THE UNIVERSITY
of EDINBURGH

An evolutionary psycholinguistic approach to the pragmatics of reference

Rachael Bailes

Thesis submitted for the degree of Doctor of Philosophy

School of Philosophy, Psychology, and Language Sciences

University of Edinburgh

2016

Abstract

Pragmatics concerns the material function of language use in the world, and thus touches on profound questions about the relationship between our cognition and the environments in which we operate. Both psycholinguistics and evolutionary linguistics have afforded greater attention to pragmatics in recent years. Though the potential of evolutionary psycholinguistics has been noted for over twenty-five years (e.g. Tooby & Cosmides, 1990; Scott-Phillips, 2010a), there has arguably been little dialogue between these two fields of study. This thesis explicitly acknowledges and investigates the adaptationist nature of functional claims in psycholinguistics, and attempts to demonstrate that psycholinguistic inquiry can provide evidence that is relevant to theories of how the cognitive architecture of linguistic communication evolved.

Chapter two reviews a broad polarisation in the pragmatic and psycholinguistic literature concerning the relative roles of linguistic convention and contextual information in comprehension. It makes explicit the theoretical approaches that reliably give rise to these polar positions across scholarly domains. It goes on to map each model of comprehension to the adaptationist particulars it may entail, and in doing so illustrates two different pictures of how linguistic cognition has developed over phylogeny. The Social Adaptation Hypothesis (SAH) holds that linguistic comprehension is performed by relevance-oriented inferential mechanisms that have been selected for by a social environment (i.e. inference-using conspecifics). In particular, the SAH holds that linguistic conventions are attended to in the same way as other ostensive stimuli and contextual information, and because of their relevance to communicative interactions. The Linguistic Adaptation Hypothesis (LAH) holds that linguistic comprehension is performed by specialised cognition that has been selected for by a linguistic environment (i.e. language-using conspecifics) that was established subsequent to, and as a consequence of, the emergence of inferential communication. In particular, the LAH

holds that linguistic conventions are a privileged domain of input for the comprehension system. The plausibility and congruence of both accounts with the current state of knowledge about the evolutionary picture necessitates empirical psycholinguistic evidence.

The remainder of the thesis presents a series of experiments investigating referential expressions relevant to the contrastive predictions of these two adaptationist accounts. The broad question that covers all of these experiments is: how sensitive is the comprehension process to linguistic input qua linguistic input, relative to various other grades of relevant contextual information?

Chapter three presents a reaction time experiment that uses speaker-specific facts about referents as referring expressions, in a conversational precedent paradigm. The experiment measures the relative sensitivity of comprehension processing to the knowledge states of speakers and the consistent use of linguistic labels, and finds greater sensitivity to linguistic labels.

Chapter four introduces a further contextual variable into this paradigm, in the form of culturally copresent associations between labels and referents. The experiment presented in this chapter compares the relative sensitivity of processing to culturally copresent common ground, the privileged knowledge state of speakers, and the consistent use of linguistic labels. The results indicated greater sensitivity to linguistic labels overall, and were consistent with the LAH.

Chapter five turns to visual context as a constraint on reference, and presents two pairs of experiments. Experiments 3 and 4 investigate the comprehension of referring expressions across congruous, incongruous, and abstract visual contexts. The experiments measured reaction time as subjects were prompted to identify constituent parts of tangram pictures. The results indicated a sensitivity to the visual context and the linguistic labels, and are broadly consistent with the SAH. If comprehension is characterised by particular sensitivities, we may expect speakers to produce utterances that lend themselves well to how hearers process them. Experiments 5 and 6 use a similar tangram paradigm to elicit referring expressions from speakers for component parts

of tangrams. The experiments measure the consistency of produced labels for the same referents across visual contexts of varied congruity. The results indicated some methodological limitations of the tangram paradigm for the study of repeated reference across contexts.

Lastly, the thesis concludes by considering the SAH and LAH in light of the empirical evidence presented and its accompanying limitations, and argues that the evidence is generally consistent with the assumptions of the LAH.

Acknowledgements

First and foremost, I must thank my excellent supervisors for bearing with me. Thanks to Prof. Holly Branigan for her attention to detail, constructive guidance, patient encouragement, and experiment expertise, all of which I'd have been lost without. Many thanks to Dr Thom Scott-Phillips for his writing advice, supportive counsel, and most of all for his cheerful willingness to discuss and develop ideas. Thanks also to Prof. Simon Kirby for his helpful feedback, and for introducing me to both the topic that led to this thesis and the idea of embarking on a PhD in the first place.

