*-mention is affected by optimal design. Do speakers include optional *that*-complementizers more often when their utterance could potentially cause a garden path effect? A secondary aim was to find out whether V. Ferreira's (2003) priming effect for optional function words extends to dialogue. Are speakers more likely to include a *that* in their utterance when they have just heard an interlocutor use a *that*?

6.7.1 *That*-mention can be primed in dialogue

In the data presented here, *that*-mention was primed by hearing an interlocutor use a *that* in their immediately preceding utterance. Speakers were more likely to include a *that* in an utterance following a prime containing a *that* from the confederate. A similar priming effect has recently been reported using a sentence-recall task (V. Ferreira, 2003). The priming effect reported here, albeit for a different kind of optional *that*, suggests that grammatical function word priming can occur in dialogue (i.e., comprehension-to-production) as well as monologue (production-to-production). That is, interlocutors can prime each other to use constructions that include or omit grammatical function words, just as they prime each other to use particular lexical items (Brennan & Clark, 1996) or syntactic structures (Branigan et al., 2000; Cleland & Pickering, 2003). It is interesting to note that the between-speaker *that*-priming effect reported for noun phrase complementizers here is as large as the within-speaker priming effect reported by V. Ferreira (2003) for a different kind of *that*-complementizer. Participants in Experiment 6 produced about 27% more *that*s in their instructions after primes containing *that*s, compared to primes without *that*s; Ferreira reported about 25% priming on average for sentence complement *that*s.

6.7.2 *That*-mention can be sensitive to optimal design?

Experiment 6 provides some preliminary evidence suggesting that *that*-mention is sensitive to ambiguity. Speakers produced about 7% more *that*s in their instructions for an array containing two referents (e.g., two penguins) than when the array contained only one referent (just one penguin). Two-referent contexts support both of the possible interpretations (goal versus modifier) for structurally ambiguous PPs, but in one-referent contexts, the visual array biases comprehension towards the modifier interpretation. Ambiguously-attached PPs (like *in the cup*, in an instruction like *Put the penguin in the cup on the star*) are therefore more likely to cause temporary comprehension difficulties for addressees when there are two potential referents in the array than when there is just one. So, the fact that speakers seem to produce more *that*s in two-referent contexts than in one-referent contexts suggests that *that*-mention may be sensitive to optimal design. If the context effect reported here is reliable, then it would suggest that language production (or specifically, optional word mention) *can* be sensitive to at least some kinds of ambiguities. This finding is surprising given the lack of ambiguity-avoidance effects in previous research (e.g., Ferreira & Dell, 2000; Kraljic & Brennan, 2003), and a number of caveats need to be considered.

Firstly, items were not rotated through conditions in the design of this experiment, and the context effect could in principle be an artefact of the different items randomly assigned to the four conditions. Moreover, the context effect only holds when we look only at speakers' use of *that*, as opposed to other optional function words. The basic analysis in Experiment 6 showed an effect of context, at least in the by-participants analysis. Speakers produced *that*s more often in two-referent contexts than in one-referent contexts. However, a secondary analysis which included function words other than *that* (e.g., *Put the penguin from the cup on the star*, or *Put the penguin which is in the cup on the star*) found no significant effect of context. Speakers did not produce a significantly greater number of optional function words in utterances that would be temporarily ambiguous without them. These data therefore do not support the claim that production is generally ambiguity-avoidant. It seems that if the production system shows a tendency towards ambiguity-avoidance, then the tendency is weak (cf. Ferreira & Dell, 2000). It may also be dependent on exposure to a particular kind of ambiguity-avoidance strategy. Participants in Experiment 6 heard the confederate use *that*s consistently in either one- or two-referent contexts. Perhaps this helped them to appreciate when and how *that*s could best be employed (cf. Horton & Gerrig, 2002). If this is the case, then whatever speakers are learning about *that*-mention seems to be fairly abstract; the questionnaire data suggest that most participants were not able to verbalise how *that*-mention in potentially ambiguous contexts might aid comprehension.

If the context effect reported here is reliable, then the findings of Experiment 6 would be at odds with previous studies investigating the use of optional function words for syntactic disambiguation. Using a sentence recall-based paradigm, Ferreira and Dell (2000) found that speakers were no more likely to include a (sentence complement) *that* in utterances that would otherwise be ambiguous than in utterances that would not be ambiguous. An unpublished referential communication study by Kraljic and Brennan (2003) also found that speakers do not include a (noun-phrase complement) *that* more often in "ambiguous" contexts, where there is a possibility for misinterpretation, than in "unambiguous" contexts (T. Kraljic, personal communication). How could the context effect found in Experiment 6 be reconciled with these null findings?

Clearly, there are several important differences between the current experiment and previous studies. Ferreira and Dell (2000) investigated optional *that*-mention in sentence complement structures (*The coach knew (that) you had missed practice*), where garden path effects seem to be relatively weak (Sturt et al., 1999). It may be that Ferreira and Dell failed to find evidence for ambiguity avoidance because speakers are aware at some level (though not necessarily consciously) that omitting a sentence complement *that* is unlikely to cause severe comprehension difficulties (but note that Ferreira and Dell's Experiment 3 tested ambiguity avoidance in reduced relative clause constructions – such as *the astronauts selected for the Apollo missions made history* – which are associated with greater comprehension difficulties; Gorrell, 1995; Pritchett, 1992). In contrast, Experiment 6 investigated ambiguity-avoidance in utterances like *Put the penguin in the cup on the star*, where biases associated with the verb *put* are known to cause stronger garden path effects (e.g., Tanenhaus et al., 1995; Trueswell et al., 1999). If the language processing system is sensitive to the relative difficulty of parsing different kinds of ambiguous constructions, then it may be more likely to include optional function words in utterances that are more likely to lead addressees up the garden path.

The sentence-recall paradigm used by Ferreira and Dell was basically non-interactive (although see their Experiment 6), and there was no obvious communicative consequence of producing syntactically ambiguous utterances. Ferreira and Dell (2000, p. 329) themselves point out that their experiments did not put speakers in a situation where it was important to avoid ambiguities:

> *However, this does not mean that speakers never avoid difficult-to-comprehend utterances. Specifically, the present evidence does not speak to the possibility that speakers cooperatively adjust the syntax of their utterances when they are aware that comprehension failure is likely.*

The issue of awareness may be crucial here. In Ferreira and Dell's experiments (unlike the experiment reported in this chapter), there was no visual context available to draw the speaker's attention to linguistic ambiguities. Indeed, there was no referential ambiguity involved, as there was in Experiment 6. Instead, for Ferreira and Dell's speakers to become aware of the potential garden path during utterance planning, all of the possible syntactic analyses for a to-be-produced utterance would have to be internally generated and compared against each other, probably by the internal monitor (Levelt, 1983; 1989). If more than one analysis is found to be compatible with the planned output then the processor can infer that the utterance it is about to produce is ambiguous. Intuitively, it seems as though this kind of production strategy would be pretty inefficient, because almost all sentences and utterances are temporarily ambiguous at some level of representation, and constantly monitoring for all kinds of ambiguity would presumably be laborious and resource-intensive. Indeed, recent unpublished evidence suggests that the production system is not very good at monitoring for purely linguistic ambiguities prior to speaking (e.g., Ferreira, Slevc, & Rogers, 2003; Kraljic & Brennan, 2003, Experiment 2). As Snedeker and Trueswell (2003, p.128) put it:

> *...it seems likely that ambiguity awareness, and attempts to disambiguate ambiguity, may be tied to the level of representation at which the ambiguity arises. One should expect that non-linguistically trained speakers will be far more aware of referential ambiguity than they are of syntactic ambiguity.*

Snedeker and Trueswell's intuition seems plausible, and may well be right. Questionnaire data from Experiment 6 suggested that participants were sometimes aware of the need to avoid the potential *referential* ambiguity. They knew, for example, that it was important to specify which penguin they were talking about in a two-penguin context. But they hardly ever mentioned the need to avoid potential syntactic ambiguities. Only two participants wrote anything that showed explicit awareness of the potential PP-attachment ambiguity.

Appealing to different levels of representation would help to explain the discrepancy between the results of Experiment 6 and those of Ferreira and Dell (2000). But in an unpublished study, Kraljic and Brennan (2003) also report a null effect of context on speakers' use of *that* to avoid PP-attachment ambiguities, in a referential communication task that was very similar the one reported here. If the production system can be sensitive to potential ambiguities, then why did Kraljic and Brennan fail to find an effect of optimal design?

A notable difference between Kraljic and Brennan's experiment and the one reported here is that although their task was interactive (in the sense that the speaker was actually trying to communicate a message to a co-present interlocutor), their participants did not dynamically alternate between the roles of speaker and addressee. Instead, Kraljic and Brennan's participants were the speaker *or* the addressee for the first ten trials, and then they switched roles for a further ten trials. So, half of their participants did not get to be on the receiving end of temporarily ambiguous or disambiguated instructions until the final part of the experiment. By this point, if they realised (consciously or otherwise) that including a *that* might make an utterance easier to understand, it was too late to put that knowledge into practice – they had already had their turn at being the speaker. In contrast, participants in Experiment 6 alternated between giving and following instructions. Perhaps this helped to serve as a reminder of how *that*-mention could be used to explicitly disambiguate potentially ambiguous instructions.