I'm grateful to the University and the School of Philosophy, Psychology and Language Sciences for funding, and for the opportunity to gain so much experience among brilliant people. A big thank you to the members of the Centre for Language Evolution and the inhabitants of DSB 1.15, past and present, for the fine collegiality and baked goods they have supplied me with over the years. I particularly want to thank Ed Baggs, Chrissy Cuskley, Stephanie De Marco, Eddie Dubourg, Mateo Obregón, Sean Roberts, Kenny Smith, Kevin Stadler, and James Thomas for helping me in various ways to make this thesis happen. Special thanks are due to Justin Quillinan for his help in coding experiments 5 and 6, and for moral support throughout the programme. Many thanks also to everyone who participated in my experiments.

Finally, it is with unreserved and profound appreciation that I acknowledge Cathy Gillespie, without whom I would have mentally, physically, and financially perished. Thank you.

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Rachael Bailes

Contents

1 Introduction **1**

1.1 Models of communication 3

1.2 How has the comprehension system been shaped by the linguistic environment? 10

1.3 Aim of the thesis 17

1.3.1 Thesis Road Map 18

2 The psychological architecture of reference in context **21**

2.1 Theoretical approaches to pragmatic competence 21

2.2 Psycholinguistic accounts of reference in context 31

2.3 Adaptationist psycholinguistics of comprehension 44

2.3.1 Psycholinguistic models in an adaptationist framework 47

2.3.2 Hypotheses 56

2.3.3 Summary 59

3 Facts about referents as precedent labels **62**

3.1 Introduction 62

3.2 Motivation 70

3.2.1 Predictions 74

3.3 Experiment 1 76

3.3.1 Participants 76

3.3.2 Design 76

3.3.3 Materials . . . 77
3.3.4 Procedure . . . 78
3.3.5 Dependent variables . . . 81
3.4 Results . . . 82
3.4.1 Reaction Time . . . 82
3.4.2 Object Selection . . . 82
3.5 Discussion . . . 83
3.5.1 Limitations and future directions . . . 87
3.6 Conclusion . . . 91

4 Speaker-specific and cultural associations in reference 93
4.1 Introduction . . . 93
4.2 Motivation . . . 98
4.2.1 Predictions . . . 101
4.3 Experiment 2 . . . 103
4.3.1 Participants . . . 103
4.3.2 Design . . . 104
4.3.3 Materials . . . 104
4.3.4 Procedure . . . 106
4.3.5 Dependent variables . . . 108
4.4 Results . . . 109
4.4.1 Reaction Time . . . 109
4.4.2 Object Selection . . . 110
4.5 Discussion . . . 115
4.5.1 Limitations and future directions . . . 117
4.6 Conclusion . . . 119

5 Reference across visual contexts 122
5.1 Introduction . . . 122
5.2 Comprehension of reference across visual contexts . . . 127

5.3 Experiment 3 132
5.3.1 Participants 132
5.3.2 Design 132
5.3.3 Materials 132
5.3.4 Visual materials: 132
5.3.5 Audio: 133
5.3.6 Procedure 134
5.3.7 Dependent variables 136
5.3.8 Results 136
5.3.9 Results Summary 138
5.4 Experiment 4 140
5.4.1 Participants 140
5.4.2 Design 140
5.4.3 Materials 141
5.4.4 Procedure 141
5.4.5 Dependent variables 144
5.4.6 Results 145
5.4.7 Comparison of experiments 3 and 4 145
5.4.8 Discussion 149
5.5 Production of reference across visual contexts 154
5.5.1 Motivation 154
5.6 Experiment 5 159
5.6.1 Participants 159
5.6.2 Design 159
5.6.3 Materials 159
5.6.4 Procedure 160
5.6.5 Dependent variables 161
5.6.6 Experiment 5 Results 161
5.6.7 Results Summary 163

5.7 Experiment 6 . . . 164
5.7.1 Participants . . . 164
5.7.2 Design . . . 164
5.7.3 Materials . . . 164
5.7.4 Procedure . . . 165
5.7.5 Dependent variables . . . 166
5.8 Experiment 6 Results . . . 167
5.8.1 Produced Labels . . . 167
5.9 Comparison of experiments 5 and 6 . . . 167
5.9.1 Results . . . 169
5.10 Discussion . . . 171
5.10.1 Limitations and future directions . . . 172
5.11 Conclusion . . . 174

6 Summary and conclusions **179**
6.1 An evolutionary psycholinguistic approach to pragmatics . . . 180
6.2 Theoretical synthesis . . . 183
6.3 Empirical evidence . . . 187
6.3.1 Methodological limitations . . . 190
6.3.2 Conclusion . . . 200
6.4 Future directions . . . 201
6.5 Overall conclusions . . . 203

References **205**

Appendices **219**
.1 Paired Tangrams (Experiments 3-6) . . . 220
.2 Paired Tangrams (Experiments 3-6) . . . 221
.3 Paired Tangrams (Experiments 3-6) . . . 222
.4 Paired Tangrams (Experiments 3-6) . . . 223

.5 Paired Tangrams (Experiments 3-6) 224

List of Figures

1.1 Code Model Communication, taken from Scott-Phillips and Kirby (2013) 4

1.2 Two Models of Inferential Communication (subset of Fig. 1.3) 8

1.3 Three Models of Inferential Communication 15

3.1 Example Visual Array used in Experiment 1 77

3.2 Boxplot of Reaction Times by Condition for Experiment 1 83

3.3 Histogram of Response Matches to Precedent Object by Condition for Experiment 1 84

3.4 Density Plot of Reaction Times by Condition for Experiment 1 85

4.1 Example Visual Array for Experiment 2 105

4.2 Boxplot of Reaction Times by Condition for Experiment 2 111

4.3 Histogram for Response Matches to Precedent Object for Experiment 2 113

4.4 Histogram for Response Matches to Cultural Object for Experiment 2 114

5.1 Example Trial from Experiment 3, showing Different Context and Maintained Label 133

5.2 Audio Prompts by Condition for Experiment 3 134

5.3 Boxplot of Reaction Times for Experiment 3 137

5.4 Histogram of Response Matches to Precedent Shape for Experiment 3 139

5.5 Example Trial from Experiment 4, showing Abstract Context and Maintained Label 142

5.6 Boxplot of Reaction Times by Condition for Experiment 4 146

5.7 Histogram of Response Matches to Precedent Shape by Condition for Experiment 4 146
5.8 Boxplot of Reaction Times for direct comparison of Experiments 3 and 4 147
5.9 Histogram of Response Matches to Precedent Shape for direct comparison of Experiments 3 and 4 148
5.10 Histogram of Produced Label Matches by Condition for Experiment 5 162
5.11 Histogram of Produced Label Matches by Condition for Experiment 6 168
5.12 Histogram of Produced Label Matches for direct comparison of Experiments 5 and 6 170

6.1 Research questions in evolutionary linguistics, taken from Scott-Phillips (2010). 181

List of Tables

2.1 Table of Scalar Implicatures, taken from Horn (2006) 30
2.2 Context-driven and Language-driven Comprehension in an Adaptationist Framework 49
3.1 Audio Prompts for each Experimental Condition in Experiment 1 . . . 78
4.1 Table of Domains of Language Use, taken from Barr and Keysar (2005) 97
4.2 Audio Prompts for each Experimental Condition in Experiment 2 . . . 106

Chapter 1

Introduction

"Well done!" is a complete sentence, but the meaning you derive from it may differ according to the situation in which it is uttered. Perhaps you have just asked how the speaker likes their steak; perhaps you just accidentally smashed an expensive ornament of theirs; or perhaps you have just delighted them with a Shakespearean soliloquy. We can go further still: perhaps you have just delivered the soliloquy, but you also hold certain beliefs about the benevolent dishonesty or passive-aggressive nature of the speaker. Unlike the simple communication systems of other animals, succesful language use requires a very complex interplay between linguistic cognition and the outside world. Honeybees use the famous waggle dance to accurately communicate the whereabouts of nectar (Von Frisch, 1967), but it is unlikely that such a broadcast by a grumpy-looking bee would ever be interpreted by the hive as a sarcastic joke or hypothetical conversation starter. Where simple animal communication is limited in function and rigidly codified, the comprehension of language may be conditioned upon an extremely rich array of contextual factors.

The observation that linguistic comprehension makes recourse to extra-linguistic information has led to the key insight that language is facilitated by a process of inference (Grice, 1975). Linguistic utterances can thus be characterised as *evidence for* a speaker's intended meaning, which must be appropriately integrated with numerous other forms of knowledge by the hearer. The inferential features of human commu-

nication may be unique in the natural world (Scott-Phillips, 2015b, 2015a). A small industry of experimental investigation has grown to address psycholinguistic questions about how this feat is achieved at the level of cognitive processing (Noveck & Sperber, 2004). At a more zoomed-out level of analysis, evolutionary linguistics has begun to investigate the role of pragmatic cognition in language evolution (Hoefler, 2009; Hoefler & Smith, 2009). In particular, inferential communication has been proposed as a necessary precursor to the emergence of language in humans (Origgi & Sperber, 2000; Bar-On, 2013; Scott-Phillips, 2014).