So far, I have assumed that speakers' tendency (in Experiment 6) to produce more *that's* in "ambiguous" contexts reflects a consideration of the addressee's communication needs. The idea is that speakers appreciate that two-referent contexts (more than one-referent contexts) could cause confusion, and that they are therefore more likely to produce disambiguated syntactic structures. But there is an alternative explanation. Ferreira and Dell (2000) noted that speakers' use of an optional *that* in sentence complement structures was affected by the accessibility of material in the critical sentences. Speakers were *less* likely to insert a *that* if the subject of the embedded clause was the same as the subject of the main verb (e.g., *You knew (that) you missed practice*), suggesting that language production operates according to a *principle of immediate mention*. In other words, the production system prefers to get highly accessible linguistic material "out there" as quickly as it can. Where the embedded subject is different from the subject of the main verb (and is therefore less accessible), *that* is more likely to be inserted, perhaps as a kind of filler while the rest of the utterance is being formulated.

An explanation based on accessibility could in principle account for the context effect in Experiment 6, although such an account is less convincing here. If speakers' use of *that's* reflects some kind of processing difficulty (e.g., because of additional encoding complexity in the two-referent array), intuition suggests that such difficulty would be more likely to show up at the beginning of the complex noun phrase, rather than mid-phrase. For example, speakers might pause or insert *um* or *er* either at the start of their utterance, or right before the head noun (e.g., *Er, put the um... penguin on the block on the star*). If it were indeed a filler at the point it appeared in the utterance, *that's* would tend to suggest difficulty formulating the first prepositional phrase, *on the block*, but it is unclear why the prepositional phrase would be harder to encode in the two-referent context than in the one-referent context. It might make sense to assume that speakers used *that's* as a filler if they were deciding whether or not to mention the container object at all. From the point of view of avoiding a *referential* ambiguity, mentioning the container is only necessary in the two-referent context. But since the analyses reported above excluded cases where the container was not mentioned (utterances without containers were coded as *others*), this cannot explain the context effect.

6.7.3 Effects of an unhelpful interlocutor

Unlike Experiment 5, Experiment 6 found no clear effects of an interlocutor's language behaviour on either priming or optimal design. This contrasts with Experiment 5, which found reliable effects of the confederate's language behaviour on the likelihood of a naïve participant being primed or engaging in optimal design. Of course, failure to find interactions with Confederate Type here could in principle be a consequence of the between-participants design; confederate effects could be obscured by between-speaker differences, if such individual differences were large, and confederate effects were weak. But on the assumption that individual differences were not enough to obscure a systematic confederate effect, the null effect of the confederate's behaviour is interesting for two reasons.

Firstly, the fact that the confederate's behaviour (helpful or unhelpful) did not appear to interact with priming in Experiment 6 suggests that the "recoding" explanation for the anti-priming effect in Experiment 5 could be right (see Chapter 5). Participants who heard unhelpful prime descriptions in that experiment may well have been mentally recoding word order in order to map it onto the box from which they were selecting a card. For example, if they heard *wavy purple star* but were trying to find the card from a box organised by colour, they might have mentally recoded the description they heard as *purple wavy star*. In effect, they may have primed themselves to produce a colour-first NP in their subsequent utterance, even though the prime was intended to push them towards producing a pattern-first word order in their target description. In Experiment 6, it seems extremely unlikely that participants would have mentally recoded "unhelpful" prime instructions by inserting or taking away a *that*. And although the priming effect was (numerically) smaller in the Unhelpful condition than the Helpful condition, nonetheless it was *positive* priming effect (more *that*s after primes containing *that*s).

The second interesting thing about Experiment 6 is that the Unhelpful confederate's behaviour had no effect on whether participants behaved helpfully themselves with respect to *that*-mention. If anything, participants in the Unhelpful condition showed a numerically larger effect of optimal design, although there was no significant difference between the Helpful and Unhelpful conditions. How can we explain this, given the "anti-helpfulness" effect observed for the Unhelpful condition in Experiment 5? One of the motivations for Experiment 6 was to investigate optimal design in a domain of language use that is likely to be outside speakers' conscious awareness. Whether or not an utterance contains a *that* is rather a subtle difference, and in all probability speakers are not using *that*-mention strategically (the questionnaire data from Experiment 6 support this idea). The context manipulation in Experiment 5 was rather less subtle, and speakers were usually explicitly aware of it. They were also pretty good at spotting whether the confederate's behaviour was helpful or unhelpful with respect to word order. Perhaps the "anti-helpfulness" effect in Experiment 5 reflects this awareness; engaging (or not engaging) in optimal design may depend on speakers being aware (a) of what would constitute a "helpful" utterance given the context, and (b) of whether their interlocutor is behaving helpfully towards them. If speakers are not explicitly aware of how *that*-mention might make comprehension easier for an addressee, nor whether their interlocutor is using *that*-mention helpfully themselves (as seems to be the case in Experiment 6), then the confederate's behaviour has no reliable effect. This seems to be an interesting line of enquiry for further research.

Although the confederate's use of *that* (helpful versus unhelpful) had no clear effects on speakers' on-line language behaviour, it did seem to have interesting consequences for how she was perceived. Participants rated the game as easier and the confederate as more friendly and more flexible when she used *that*-mention in a helpful way. Whether the confederate used *that*s in one- or two-referent contexts is rather a subtle behavioural difference, and speakers did not explicitly mention this aspect of her behaviour in their responses to the questionnaire. Nonetheless, it seems that such small and apparently inconsequential differences in linguistic behaviour can affect how people feel about an interlocutor, and about the interactions they have with them.

6.8 Chapter Summary

If the results of Experiment 6 prove to be reliable, then two interesting findings have emerged. The first is that the use of optional function words (specifically, optional *that*s) can be primed in a dialogue task, replicating the results of a recent sentence-recall experiment (V. Ferreira, 2003). Additionally, speakers seem to be able to use the *that*-complementizer to avoid ambiguity in their speech. Participants included more *that*s when they gave instructions that had the potential to garden-path their addressee, given the referential context. This finding is at odds with evidence from previous studies that have investigated ambiguity avoidance and optional word use (e.g., Ferreira & Dell, 2000; Kraljic & Brennan, 2003), and deserves further attention. The confederate's language behaviour had no interactions with either priming or context effects, but it *did* have an interesting impact on participants' attitudes towards their interlocutor, and towards the task. The communication game was rated as easier and the confederate was rated as more friendly and more flexible when she used *that*-mention in a "helpful" way.

CHAPTER 7
Conclusions

7.1 Summary of the Findings

Experiment 1 investigated the effect of cognitive load on optimal design at the level of conceptualisation and lexical choice. Does having two conversations at once make it more likely that speakers will struggle to design referring expressions "for" the current addressee? The results of Experiment 1 suggest that speakers are remarkably good at recalling appropriate conceptualisations established with different interlocutors, even when they are having to switch between two sets of conversational precedents or conceptual pacts. When communication is likely to fail unless the speaker remembers the "right" perspective, they show good evidence of engaging in optimal design. So, knowledge of an interlocutor's perspective seems to have a strong influence on the production of referring expressions, even under conditions of increased cognitive load.

Experiments 2 and 3 investigated optimal design in the production of simple noun phrases, to find out whether grammatical encoding (and specifically word order) shows evidence of optimal design. In the production literature, researchers have mostly focused on optimal design at the level of conceptualisation and lexical choice (e.g., Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992), although some studies also consider effects of an interlocutor's perspective on phonological encoding (Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Bard & Aylett, in press; Gregory, Healy, & Jurafsky, submitted). To date, evidence relating to grammatical aspects of production suggests that optimal design may not have much of a role to play (Arnold, Wasow, Asudeh, & Alrenga, 2004; Branigan, McLean, & Reeve, 2003; Brown & Dell, 1987; Dell & Brown, 1991; Ferreira & Dell, 2000; Kraljic & Brennan, 2003). However, Experiments 2 and 3 suggest that the form of referring expressions may be sensitive to optimal design. Speakers produced word orders that were "incrementally helpful" for their addressee, in the sense that information in the noun phrase could be mapped straight onto the array from which the addressee was choosing picture cards. Crucially, speakers only produced these helpful word orders once they knew, from their own experience as an addressee, what would constitute a helpful or unhelpful utterance for the partner.