While it is generally agreed that linguistic and contextual knowledge are both important components of meaning comprehension, their relative contributions to the process are more contested. This has implications for our understanding of both cognitive processing and the evolutionary picture of linguistic communication. The question of whether language comprehension rests more upon associative linguistic meaning or inferred contextual meaning characterises a long-standing and continuing disagreement in pragmatics (Grice, 1957; Strawson, 1964; Sperber & Wilson, 1986; Horn, 2006a). Over the last thirty years, this division has resurfaced and continues in psycholinguistics, where the development of advanced tools such as eye-tracking has facilitated a large number of empirical investigations into the comprehension of language and dialogue in context (see Trueswell & Tanenhaus, 2005; Barr & Keysar, 2006, for reviews). Most recently, this same debate has become pertinent to the growing field of evolutionary linguistics, with arguments for the evolution of fundamentally context-driven language comprehension in humans (Scott-Phillips, 2014) and, more broadly, an interest in how the pragmatic phenomena particular to human communication relate to more formalist models of language evolution (Kotchoubey, 2005; Hoefler, 2009). Despite this common point of interest in evolutionary and psycho-linguistics, there is notably little dialogue between these domains of study, much less a synthesised approach to the issue.

This thesis advocates and attempts such an approach, and aims to demonstrate that psycholinguistics and evolutionary linguistics can make relevant and interesting contri-

butions to each other in a reciprocal fashion. Psycholinguistic theories about processing may imply certain adaptationist particulars; an evolutionary framework provides the means to make these features explicit, and to assess their coherence with the state of knowledge about how linguistic cognition evolved in humans. On the other side of the coin, claims about the evolution of linguistic communication predict certain features of linguistic processing, and these predictions require empirical, psycholinguistic evidence. In short, psycholinguistic processing is a valuable object of study for evolutionary linguists, and evolutionary theory is a relevant tool for psycholinguists. In the following chapters, I will set out an evolutionary psycholinguistic approach to the current, polarised debate surrounding the pragmatics of linguistic comprehension, formalise the theoretical approaches that give rise to this polarisation across scholarly domains, and contribute to the new and unfolding conversation about inferential communication in evolutionary linguistics.

1.1 Models of communication

There are two broad approaches to thinking about linguistic communication, each with different implications. The first, and most widely established, is commonly termed the *code model* (see Sperber & Wilson, 1986; Blackburn, 1999; Scott-Phillips, 2010b; Scott-Phillips & Kirby, 2013). In broad terms, the code model conceives of language as a set of encoded mappings between signals and meanings. Code model communication takes place when an encoded signal is transmitted across a channel, and 'decoded' by the receiver in order to retrieve the associated meaning. This way of thinking about linguistic communication pervades everyday speech in phrases like "*put* your thoughts *into* words" and "*get* your message *across*" (Reddy, 1979), and is therefore intuitive to most speakers of English. A similar conception, specifically designed to characterise the transmission of digital strings across a noisy channel, is the standard model of communication in information theory (Weaver, 1949) and related disciplines such as artificial intelligence. Diagrams of both schemas of code model communication can

be seen below in Fig. 1.1. The code model is widely used in language evolution research; it is a useful schema for computational simulations that abstract away from the real-world instantiations of linguistic communication in order to address issues like the biological evolution of signalling systems (e.g. MacLennan & Burghardt, 1993; Ackley & Littman, 1994), the cultural transmission and evolution of signal structures (e.g. Batali, 1998; Kirby, 2001; K. Smith, 2004; Brighton, Smith, & Kirby, 2005), and the dynamic relationships between biology and culture (e.g. Oliphant, 1996).