Experiments 4, 5 and 6 investigated the relationship between optimal design and another process known to have pervasive effects in language production: syntactic priming. Previous research demonstrates that speakers can be primed to produce different kinds of utterances in dialogue simply by hearing an interlocutor use a particular syntactic construction (e.g., Branigan, Pickering & Cleland, 2000; Cleland & Pickering, 2003). In Experiments 4, 5, and 6, priming effects were more robust than effects of optimal design, which suggests that priming has a more profound influence on what speakers say (or rather *how* they say it) than beliefs about an interlocutor's perspective. These observations about the magnitude of priming and helpfulness effects are important for future research; clearly we should be sensitive to the fact that priming could, in principle, obscure speakers' attempts to dynamically adapt to the addressee's perspective. Moreover, we should take care to design experiments which allow us to investigate the various sub-components of language production separately, as well as studying the interactions and relationships between them when they act in concert. However, even in the context of strong priming effects, evidence for optimal design seems to emerge. Clearly priming is not strong enough to 'knock out' optimal design in production, or at least, not in the kind of task-oriented dialogue studied here. When speakers are put in a situation where it is clear how they can adapt what they say to meet a partner's "needs" (given the task at hand), or where their partner models a particular kind of "helpful" linguistic behaviour, then they tend to show evidence of tailoring utterances to the addressee's perspective.

Chapter 6 addressed a somewhat different question to the previous two chapters. In Experiment 6, the confederate priming technique was again used to assess the relative influence of priming and optimal design on grammatical encoding. But this time the issue was whether speakers included optional function words (specifically, the noun phrase complementizer *that*) to disambiguate utterances that would otherwise be temporarily ambiguous in the communicative context. Whether a partner uses or doesn't use an optional function word is almost certainly less salient than hearing them produce an unusual syntactic form, like *the square that's red*. So, ambiguity avoidance is an interesting issue, because using optional words to disambiguate syntactic form is likely to be outside speakers' conscious awareness. Previous experimental studies had failed to find consistent evidence for ambiguity-avoidance in production, even though speakers have several strategies available to them (Arnold et al., 2004; Ferreira & Dell, 2000; Kraljic & Brennan, 2003). However, Experiment 6 presented some preliminary evidence that speakers may in fact use optional grammatical words in a helpful way, even though off-line questionnaire data suggests that they are not *explicitly aware* of doing so. This finding requires replication before any firm conclusions can be drawn. If it is reliable, however, this would present a strong challenge to the view that *only* conscious, strategic aspects of production can be influenced by beliefs about an interlocutor.

Experiments 5 and 6 also addressed the issue of how an interlocutor's language behaviour can affect the likelihood of a speaker engaging in optimal design. The two experiments came up with quite different results. In Experiment 5, participants who interacted with an Unhelpful partner were significantly "unhelpful" themselves. That is, they tended to produce the *opposite* (prenominal adjective) order to the one that would be most helpful for their partner to hear. In Experiment 6, however, the confederate's behaviour had no direct effect on the participant's use of *that*-mention. These different patterns of results might best be explained by considering the role of conscious, strategic processing in optimal design.

Conscious awareness of what would make for a helpful utterance does not seem to be necessary for optimal design. If it were, then presumably speakers in Experiment 6 would not have used more *that*s in potentially ambiguous contexts than in unambiguous contexts (and they did). But conscious awareness of one's *partner* being helpful or unhelpful seems to have interesting effects. When speakers were explicitly aware of a partner being linguistically unhelpful (Experiment 5), they tended to behave unhelpfully themselves in the same way, even though they knew how they could tailor utterances to the partner's perspective. In fact, in this situation, speakers went out of their way to say *unhelpful* things, perhaps as a kind of conscious "tit-for-tat" strategy. But when the unhelpful aspect of a partner's language behaviour was outside conscious awareness (Experiment 6), this had a much more subtle effect. Participants in a dialogue with an unhelpful interlocutor did not say more unhelpful things themselves. However, they *did* make less positive social judgements about their conversational partners, in comparison with participants who interacted with a helpful interlocutor.

Finally, Experiments 4, 5 and 6 demonstrate that grammatical aspects of utterance production can be primed by a partner's use of particular syntactic constructions and word orders. This is a not a new finding. But Experiment 4 replicates a priming effect for noun phrase structure that has only recently been reported (Cleland & Pickering, 2003), and Experiment 5 presents a challenge for current models of syntactic priming based on the activation of combinatorial information (e.g., Pickering & Branigan, 1998). Such models cannot account for priming effects in prenominal adjective order, because noun phrases like *orange dotty square* and *dotty orange square* would be represented by the same combinatorial node (see 5.11.4). Experiment 6 suggests that function word priming effects reported for monologue (V. Ferreira, 2003) extend to dialogue.

7.2 Directions for Future Research

The role of common ground and optimal design in language production is an issue that has only recently begun to receive serious attention in psycholinguistics. This thesis adds to the growing body of literature suggesting that speakers adapt not only what they say, but also how they say it, to meet the needs of their addressee. The experiments reported here suggest that utterances can be adapted to addressees at several different levels of the production process, including conceptualisation, choice of syntactic form, and generation of word order. However, there are many questions still to be answered.

For example, we cannot definitively rule out an explanation for the findings of Experiments 2 – 5 based on what is salient or accessible for the speaker, rather than what is helpful for the addressee. How do we know that choice of word order or syntactic form here indeed reflects a consideration of the addressee's perspective? The experiments reported in this thesis do not "unconfound common ground" (Keysar, 1997), because utterances that are incrementally helpful for the addressee also put information salient to the speaker in earlier word order positions. In other words, it is possible that what *look like* adaptations to addressees in these experiments are simply reflecting what is easiest for, or most salient to, the speaker (cf. Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996; Pickering & Garrod, in press). It seems unlikely that this is the only explanation for speakers' use of helpful grammatical forms here; questionnaire data suggested that participants were generally aware of how word order or syntactic structure could be used "helpfully", and that they tried to make strategic adaptations to the addressee's perspective (sometimes part-way through their descriptions; see Chapter 5 for discussion). However, further research is needed to pin down the precise nature of these effects. Are adaptations (particularly at the level of grammatical encoding) indeed tailored "for" the addressee? How could we find out?

One way would be to use a variation on the "secret object" paradigm (cf. Hanna, Tanenhaus, & Trueswell, 2003; Horton & Keysar, 1996; Keysar, Barr, Balin, & Paek, 1998; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003; see also Branigan et al., 2003), to create a disparity between the speaker and addressee's (conceptual) perspectives, but such paradigms bring their own set of methodological issues with them. For example, setting up an experimental situation such that the addressee explicitly knows the speaker's knowledge is different from his own is difficult without focussing the addressee's attention on the secret object. This object presumably then becomes conceptually salient in the addressee's model of the visual context, making it difficult for him to completely ignore that object, even if he knows it is an unlikely (or even impossible) referent for the speaker's subsequent description. Moreover, if the difference between interlocutors' perspectives is at the conceptual level (e.g., whether they know about the existence of a secret shape or not), then it becomes even harder to know where optimal design effects are arising – at the level of grammatical encoding, or higher up the production process, during conceptualisation and message generation.

Another way to find out whether word order and syntactic choices are genuine listener adaptations would be to investigate whether speakers are more likely to produce helpful utterances when there is a communicative reason to do so. For example, we could compare situations where the Matcher is selecting cards from a box (as in Experiments 2, 3 and 5) with situations where the Matcher simply has to verify whether or not the speaker's description matches a picture presented to them in isolation. On the assumption that different word orders have no incremental effect on Matchers' ability to carry out the verification task, then there is no communicative "need" for the speaker to attend to the addressee's perspective. If speakers are more likely to produce helpful word orders or syntactic structures in the "box" task than in the verification task, then we can assume that grammatical choices are (at least partly) driven by a consideration of what would make the interaction more efficient.

Further research might also refine our understanding of the effect of an interlocutor's behaviour (linguistic or otherwise) on a speaker's language use. For example, from the data presented here, we cannot tell for sure why speakers who interacted with an Unhelpful interlocutor in Experiment 5 behaved unhelpfully themselves. They may have been following some kind of conscious "tit-for-tat" strategy (e.g., Axelrod, 1984). They might have thought they were producing the kinds of descriptions that they believed their partner preferred. Alternatively they might have been non-consciously primed to behave unhelpfully by the confederate's behaviour (cf. Bargh, Chen, & Burrows, 1996; Chen & Bargh, 1997). The off-line questionnaire data from Experiment 5 are interesting, because participants' responses suggest that they were well aware of what would make for a helpful utterance, even when they tended to produce *unhelpful* word orders. This might be preliminary evidence for a conscious tit-for-tat strategy, but further evidence is needed; for example, post-experiment questionnaires could explicitly ask whether participants did anything to make the game *harder* for their partner.

Attending to the addressee's perspective may be important for successful communication, but it is obviously not an automatic, resource-free process. If it were, all speakers would produce perfectly formed utterances, all of the time; clearly this is not the case. In fact, some speakers show no evidence of engaging in optimal design, at least not in the goal-directed dialogues investigated here. The fact that there was more variability in the participant sample than the item sample in some of the experiments suggests that optimal design may be a dimension along which participants differ. These differences between participants, and the factors underlying them, may be a fruitful area for future research. For example, to what extent do cognitive, social, and personality-based differences between speakers contribute to the likelihood of them attending to the addressee's perspective in dialogue? How do these *individual* differences relate to differences between *pairs* of interlocutors, and the nature of their conversational exchanges? For example, if a speaker likes his interlocutor and feels motivated to behave in a linguistically helpful way towards her in a spatial perspective-taking task, does this override the fact that he has lower than average spatial ability, and would not normally try to tailor his spatial descriptions to an addressee's perspective (Schober & Brennan, 2003)? The questionnaire data from Experiment 6 suggest that an interlocutor's language behaviour has important consequences for speakers' affective responses to them, and to the interaction as a whole, so it is important to take such social, interactional and individual factors into account in future dialogue research. It may also be important to approach such questions using within-participants designs, in order to minimise variation between experimental conditions (see 5.11.3).