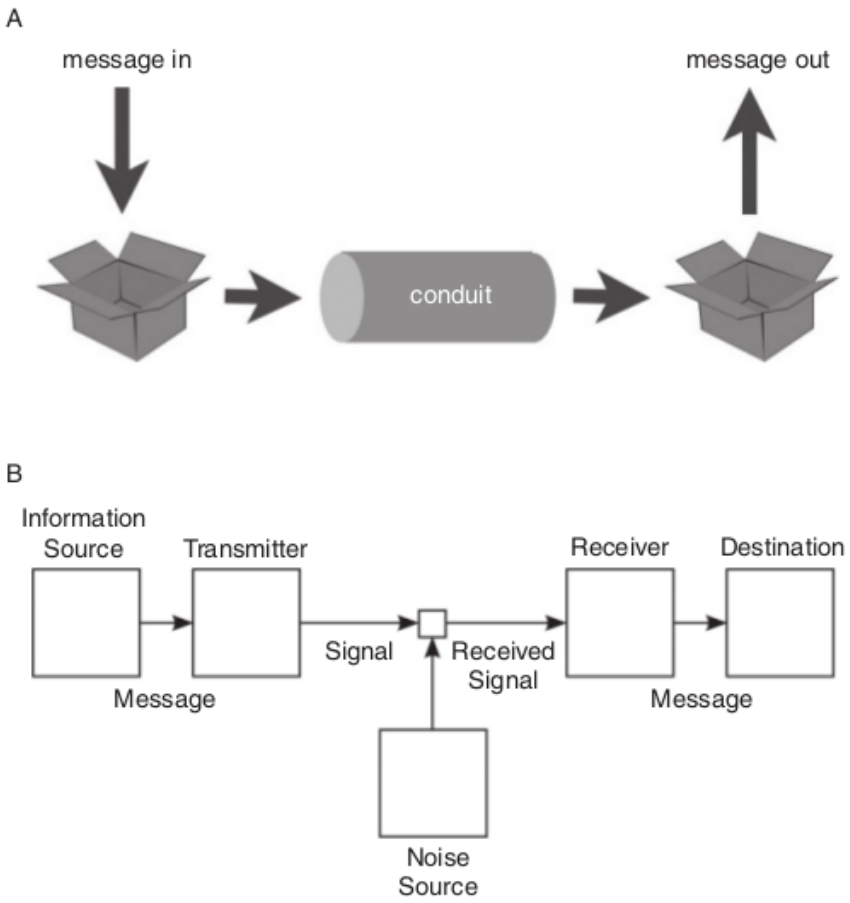


Figure 1.1: Code model communication, taken from Scott-Phillips and Kirby (2013). In A, the commonly used *conduit metaphor* of language, a message is packaged up and sent along a conduit, where it is then received and the message is retrieved at the other end. In B, the *Shannon-Weaver model* of communication, an encoded signal is transmitted across a channel, which may include noise, and the message is extracted from the signal upon reaching its destination.

This conception of communication, where meanings are encoded in signals and unpacked by the receiver, cannot account for the contribution of context to a hearer's interpretation. An alternative model, an *inferential model* of linguistic communication, has been developed and modified by pragmaticists in order to address this. Early conceptions of inferential communication, seeded by the work of Paul Grice (1975) are often characterised in the literature as a 'two step' model (e.g. Recanati, 2004). The two-step model of inference consists of first decoding the linguistic utterance for a linguistic meaning that it represents. Linguistic meaning can be understood as the conventional, associative, literal meanings of words and utterances that constitute a hearer's linguistic knowledge. The second stage involves the additional integration of contextual information, selected *on the basis of* the linguistic meaning that has been derived. In other words, decoding the linguistic utterance triggers appropriate inferences, and the result of this process is comprehension. This two step model of inference is represented by Model A in Fig. 1.3.

The two-step model and the code model encounter significant problems when it comes to the origins of linguistic communication. Specifically, the structure of the conventional linguistic code could not have emerged like natural codes do, without prior inferential abilities (Wharton, 2003; Scott-Phillips, 2014). Put another way: if inference is a second step of comprehension that is designed to deal with the meanings of linguistic signals, then how were those linguistic codes established? In nature, communication systems are established by the gradual coevolution (either through *ritualisation* or *sensory manipulation*) of a signal behaviour and a response (Bradbury & Vehrencamp, 1998; J. M. Smith & Harper, 2003). One example of the ritualisation process is the evolution of territorial urination signals in some mammals: an animal urinates as it approaches the periphery of an area in which it feels safe (a common mammalian fear response), and others are then able to use this as a reliable marker of the animal's territory in order to keep their distance. Given this incidental foothold, the capacity for signals and responses can coevolve around the function of territory marking until they become adapted to one another in a system of communication. The rigid

function that characterises these natural code systems stands in contrast to the combinatorial and richly context-dependent system of human language, with no clear route from the former to the latter. Combinatoriality is a defining feature of human language (Hockett, 1960), but it is rare elsewhere in nature. This is because ritualisation and sensory manipulation processes such as the mammalian territory-marking scenario can, in principle, begin from any behaviour in an animal's repertoire. The development of a combinatorial signal, however, requires that existing signals - that is, *a smaller subset* of the animal's repertoire - are re-used for a different function. They follow a similarly slow process of coevolution, and are simply less likely to occur from the outset. It quickly becomes very unlikely that the processes which give rise to natural codes in nature could be responsible for the flexible and vastly combinatorial inventories of human languages. Once cultural conventions have been established, it is possible that cultural evolutionary processes could give rise to combinatoriality (see Verhoef, 2012); however this raises the question of how conventional codes (rather than natural codes) become established.

There are other problems with a 'code-first' account of language emergence. The arbitrary nature of linguistic conventions obscures any functional foothold from which a natural code process could have arisen. Sensory theories of protolanguage propose that stable, cross-modal regularities in human sensation and perception may have grounded protolinguistic conventions. For example, population-level cross-modal associations between some linguistic sounds and tastes could have been exploited in the establishment of a shared lexicon for foods (Simner, Cuskley, & Kirby, 2010). However, sounds do not indicate tastes in the indexical way that urine indicates the presence of an animal, so it is likely that using shared sensory associations to establish a convention still requires metapsychological inference at the outset. Indeed, the study of protolanguage necessarily assumes some cognitive pre-adaptations for communication (see Scott-Phillips & Kirby, 2010, for a clear delineation of stages in language evolution study). Humans are also capable of using their inferential cognition to communicate without using language at all; if someone tilts a bag of crisps toward their friend,

their friend understands that this signals an offer and responds by taking one. If the signaller then tilts the bag again and shakes it a little, they signal that the friend should take more; the friend may decline by raising a hand. Mundane exchanges like this are not linguistic, but they are nonetheless communicative. If inferential cognition developed after codified language in order to process the linguistic code, exchanges such as these lack explanation; one would first need to explain how conventional codes (not natural codes) were established and used without prior inferential capabilities, explain the means by which inference emerged and became specifically adapted for the use of linguistic conventions (and indeed explain why this was necessary if humans were already able to use these codes), and then identify the means by which this linguistic inference was extended to multiple other domains of use. Such an account does not present itself in the literature at this time. In short, there are no coherent evolutionary accounts of how a linguistic code could have been established before inferential abilities.

One solution is *ostensive-inferential* communication, which has been proposed as a necessary precursor for the emergence of language in humans (Sperber, 2000; Scott-Phillips, 2014). A high-level characterisation is enough to situate this model in the theoretical landscape; technical details will be elaborated further in chapter 2. Ostensive-inferential communication is defined in terms of the expression (ostension) and recognition (inference) of intentions. In order to interpret the *intended meaning* of an utterance, the hearer must refer to a range of relevant contextual information, which includes the knowledge they share with the speaker. In this model, linguistic utterances are one of many sources of evidence for a speaker's intended meaning, all of which are processed by a central inferential system. A schema of the ostensive-inferential model can be seen in Model B of Fig. 1.2, overleaf.

Ostensive-inferential comprehension is achieved by the continual, parallel integration of relevant contextual information. According to this model, the linguistic knowledge hearers use to 'decode' utterances functions only as an input to this central inferential process. The conventional, associative meanings that constitute linguis-

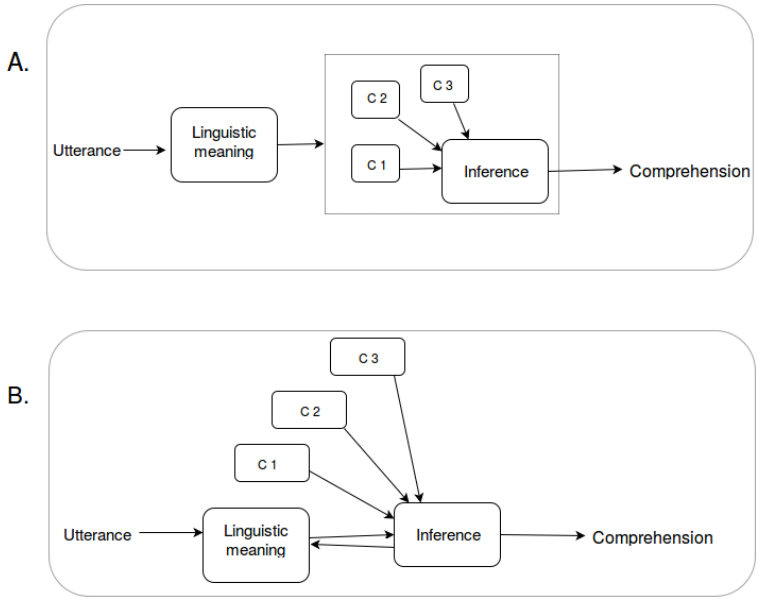


Figure 1.2: Two models of inferential communication. Model **A** represents a *'two step' model* of inferential communication whereby the associative linguistic meanings activated by a linguistic utterance are then supplemented by additional contextual information. Note that the context items integrated in comprehension are triggered by the linguistic meaning. If the linguistic meaning is the appropriate output, that this is the case must itself be inferred. This model represents an early formulation of inferential comprehension. Model **B** represents a reformulation of A, whereby linguistic meaning is one of many stimuli that feeds into a central inferential system; the linguistic meaning in this model is never a candidate output; instead, linguistic meaning constitutes a logical form which is subject to modification by contextual inference, and the output is an optimal integration of all relevant inputs. This model is known as the *ostensive-inferential* model.