Experiment 6 presents some interesting preliminary evidence that speakers can avoid temporary syntactic ambiguities where it would be helpful to do so. There are good reasons to be cautious in interpreting this finding. But if the finding replicates, this would present a serious challenge to the idea that the production system is fundamentally insensitive to grammatical ambiguities.

Finally, the experiments presented here are, by necessity, agnostic with respect to the question of how early (or how late) optimal design has an effect on production (Horton & Gerrig, 2002). It might guide utterance planning from the earliest moments of processing, or it may only be used later, as a checking mechanism that monitors for violations of common ground (Horton & Keysar, 1996). Since these experiments seek evidence for optimal design in the final product – an overt utterance – we cannot draw any inferences about how quickly a model of the listener was brought to bear. Informal observations suggest that optimal design did not always guide initial planning in these experiments. For example, speakers sometimes overtly corrected themselves after they started to produce an unhelpful description (see Chapter 4). Presumably, the refinement of research paradigms, and the increased use of technologies (such as eye-tracking) that allow us to look for evidence of optimal design on a moment-by-moment basis will help to understand optimal design in language production at a finer grain than the experiments reported here (Schober & Brennan, 2003). To date, such techniques have only been applied to understanding the role of optimal design in *comprehension* (e.g., Barr & Keysar, 2002; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Keysar et al., 1998; 2000; 2003; Metzing & Brennan, 2003), but production studies cannot be far behind.

This thesis set out to examine the extent to which language production in dialogue displays optimal design. Six experiments showed that speakers are sensitive to their addressees' needs in goal-oriented tasks. They adapt word order, syntactic form, and the content of referring expressions to the addressee's perspective if it is obvious how (and why) they should do so, or if they can make some association between different kinds of context and alternative linguistic forms. Speakers must be motivated to do so, however, perhaps by the interlocutor's language behaviour. Speakers' sensitivity to the addressee's perspective co-exists with a sensitivity to the linguistic structure of recently processed utterances: speakers repeat syntax that they have just heard. The results suggest that optimal design is a complex process which involves a trade-off between what is easy for the speaker to produce (e.g., via priming) and what will be easily understood by the addressee; good communication occurs when this trade-off is successfully acheived.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 714-735.

Altmann, G. T .M. (1997). *The Ascent of Babel: An exploration of language, mind, and understanding.* Oxford: Oxford University Press.

Anderson, A., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S. D., Kowto, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The H.C.R.C. Map Task Corpus. *Language and Speech, 34*, 351-366.

Anderson, A. H., & Boyle, E. (1994). Forms of introduction in dialogues: Their discourse contexts and communicative consequences. *Language and Cognitive Processes, 9*, 101-122.

Ariel, M. (1990). *Accessing noun-phrase antecedents.* London: Croom Helm.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language, 51*, 55-70.

Austin, J. L. (1962). *How to do things with words* (2nd Ed.). Cambridge, MA: Harvard University Press.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Bard, E. G., & Aylett, M. P. (in press). Referential form, word duration, and modeling the listener in spoken dialogue. In J. Trueswell & M. Tanenhaus (Eds.), *Processing world-situated language: Bridging the language-as-product and language-as-action traditions.* Cambridge, MA: MIT Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230-244.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391-418.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941-952.

Beach, C. M. (1991). The interpretation of prosodic patterns points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644-663.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*, 145-204.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp.279-362). New York: Wiley.

Blakar, R. (1984). *Communication: A social perspective on clinical issues.* Oslo, Norway: Universitetsforlaget.

Bly, B. (1993). *Uncooperative language and the negotiation of meaning.* Unpublished doctoral dissertation, Stanford University.

Bock, J. K. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1-47.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*, 119-137.

Bock, J. K. (1989). Closed-class immanence in sentence production. *Cognition, 31*, 163-186.

Bock, J. K. (1995). Sentence production: From mind to mouth. In J.L. Miller & P.D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181-216). San Diego: Academic Press.

Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177-192.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information accessibility in sentence production. *Journal of Verbal Learning and Verbal Behavior, 19*, 467-484.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945-984). New York: Academic Press.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1-39.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150-171.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47-67.

Boland, J. E., & Bohem-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language, 39*, 684-719.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*, 119-147.

Boyle, E., Anderson, A., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech, 37*, 1-20.

Branigan, H. P. (in press). Full alignment at some but not all representations in dialogue. A reply to Pickering & Garrod (in press). To appear in *Behavioral and Brain Sciences.*

Branigan, H. P., & Feleki, E., (1999). Conceptual accessibility and serial order in Greek speech production. *Paper presented at The 21st Annual Conference of the Cognitive Science Society, Vancouver, July 1999.*

Branigan, H. P., McLean, J. F., & Reeve, H. (2003). Something Old, Something New: Addressee knowledge and the Given-New contract. *Paper presented at the 25th Annual Meeting of the Cognitive Science Society, Boston, MA, 31 July – 2 August 2003.*

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review, 6*, 635-640.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75*, B13-B25.

Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research, 24*, 489-506.

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding.* Unpublished doctoral dissertation, Stanford University.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1482-1493.

Brennan, S. E., & Metzing, C. A. (in press). Two steps forward, one step back: Partner-specific effects in a psychology of dialogue. A reply to Pickering & Garrod (in press). To appear in *Behavioral and Brain Sciences.*

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. In *Proceedings of the International Joint Conference on Work Activities, Coordination and Collaboration (WACC-99)*, pp.227-235. San Francisco, CA.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251-283.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*, 441-472.

Carletta, J., & Mellish, C. (1996). Risk-taking and recovery in task-oriented dialogue. *Journal of Pragmatics, 26*, 71-107.

Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45*, 58-81.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K., Filip, H., & Carlson, G. (2002). Circumscribing referential domains in real time language comprehension. *Journal of Memory and Language, 47*, 30-49.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research, 29*, 217-229.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*, 541-560.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368-407.

Cinan, S. (2003). Executive processing in free recall of categorized lists. *Learning and Motivation, 34*, 240-261.

Clark, H. H. (1992). *Arenas of Language Use*. London: The University of Chicago Press Ltd.

Clark, H. H. (2002). Speaking in time. *Speech Communication, 36*, 5-13.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and Performance IX* (pp. 313-330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*, 332-373.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension*. Norwood, NJ: Ablex.

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

Clark, H. H., & Murphy, G. L. (1983). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Language and Comprehension* (pp. 287-299). Amsterdam: North-Holland Publishing Company.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language, 26*, 209-225.

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

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22*, 245-258.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language, 49*, 214-230.

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. *Psychonomic Bulletin and Review, 9*, 126-131.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The Effects of Divided Attention on Encoding and Retrieval Processes in Human Memory, *Journal of Experimental Psychology: General, 125*, 159-180.

Cupples, L. (2002). The structural characteristics and on-line comprehension of experiencer-verb sentences. *Language and Cognitive Processes, 17*, 125-162.

Dale, R., & Reiter, E. (1995). Computation interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science, 19*, 233-263.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour, 19*, 450-466.

Danet, B. (1980). 'Baby' or 'fetus'?: Language and the construction of reality in a manslaughter trial. *Semiotica, 32*, 187-219.

Danks, J. H., & Glucksberg, S. (1971). Psychological scaling of adjective orders. *Journal of Verbal Learning and Verbal Behavior, 10*, 63-67.

Danks, J. H., & Schwenk, M. A. (1972). Prenominal adjective order and communication context. *Journal of Verbal Learning and Verbal Behavior, 11*, 183-187.

Dell, G. S. (1986). A spreading activation theory of lexical access in speech production. *Psychological Review, 83*, 283-321.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the 'model of the listener'. In D. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105-129). New York: Academic Press.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11*, 159-184.

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied, 3*, 105-125.

Donaldson, M. (1978). *Children's minds*. London: Fontana Press.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409-436.

Eikmeyer, H. J., & Ahlsen, E. (1998). The cognitive process of referring to an object: A comparative study of German and Swedish. In T. Haukioja (Ed.), *Papers from the 16th Scandinavian Conference of Linguistics* (Publications of the Department of Finnish and General Linguistics of the University of Turku, Vol. 60, pp. 75-86).

Epley, N., & Gilovich, T. (1999). Just going along: Nonconscious priming and conformity to social pressure. *Journal of Experimental Social Psychology, 35*, 578-589.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*, 104-113.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164-203.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language, 48*, 379-398.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725-745.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296-340.

Ferreira, V. S., Slevc, R., & Rogers, E. S. (2003). How do speakers avoid linguistic ambiguity? *Paper presented at Architectures and Mechanisms for Language Processing-2003, Glasgow, Scotland, 25-27 August 2003.*

Ford, W., & Olson, D. R. (1975). The elaboration of the noun phrase in children's description of objects. *Journal of Experimental Child Psychology, 19*, 371-382.

Fowler, C., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perceptions and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology, 19*, 509-526.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378-391.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58-93.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp.133-177). New York: Academic Press.

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol.1, pp. 177-220). New York: Academic Press.

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

Garrod, S., & Clark, A. (1993). The development of dialogue coordination skills in schoolchildren. *Language and Cognitive Processes, 8*, 101-126.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53*, 181-215.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation.* Cambridge: Cambridge University Press.

Glucksberg, S., Krauss, R. M., & Higgins, E. T. (1975). The development of referential communication skills. In F. E. Horowitz (Ed.), *Review of child development research* (Vol. 4, pp.305-345). Chicago: University of Chicago Press.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology, 3*, 333-342.

Goffman, E. (1981). *Forms of talk.* Philadelphia: University of Pennsylvania Press.

Gorrell, P. (1995). *Syntax and parsing.* Cambridge: Cambridge University Press.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161-178.

Gregory, M. F., Healy, A. F., & Jurafsky, D. (submitted). *Common ground in production: Effects of mutual knowledge on word duration.* Unpublished manuscript.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp.225-242). New York: Academic Press.

Grodner, D., & Sedivy, J. C. (2003). The effect of speaker-specific information in on-line pragmatic inferencing. *Poster presented at the 25th Annual meeting of the Cognitive Science Society, Boston, MA, 31 July – 2 August, 2003.*

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science, 28*, 105-115.

Hanna, J. E., & Tanenhaus, M., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43-61.

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic facilitation in agrammatic sentence production. *Brain and Language, 62*, 221-254.

Hartsuiker, R. J., Kolk, H. H. J., & Huiskamp, P. (1999). Priming word order in sentence production. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52*, 129-147.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*, B27-B39.

Hatch, E. (1983). *Psycholinguistics: A second language perspective.* Rowley, MA: Newbury House.

Hawkins, J. A. (1994). *A performance theory of order and constituency.* Cambridge: Cambridge University Press.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language, 47*, 589-606.

Horton, W. S., & Gerrig, R. J. (submitted). *The impact of memory demands on audience design during language production.* Unpublished manuscript.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of Psycholinguistic Research, 21*, 485-496.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology, 82*, 449-462.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*, 182-195.

Isaacs, E., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*, 26-37.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824-843.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness.* Harvard University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Just, M. A., & Carpenter, P. A. (1993). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kemmerer, D. (2000). Selective impairment of knowledge underlying prenominal adjective order: Evidence for the autonomy of grammatical semantics. *Journal of Neurolinguistics, 13*, 57-82.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*, 201-258.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185-209.

Keysar, B. (1997). Unconfounding common ground. *Discourse Processes, 24*, 253-270.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32-38.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language, 39*, 1-20.

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13*, 207-212.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89,* 25-41.

Kingsbury, D. (1968). *Manipulating the amount of information obtained from a person giving directions.* Unpublished honors thesis, Harvard University.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 4,* 391-405.

Kraljic, T., & Brennan, S. E. (2003). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Paper presented at the 16th Annual C.U.N.Y. Conference on Human Sentence Processing, Cambridge, MA, 27-29 March 2003.*

Krauss, R. M., & Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America, 41,* 286-266.

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition, 9,* 2-24.

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 655-701). New York: Guilford.

Krauss, R. M., Garlock, C. M., Bricker, P. D., & McMahon, L. E. (1977). The role of audible and visible back-channel responses in interpersonal communication. *Journal of Personality and Social Psychology, 35,* 523-529.

Krauss, P. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development, 40,* 255-266.

Krauss, P. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American, 236,* 100-105.

Krauss, R. M., Vivekananthan, P. S., & Weinheimer, S. (1968). "Inner speech" and "external speech": Characteristics and communication effectiveness of socially and nonsocially encoded messages. *Journal of Personality and Social Psychology, 9,* 295-300.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1,* 113-114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4,* 343-346.

Krauss, R. M., & Weinheimer, S. (1967). Effect of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior, 6,* 359-363.

Kraut, R. E., Lewis, S. H., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology, 43,* 718-731.

Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. In *Proceedings of the ACM Conference on Computer Supported Collaborative Work.* Cambridge, MA: ACM.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology, 99*, 199–208.

Lakoff, R. (1974). Remarks on *this* and *that. Papers from the Chicago Linguistic Society, 10*, 345-356.

Lantz, D., & Stefflre, V. (1964). Language and cognition revisited. *Journal of Abnormal and Social Psychology, 69*, 472-481.

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*, 2018-2024.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*, 78-106.

Lewis, D. K. (1969). *Convention.* Cambridge, MA: Harvard University Press.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Lloyd, P. (1991). Strategies used to communicate route directions by telephone: A comparison of the performance of 7-year olds, 10-year olds, and adults. *Journal of Child Language, 18*, 175-189.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*, 550-557.

Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*, 177-193.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*, 692-715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157-201.

Maclay, H. S. (1962). A descriptive approach to communication. In N. F. Washburne (Ed.), *Decisions, values and groups* (pp. 201-226). New York: Pergamon Press.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7*, 181-191.

Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition.* Cambridge, New York: Cambridge University press.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*, 331-354.

McAllister, J., Potts, A., Mason, K., & Marchant, G. (1994). Word duration in monologue and dialogue speech. *Language and Speech, 37*, 393-405.

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological and metrical determinants of serial position. *Cognitive Psychology, 25*, 188-230.

Metzing, C. A., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*, 201-213.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*, 477-496.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329-336.

Naveh-Benjamin, M., Craik, F. I. M., Guez, J. & Dori, H. (1998). Effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1091–1104.

Naveh-Benjamin, M., & Guez, J. (2000). Effects of divided attention on encoding and retrieval processes: Assessment of attentional costs and a componential analysis, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1461-1482.

Newport, E., Gleitman, H., & Gleitman, E. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow and C. A. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition.* New York: Cambridge University Press.

Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologia, 64*, 245-259.

Noordman, L., Maes, F., & Arts, A. (2003). Overspecified descriptions in instructive communication. *Paper presented at the 13th Conference of the European Society of Cognitive Psychology, Granada, Spain, 17-20 September 2003.*

Oller, J. W., & Sales, B. D. (1969). Conceptual restrictions on English: A psycholinguistic study. *Lingua, 23*, 209-232.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257-273.

O'Neill, D. K. (1996). Two year-old children's sensitivity to parent's knowledge state when making requests. *Child Development, 67*, 659-677.

Osgood, C. E. (1971). Where do sentences come from? In D.D. Steinberg & L.A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology* (pp. 497-529). London: Cambridge University Press.

Osgood, C. E., & Sebok, T. A. (Eds.) (1954). *Psycholinguistics: A survey of theory and research problems.* Bloomington, IA: Indiana University.

Pasupathi, M., Stallworth, L. M., & Murdoch, K. (1998). How what we tell becomes what we know: Listener effects on speakers' long-term memory for events. *Discourse Processes, 26*, 1-25.

Pechmann, T. (1984). *Uberspezifikation und Betonung in referentieller Kommunikation* [Overspecification and stress in referential communication]. Unpublished doctoral dissertation: Mannheim University, Germany.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89-110.

Pechmann, T. (1994). *Sprachproduktion: Zur Generierung komplexer Nominalphrasen* [Speech production: On generating complex noun phrases]. Opladen: Westdeutscher verlag.

Pechmann, T., & Zerbst, D. (2002). The activation of word class information during speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 233-243.

Perner, J., & Leekham, S. (1986). Belief and quantity: Three-year olds' adaptation to listener's knowledge. *Journal of Child Language, 13*, 305-315.

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A.P. (in press). The personal norm of reciprocity. To appear in *European Journal of Personality*.

Piaget, J. (1926). *The Language and Thought of the Child.* London: Routledge and Kegan Paul.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633-651.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*, 136-141.

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language, 46*, 586-605.

Pickering, M. J., & Garrod, S. (in press). Toward a mechanistic psychology of dialogue. To appear in *Behavioral and Brain Sciences.*

Pinker, S. (1994). *The Language Instinct.* London: Penguin.

Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition, 66*, 183-189.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265-282.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language, 42*, 168-182.

Prentice, J. L. (1967). Effects of cueing actor vs. cueing object on word order in sentence production. *Psychonomic Science, 8*, 163-164.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956-2970.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics*. New York: Academic Press.

Pritchett, B. L. (1992). *Grammatical Competence and Parsing Performance*. Chicago, IL: University of Chicago Press.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28*, 363-392.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes, 11*, 165-192.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374.

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

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*, 175-194.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language.* Wiley.

Schafer, A. J., Speer, S., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*, 169-182.

Schenkein, J. (1980). A taxonomy of repeating action sequences in natural conversation. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 21-47). Academic Press.