tic knowledge are not possible outputs of comprehension, but instead yield a 'logical form' that must be integrated with contextual information in order to achieve comprehension (Sperber & Wilson, 2002). A comprehension process where the integration of relevant context is front and centre addresses the problem of how linguistic conventions could have been established. A key feature of ostensive-inferential comprehension is that it makes the establishment of linguistic codes possible (Scott-Phillips, 2014). Interlocutors can use metapsychological, social cognition to draw conclusions about what a speaker or gesturer might intend by their speech or gesture. The ability to infer about communicative behaviours means that conventions can eventually become established and exploited. This inference-driven 'de novo' formation of conventions stands in contrast to the evolution of natural codes, which are fixed to particular functions and follow a lengthy process of ritualisation/sensory manipulation (Wharton, 2003). An

inference-first account of language emergence is consistent with the current state of knowledge regarding the evolution of social intelligence in humans (see Dunbar, 2003, for a review). Briefly: as human social groups grew in complexity, social cognition became more powerful and complex in turn, including metapsychological cognition (Brothers, 1990; Dunbar, 1998b), thus providing the cognitive wherewithal for more complex communication (Frith & Frith, 2010). The ostensive-inferential model thus lends itself to a clear evolutionary picture of the emergence of inferential communication and, in turn, conventional language, from social-cognitive substrates. The use of inference to establish communicative conventions is proposed to have eventually given rise to the languages we observe today, through the repeated use and cultural propagation of those conventions (Scott-Phillips, 2014). Once they have been created, cultural transmission itself can significantly shape the structure, efficiency and complexity of linguistic conventions (Kirby & Hurford, 2002; K. Smith, Kirby, & Brighton, 2003; Kirby, Cornish, & Smith, 2008; Steels, 2011).

A proposed consequence of this emergence is the natural selection of mechanisms that contribute to the effective performance of ostension and inference (Sperber, 2000; Sperber & Wilson, 2002; Scott-Phillips, 2014). One such mechanism is a dedicated module for the processing of ostensive communication, of which linguistic 'decoding' is a part (Sperber & Wilson, 2002). Such mechanisms should operate only as input to a central inferential process, and so 'linguistic meaning' is never a potential output of comprehension. In fact, the form of linguistic meaning is itself subject to the interpretation of central inference, as represented by the two-way arrows of influence between linguistic meaning and inferential processing in Model B of Fig. 1.2. On this view, a hearer could never arrive at a 'literal meaning' interpretation because literal meaning does not exist in the familiar sense. Rather, comprehension of linguistic utterances is fundamentally integrated with and *determined by* relevant contextual knowledge. Thus, linguistic meaning is not a privileged domain of input to comprehension, but rather one of an array of concurrent inputs.

1.2 How has the comprehension system been shaped by the linguistic environment?

The evolutionary problem for two-step and code models of communication is that the establishment of conventions requires inference; a signal that is not grounded in a natural process of ritualisation or sensory manipulation could potentially mean anything. Natural codes such as mating calls have narrowly specified meanings that have been shaped by a long evolutionary process that grounds such signals in a particular function. Without this grounding, arbitrary conventions are radically underdetermined. This radical underdeterminacy of intended meaning is precisely the problem that inference appears well-designed to solve. Inference is a means of reaching an interpretation that has been narrowed from *literally anything* to a certain *something in particular*, along well-motivated (or even reasoned) lines. Inference allows interlocutors to reach an interpretation of a communicative behaviour, and eventually to converge upon shorthand conventions for particular meanings. It thus bridges the gap between non-linguistic and linguistic communication in humans. But if ostension and inference can already facilitate the successful comprehension of radically underdetermined signals of any type or modality, why should language emerge at all?

Language can be conceived as a tool for narrowing the scope of potential meanings a signaller may intend. The emergence of language made ostensive-inferential communication 1) far more expressively powerful, and 2) far more precise (Scott-Phillips, 2014). The expressive power of language is infinite; the finite elements of language, and the meanings they represent, can be recombined endlessly. The structural properties of language that facilitate this are arguably the most studied subject in evolutionary linguistics. Of particular interest for the current exploration is the second contribution of language to the existing ostensive-inferential system: *precision*. The conventional, associative meanings of linguistic elements are stable and similar between users of the same language, and this alone has the potential to reduce the vast space of possible

interpretations - a particular word has a particular scope.

In other words, protolinguistic conventions emerged between conspecifics - and subsequently proliferated and developed into fully fledged languages through cultural processes - as part of a strategy to narrow down the radical ambiguity of communicative behaviours that were not grounded in an evolutionary ritualisation process. Language would not be able to make the significant contribution of *precision* if linguistic meaning did not function to narrow the range of possible interpretations more than other kinds of information. This point is intuitive; inferred communications between speakers of different languages are especially laborious, and require more trial and error in interpretation, when compared to a conversation between two fluent language users. Common conventions give interlocutors similar representations that can be employed for more precise, efficient communication. This fact about the function of language holds a subtle implication: the linguistic environment functions to relieve the problem of radical underdeterminacy that necessitated ostensive-inferential cognition in the first place. As a result, the efficient use of linguistic meaning itself may play a more significant role in the comprehension process than the ostensive-inferential model, as currently conceived, may suggest.