Schiffer, S. (1972). *Meaning.* Oxford, England: Clarendon Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1-24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minmized in conversations about locations? *Discourse Processes, 20*, 219-247.

Schober, M. F. (1998). Different kinds of conversational perspective-taking. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and Cognitive Psychological Approaches to Interpersonal Communication*, pp. 145-174. Mahwah, N.J.: Lawrence Erlbaum.

Schober, M. F. (in press). Just how aligned are interlocutors' representations? A reply to Pickering and Garrod (in press). To appear in *Behavioral and Brain Sciences.*

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of Discourse Processes* (pp. 123-164). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schober, M. F., & Carstenson, L. L. (2001). *Do age and relationship matter in conversations about unfamiliar things?* Unpublished manuscript.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

Schriefers, H., & Pechmann, T. (1988). Incremental production of referential noun-phrases by human speakers. In M. Zock & G. Sabah (Eds.), *Advances in Natural Language Generation* (pp.172-179). London: Pinter.

Searle, J. R. (1969). *Speech acts. An essay in the philosophy of language.* Cambridge: Cambridge University Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation: Evidence from the processing of adjectives. *Cognition, 71*, 109-147.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of the listener. *Monographs of the Society for Research in Child Development, 38*, 1-37.

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production: An online study. *Cognition, 78*, 123-164.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103-130.

Snow, C. E. (1972). Mother's speech research: From input to interaction. In C.E. Snow & C.A. Ferguson (Eds.), *Talking to children: Language input and acquisition.* New York: Cambridge University Press.

Sonnenschein, S. (1982). The effects of redundant communications on listeners: When more is less. *Child Development, 53*, 717-729.

Sonnenschein, S. (1984). The effects of redundant communication on listeners: Why different types may have different effects. *Journal of Psycholinguistic Research, 14*, 147-166.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55*, 1936-1945.

Sridhar, S. N. (1988). *Cognition and Sentence Production: A cross-linguistic study.* New York: Springer-Verlag.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics.* Academic Press.

Steedman, M. (1989). Grammar, interpretation, and processing from the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process.* Cambridge, MA: MIT Press.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40*, 136-150.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eyetracking. *Language & Cognitive Processes, 11*, 583-588.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception*. San Diego, CA: Academic Press.

Tannen, D. (1987). Repetition in conversation: Toward a poetics of talk. *Language, 63*, 574–605.

Tannen, D. (1989). *Talking voices: Repetition, dialogue and imagery in conversational discourse.* Cambridge: Cambridge University Press.

Tatar, D. (1997). *Social and personal consequences of a preoccupied listener.* Unpublished doctoral dissertation, Stanford University.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.

Trueswell, J. C., Sekerina, I., & Hill, N. M., & Logrip, M. L. (1999). The kindergarten path effect: Studying on-line processing in young children. *Cognition, 73*, 89-134.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 528-553.

Vallduvi, E. (1992). *The informational component.* New York: Garland.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension.* Academic Press.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes, 10*, 457-486.

Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics, 19*, 29-58.

Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation.* Unpublished doctoral dissertation, Stanford University.

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

Yaguchi, M. (2001). The function of non-deictic *that* in English. *Journal of Pragmatics, 33*, 1125-1155.

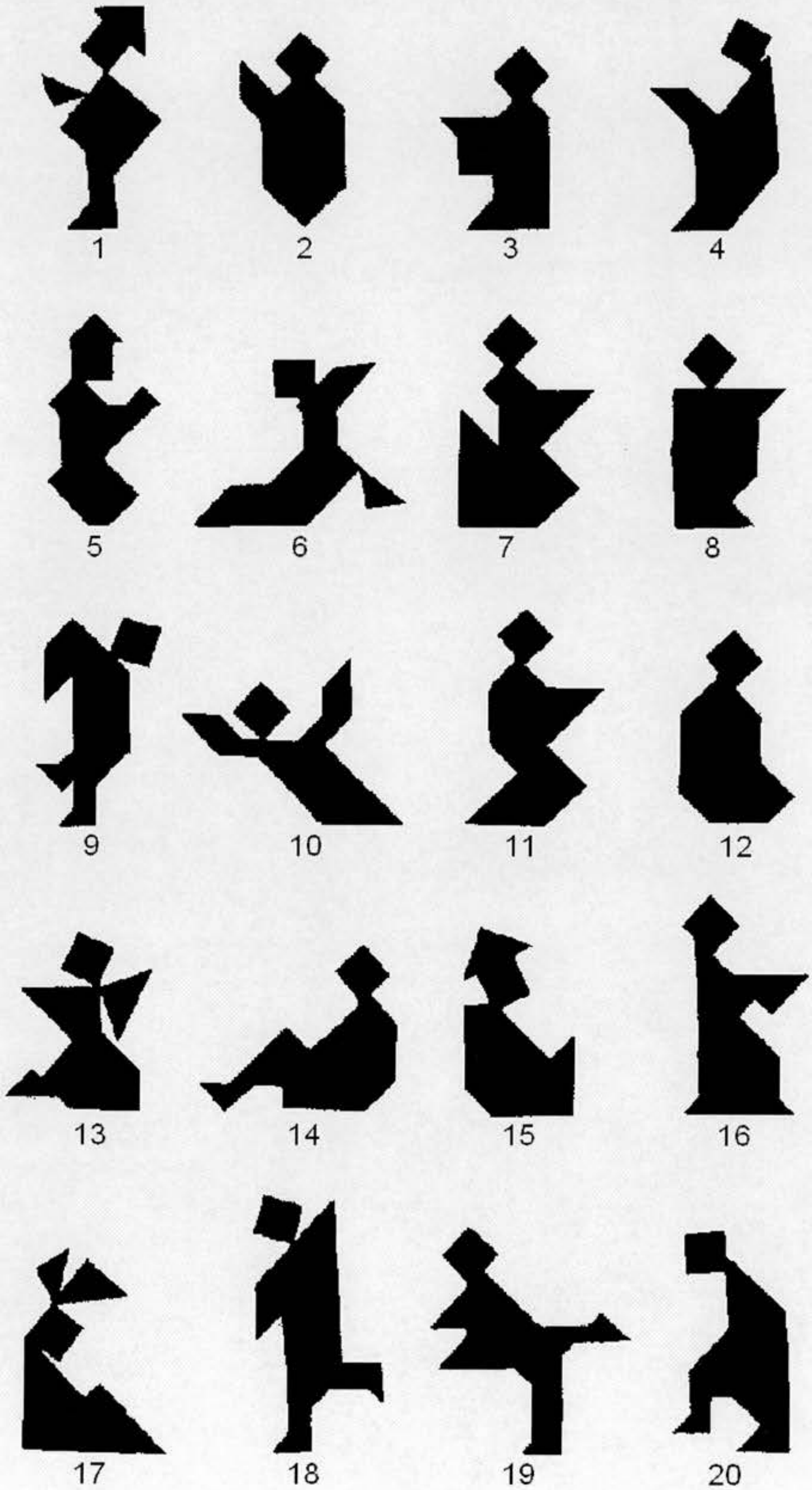
Yngve, V. H. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* (pp.567-578). Chicago: Chicago Linguistic Society.

Yule, G. (1997). *Referential Communication Tasks* (Second Language Acquisition Research Series: Theoretical and Methodological Issues). Mahwah, N.J.: Lawrence Erlbaum Associates.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.

APPENDIX A

Experiment 1: Tangram Figures



APPENDIX B

Experiments 2—6: Experimental Items

B1: Experimental items for Experiments 2 & 3

For each pair of participants, the following list of 64 items was randomised, and the top 48 were used to make the Director's instructions:

1. Stripy orange square
2. Stripy orange circle
3. Stripy orange star
4. Stripy orange triangle
5. Chequered orange square
6. Chequered orange circle
7. Chequered orange star
8. Chequered orange triangle
9. Dotty orange square
10. Dotty orange circle
11. Dotty orange star
12. Dotty orange triangle
13. Wavy orange square
14. Wavy orange circle
15. Wavy orange star
16. Wavy orange triangle
17. Stripy turquoise square
18. Stripy turquoise circle
19. Stripy turquoise star
20. Stripy turquoise triangle
21. Chequered turquoise square
22. Chequered turquoise circle
23. Chequered turquoise star
24. Chequered turquoise triangle
25. Dotty turquoise square
26. Dotty turquoise circle
27. Dotty turquoise star
28. Dotty turquoise triangle
29. Wavy turquoise square
30. Wavy turquoise circle
31. Wavy turquoise star
32. Wavy turquoise triangle
33. Stripy yellow square
34. Stripy yellow circle
35. Stripy yellow star
36. Stripy yellow triangle
37. Chequered yellow square
38. Chequered yellow circle
39. Chequered yellow star
40. Chequered yellow triangle
41. Dotty yellow square
42. Dotty yellow circle
43. Dotty yellow star
44. Dotty yellow triangle
45. Wavy yellow square
46. Wavy yellow circle
47. Wavy yellow star
48. Wavy yellow triangle
49. Stripy purple square
50. Stripy purple circle
51. Stripy purple star
52. Stripy purple triangle
53. Chequered purple square
54. Chequered purple circle
55. Chequered purple star
56. Chequered purple triangle
57. Dotty purple square
58. Dotty purple circle
59. Dotty purple star
60. Dotty purple triangle
61. Wavy purple square
62. Wavy purple circle
63. Wavy purple star
64. Wavy purple triangle

B2: Prime-target pairs for Experiment 4

CONFEDERATE'S PRIME NP	PARTICIPANT'S TARGET CARD
1. The square that's blue	RED CIRCLE
2. The blue triangle	GREEN BRIDGE
3. The green circle	YELLOW BRIDGE
4. The triangle that's green	RED SQUARE
5. The square that's green	BLUE BRIDGE
6. The yellow circle	GREEN SQUARE
7. The blue bridge	YELLOW SQUARE
8. The red triangle	BLUE CIRCLE
9. The bridge that's red	YELLOW TRIANGLE
10. The circle that's blue	RED TRIANGLE
11. The yellow square	BLUE TRIANGLE
12. The triangle that's yellow	RED BRIDGE
13. The red square	GREEN TRIANGLE
14. The circle that's red	BLUE SQUARE
15. The bridge that's yellow	GREEN CIRCLE
16. The green bridge	YELLOW CIRCLE
17. The bridge that's red	BLUE TRIANGLE
18. The blue bridge	RED SQUARE
19. The red square	YELLOW BRIDGE
20. The circle that's red	GREEN SQUARE
21. The triangle that's red	BLUE CIRCLE
22. The yellow square	RED TRIANGLE
23. The triangle that's yellow	BLUE BRIDGE
24. The blue triangle	YELLOW SQUARE
25. The bridge that's yellow	GRREN TRIANGLE
26. The green bridge	RED CIRCLE
27. The square that's blue	GREEN CIRCLE
28. The circle that's blue	YELLOW TRIANGLE
29. The green circle	RED BRIDGE
30. The triangle that's green	BLUE SQUARE
31. The yellow circle	GREEN BRIDGE
32. The square that's green	YELLOW CIRCLE

B3: Prime-target pairs for Experiment 5

Confederate's prime NP (Helpful condition)
Confederate's prime NP (Unhelpful condition)
PARTICIPANT'S TARGET CARD

1. wavy purple circle / yellow stripy triangle / yellow wavy star / dotty orange triangle
   purple wavy circle / stripy yellow triangle / wavy yellow star / orange dotty triangle
   BLUE CHEQUERED TRIANGLE / STAR / SQUARE / CIRCLE
2. chequered orange square / orange chequered star / wavy yellow triangle / yellow stripy star
   chequered orange square / orange chequered star / yellow wavy triangle / stripy yellow star
   BLUE DOTTY STAR / CIRCLE / SQUARE / TRIANGLE
3. yellow dotty square / wavy yellow square / orange dotty circle / chequered purple star
   dotty yellow square / yellow wavy square / dotty orange circle / purple chequered star
   BLUE STRIPY STAR / CIRCLE / SQUARE / TRIANGLE
4. orange stripy circle / stripy purple circle / dotty purple triangle / purple chequered star
   stripy orange circle / purple stripy circle / purple dotty triangle / chequered purple star
   BLUE WAVY TRIANGLE / STAR / SQUARE / CIRCLE
5. purple stripy square / wavy blue triangle / yellow dotty triangle / stripy blue triangle
   stripy purple square / blue wavy triangle / dotty yellow triangle / blue stripy triangle
   ORANGE CHEQUERED TRIANGLE / STAR / CIRCLE / SQUARE
6. blue wavy circle / purple chequered circle / chequered blue star / stripy yellow square
   wavy blue circle / chequered purple circle / blue chequered star / yellow stripy square
   ORANGE DOTTY SQUARE / STAR / CIRCLE / TRIANGLE
7. purple wavy star / wavy purple square / dotty yellow circle / blue chequered triangle
   wavy purple star / purple wavy square / yellow dotty circle / chequered blue triangle
   ORANGE STRIPY TRIANGLE / STAR / SQUARE / CIRCLE
8. yellow chequered circle / stripy purple star / dotty yellow star / purple dotty square
   chequered yellow circle / purple stripy star / yellow dotty star / dotty purple square
   ORANGE WAVY SQUARE / TRIANGLE / CIRCLE / STAR
9. wavy blue star / blue dotty triangle / stripy orange triangle / orange dotty star
   blue wavy star / dotty blue triangle / orange stripy triangle / dotty orange star
   PURPLE CHEQUERED TRIANGLE / STAR / CIRCLE / SQUARE
10. orange wavy triangle / chequered yellow star / chequered blue circle / orange stripy square
    wavy orange triangle / yellow chequered star / blue chequered circle / stripy orange square
    PURPLE DOTTY / CIRCLE / SQUARE / STAR / TRIANGLE
11. chequered yellow triangle / yellow wavy circle / dotty blue circle / orange wavy square
    yellow chequered triangle / wavy yellow circle / blue dotty circle / wavy orange square
    PURPLE STRIPY CIRCLE / SQUARE / TRIANGLE / STAR
12. stripy yellow circle / yellow chequered square / blue stripy star / dotty blue square
    yellow stripy circle / chequered yellow square / stripy blue star / blue dotty square
    PURPLE WAVY SQUARE / TRIANGLE / CIRCLE / STAR
13. dotty orange square / blue wavy square / blue stripy circle / wavy orange star
    orange dotty square / wavy blue square / stripy blue circle / orange wavy star
    YELLOW CHEQUERED STAR / CIRCLE / TRIANGLE / SQUARE
14. purple wavy triangle / orange chequered triangle / stripy orange star / stripy blue square
    wavy purple triangle / chequered orange triangle / orange stripy star / blue stripy square
    YELLOW DOTTY SQUARE / STAR / TRIANGLE / CIRCLE
15. dotty purple star / blue dotty star / blue chequered square / wavy orange circle
    purple dotty star / dotty blue star / chequered blue square / orange wavy circle

YELLOW STRIPY SQUARE / TRIANGLE / CIRCLE / STAR

16. chequered purple triangle / chequered orange circle / purple stripy triangle / purple dotty circle
    purple chequered triangle / orange chequered circle / stripy purple triangle / dotty purple circle
    YELLOW WAVY CIRCLE / TRIANGLE / STAR / SQUARE

B4: Experimental prime-target pairs for Experiment 6

Confederate's prime instruction (Helpful condition)
Confederate's prime instruction (Unhelpful condition)
PARTICIPANT'S TARGET INSTRUCTION

1. Put the tea-bag in the jar on the triangle
   Put the tea-bag that's in the jar on the triangle
   KEY ON BREAD ON CIRCLE
2. Put the pig that's in the cup on the square
   Put the pig in the cup on the square
   PENGUIN ON BLOCK ON STAR
3. Put the key in the cup on the moon
   Put the key that's in the cup on the moon
   TEA-BAG ON BLOCK ON ARROW
4. Put the penguin that's in the jar on the cross
   Put the penguin in the jar on the cross
   PIG ON BREAD ON HEART
5. Put the tea-bag that's in the cup on the star
   Put the tea-bag in the cup on the star
   KEY ON BLOCK ON TRIANGLE
6. Put the pig in the jar on the circle
   Put the pig that's in the jar on the circle
   PENGUIN ON BREAD ON SQUARE
7. Put the key that's in the jar on the arrow
   Put the key in the jar on the arrow
   TEA-BAG ON BLOCK ON CROSS
8. Put the penguin in the cup on the heart
   Put the penguin that's in the cup on the heart
   PIG ON BREAD ON MOON
9. Put the tea-bag in the cup on the square
   Put the tea-bag that's in the cup on the square
   KEY ON BLOCK ON STAR
10. Put the pig that's in the jar on the triangle
    Put the pig in the jar on the triangle
    PENGUIN ON BREAD ON CIRCLE
11. Put the penguin that's in the cup on the moon
    Put the penguin in the cup on the moon
    PIG ON BLOCK ON ARROW
12. Put the key in the jar on the cross
    Put the key that's in the jar on the cross
    TEA-BAG ON BREAD ON HEART
13. Put the tea-bag that's on the block on the circle
    Put the tea-bag on the block on the circle

KEY IN JAR ON SQUARE

14. Put the pig on the bread on the star
    Put the pig that's on the bread on the star
    PENGUIN INCUP ON TRIANGLE
15. Put the penguin on the block on the arrow
    Put the penguin that's on the block on the arrow
    TEA-BAG IN CUP ON MOON
16. Put the key that's on the bread on the heart
    Put the key on the bread on the heart
    PIG IN JAR ON CROSS
17. Put the tea-bag on the bread on the triangle
    Put the tea-bag that's on the bread on the triangle
    PENGUIN INCUP ON STAR
18. Put the pig that's on the block on the square
    Put the pig on the block on the square
    KEY IN JAR ON CIRCLE
19. Put the key on the block on the moon
    Put the key that's on the block on the moon
    TEA-BAG IN JAR ON ARROW
20. Put the penguin that's on the bread on the cross
    Put the penguin on the bread on the cross
    PIG IN CUP ON HEART
21. Put the tea-bag that's on the bread on the star
    Put the tea-bag on the bread on the star
    KEY IN CUP ON TRIANGLE
22. Put the pig on the block on the circle
    Put the pig that's on the block on the circle
    PENGUIN IN JAR ON SQUARE
23. Put the key that's on the block on the arrow
    Put the key on the block on the arrow
    TEA-BAG IN JAR ON CROSS
24. Put the penguin on the bread on the heart
    Put the penguin that's on the bread on the heart
    PIG IN CUP ON MOON

B5: Filler prime-target pairs for Experiment 6

Confederate's prime instruction (Helpful & Unhelpful conditions)
PARTICIPANT'S TARGET INSTRUCTION

1. Put the rabbit in the glass on the moon
   PAPERCIP ON ARROW
2. Put the sharpener on the saucer on the cross
   SHEEP ON HEART
3. Put the sheep in the glass on the triangle
   RABBIT ON STAR
4. Put the paperclip on the square
   SHARPENER ON SAUCER ON CIRCLE
5. Put the sharpener in the glass on the heart

SHEEP ON SAUCER ON MOON

6. Put the rabbit on the arrow
   PAPERCLIP ON CROSS
7. Put the paperclip on the saucer on the circle
   SHARPENER ON TRIANGLE
8. Put the sheep in the glass on the star
   RABBIT ON SQUARE
9. Put the rabbit on the cross
   PAPERCLIP ON HEART
10. Put the sharpener in the glass on the moon
    SHEEP ON SAUCER ON ARROW
11. Put the sheep on the square
    RABBIT ON CIRCLE
12. Put the paperclip in the glass on the triangle
    SHARPENER ON SAUCER ON STAR
13. Put the sharpener on the arrow
    SHEEP IN GLASS ON CROSS
14. Put the rabbit on the saucer on the heart
    PAPERCLIP ON MOON
15. Put the paperclip on the saucer on the star
    SHARPENER IN GLASS ON SQUARE
16. Put the sheep on the circle
    RABBIT ON TRIANGLE
17. Put the rabbit on the saucer on the moon
    PAPERCLIP IN GLASS ON ARROW
18. Put the sharpener on the cross
    SHEEP ON HEART
19. Put the sheep on the triangle
    RABBIT ON STAR
20. Put the paperclip on the saucer on the square
    SHARPENER IN GLASS ON CIRCLE
21. Put the sharpener on the saucer on the heart
    SHEEP ON MOON
22. Put the rabbit on the arrow
    PAPERCLIP IN GLASS ON CROSS
23. Put the paperclip on the circle
    SHARPENER ON SAUCER ON TRIANGLE
24. Put the sheep on the star
    RABBIT IN GLASS ON SQUARE

APPENDIX C

Experiment 3: Post-Experiment Questionnaire

Thanks for taking part in this experiment! Now we'd like you to think about your experience as a participant in this study.

Please use the space below to write down what you think the experiment might be about. We are interested to find out *your thoughts* about the purpose of the experiment, so there are **no wrong answers**! Please have a guess even if you feel you can't say very much about it.

Now you can turn over the page to fill in the rest of the questionnaire.

1. In your opinion, how good were you and your partner at building up the grids together? (circle one)

very good good okay not very good poor

<u>When you were the Director</u>:

2. Do you think you mis-described any of the cards? (circle one) YES NO

3. Did you do anything special to make it easy for your partner to build up the grids? What did you do? Did you do this all the way through the experiment, or just sometimes?

4. Do you think you could have been a more 'helpful' Director? If so, what could you have done to make it easier for the Matcher?

<u>When you were the Matcher</u>:

5. Do you think you made any mistakes putting the cards on the grid? YES NO

6. Did your partner do anything special that made it easy for you to build up the grids? What did they do? Did they do this all the way through the experiment, or just sometimes?

7. Thinking back, was there anything your partner could have done to make your job as Matcher easier?

Thanks again for your help!

APPENDIX D

Experiment 4: Post-Experiment Questionnaire

Thanks for taking part in this experiment! Now we'd like you to think about your experience as a participant in this study.

Please use the space below to write down what you think the experiment might be about. We are interested to find out *your thoughts* about the purpose of the experiment, so there are **no wrong answers**! Please have a guess even if you feel you can't say very much about it.

Now you can turn over the page to fill in the rest of the questionnaire.

1. In your opinion, how good were you and your partner at building up the grids together? (circle one)

very good　　good　　okay　　not very good　　poor

2. Do you think you mis-described any of the cards? (circle one)　YES　　NO

3. Did you do anything special to make it easy for your partner to build up the grids? What did you do? Did you do this all the way through the experiment, or just sometimes?

4. Do you think you could have been more 'helpful' when you described your cards? If so, what could you have done to make it easier for your partner?

5. Do you think you made any mistakes putting the blocks on the grid?　YES　　NO

6. Did your partner do anything special that made it easy for you to build up the grids? What did they do? Did they do this all the way through the experiment, or just sometimes?

7. Thinking back, was there anything your partner could have done to make it easier for you to select blocks and put them in the right places on the grid?

Thanks again for your help!

APPENDIX E

Experiment 5: Post-Experiment Questionnaire

Thanks for taking part in the experiment! Now we'd like you to think about your experience as a participant in this study.

Please use the space below to write down what you think the experiment might be about. We are interested to find out *your thoughts* about the purpose of the experiment, so there are **no wrong answers**! Please have a guess even if you feel you can't say very much about it.

Now you can turn over the page to fill in the rest of the questionnaire.

1. In your opinion, how good were you and your partner at building up the grids together? (circle one)

excellent good okay not very good poor

2. Do you think you mis-described any of the cards? (circle one) YES NO

3. When you were describing the instruction cards, did you do anything special to make the game easy for your partner? What did you do? Did you do this all the way through the experiment, or just sometimes?

4. Do you think you could have been more 'helpful' when you described your instruction cards? If so, what could you have done to make it easier for your partner?

5. Do you think you made any mistakes putting cards on your grid? YES NO

6. When your partner was describing, did they do anything special that made it easy for you to play the game? What did they do? Did they do this all the way through the experiment, or just sometimes?

7. Thinking back, was there anything your partner could have done to make it easier for you to select cards from the red boxes?

9. Have you met your partner in this experiment before? (circle one) YES NO

10. If you have met them before, how well do you know them? (close friend, classmate, acquaintance…) ____________________

11. If you haven't met them before, did they seem to be:

extremely likeable fairly likeable okay fairly dislikeable extremely dislikeable

Thanks again for your help!

APPENDIX F

Experiment 6: Post-Experiment Questionnaire

Thanks for taking part in the experiment! Now we'd like you to think about your experiences as a participant in the study. This questionnaire is completely anonymous. The experimenter will only identify you with a number. Please answer the questions in the order that they appear.

Use the space below to write down what you think the experiment might be about. We are interested in your opinions; there are **no wrong answers**!

Some of the following questions have a rating scale below them, such as:

What is the weather like today?

Very Cloudy • • • • • • • • • • Very Sunny

For these questions, please circle the point on the scale corresponding to your judgement. If, for example, you thought it was a very cloudy day, you would circle the first point. On the other hand, if you thought that it was quite a sunny day, you might circle the seventh or eighth dot.

Now you can turn over the page to fill in the rest of the questionnaire.

1. Please circle the dot that best describes the communication game.

Not Enjoyable	•	•	•	•	•	•	•	•	•	•	Enjoyable
Easy	•	•	•	•	•	•	•	•	•	•	Difficult
Simple	•	•	•	•	•	•	•	•	•	•	Complicated

2. When you were **following your partner's instructions**, was there anything about the game that made it difficult? What was it?

3. Was the game easier on some turns than on others? If so, why?

4. Thinking back, was there anything your partner could have done to make it easier for you to follow their instructions? What was it?

5. When you were **giving instructions**, did you do anything special to make the game easy for your partner? What did you do? Did you do this all the way through the experiment, or just sometimes?

6. Do you think you could have been more 'helpful' in the way that you gave instructions? If so, what could you have done to make the game easier for your partner?

7. How similar or different were the following:

Your instructions and your partner's instructions?

Very Different • • • • • • • • • • Very Similar

Your word choices and your partner's word choices?

Very Different • • • • • • • • • • Very Similar

8. Have you met your partner in this experiment before? (circle one) YES NO

9. If you have met them before, how well do you know them? (close friend, classmate, acquaintance...)

10. Please circle the dot that best describes how your partner played the game.

friendly	• • • • • • • • • •	unfriendly
unintelligent	• • • • • • • • • •	intelligent
inflexible	• • • • • • • • • •	flexible
fast	• • • • • • • • • •	slow
part of a team	• • • • • • • • • •	not part of a team
rigid	• • • • • • • • • •	not rigid
uncooperative	• • • • • • • • • •	cooperative
skilled	• • • • • • • • • •	unskilled

11. Your age: ___________

12. Gender: FEMALE MALE

13. Are you a native speaker of British or American English? YES NO

14. What course are you registered for? ______________________________

Thanks again for your help!