Linguistic meaning constitutes a valuable input to the ostensive-inferential model. Recall that a linguistic utterance yields a *'logical form'*, which provides input to the comprehension process alongside other relevant contextual parameters (Sperber & Wilson, 2002). The derived logical form may also be supplemented, or 'enriched', with contextual information in order to derive a proposition which can then contribute to interpretation (Sperber & Wilson, 1986; Carston, 1988; Wilson & Sperber, 1993; Sperber & Wilson, 2002, see also §2.1). In other words, the fully propositional form of a linguistic utterance can only be derived by enriching the logical form with relevance-oriented information. This relationship of mutual influence between linguistic meaning and inference is represented by bi-directional arrows in Model B of Fig. 1.2. Conceived this way, linguistic meaning is itself responsive to contextual information. As such, the comprehension of linguistic utterances - and the linguistic meaning itself - is

fundamentally context-driven.

The status of linguistic meaning as a logical form places a necessary limit upon its role in the ostensive-inferential model of comprehension. Specifically, this form of representation functions only as an input to the inferential process; it can never constitute an output. Note that this applies even if a speaker's intended meaning is entirely accounted for by the linguistic meaning. Cases like these, known in the ostensive-inferential approach as *explicatures*, must also be established by the use of inference (Bach, 2010).

If linguistic meaning were to play a more significant role in supporting the comprehension process, it would need to involve different properties to the logical form in the ostensive-inferential account. A defining aspect of the inferential-ostensive model is that linguistic meaning must be augmented by the inferential process; the logical form is limited to the role of *input*. If linguistic meaning were to support a full comprehension response, it must entail some properties that can constitute an *output* of comprehension. There are some indications from neurolinguistic studies that this is plausible. In particular, there is a growing body of evidence in support of embodied semantic representations (see Aziz-Zadeh & Damasio, 2008, for a short review), which suggests that conceptual representations (including more abstract concepts) share sensori-motor circuitry with enactment (Gallese & Lakoff, 2005). For example, the passive reading of action words that relate to particular areas of the body, such as *"pick", "lick"*, and *"kick"*, has been shown to elicit the activation of corresponding areas in the motor and premotor cortex (Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller, 2005). Similar correspondences in cortical activity between enaction and conceptualisation have been found at the phoneme level (D'Ausilio et al., 2009) and the sentence level (Tettamanti et al., 2005). It can be argued that this kind of neural activation is elicited by the *outcome* of comprehension, rather than the process. Findings of neurofunctional double dissociation for recognising phonemes (that is, impaired recognition in cases where the corresponding motor substrate is compromised) give some indication that these neural substrates may play a more causal role in the process (D'Ausilio et

al., 2009). In principle, then, the representation of linguistic meaning could plausibly involve some qualities that lend themselves to the status of *output* in comprehension.

Linguistic associations emerged to disambiguate communicative exchanges in concert with ostensive-inferential abilities. In order for these linguistic associations to develop a more prominent role in the comprehension process, they must be reliable enough to constitute a special domain of information. In other words, a given individual's linguistic representations ought to be a reliable enough proxy for the meaning that other interlocutors intend to communicate. If the linguistic environment - that is, the inventory of linguistic associations that a speech community acquires - represents a stable resource that can be exploited in comprehension, we may expect this fact to have shaped the comprehension system over phylogeny. The properties of the linguistic environment itself may provide the wherewithal for changing the status of linguistic representations in the comprehension process.

Linguistic systems are subject to processes that make them exceptionally well-designed tools for communication (e.g. Kemp & Regier, 2012; Regier, Kemp, & Kay, 2015), and they constitute a cultural environment that has impacted human evolution. Languages can evolve over time into structured inventories that are expressive and learnable for the population (Kirby et al., 2008; Cornish, Tamariz, & Kirby, 2009). Cultural transmission processes have also been found to result in robust semantic categories; linguistic items come to denote underspecified, yet stable, scopes of meaning along dimensions that are salient to the task domain (Silvey, Kirby, & Smith, 2015). In short, the linguistic environment has culturally evolved to become very good at facilitating communication, and exhibits robust structural qualities that include its semantic associations. The successful acquisition and use of symbol-meaning mappings can be underpinned by fundamentally associative mechanisms that may be particular to humans (K. Smith, 2001, 2003). Human adaptation to the cultural environment is implicated in a range of cognitive and physiological traits, some as recently as 6000 years ago (Kingdon, 1993; Aoki, 1986). Though there is no specific consensus, humans have certainly had linguistic culture for far longer than this, with most estimates ranged be-

tween 50,000 and 250,000 years (Botha & Knight, 2009; Perreault & Mathew, 2012). During this time, humans have become physiologically adapted to the production of speech (Fitch, 2000; Ghazanfar & Rendall, 2008), and cognitively adapted for the use of language. The wide range of cognitive abilities employed in language include the associative fast-mapping of word meanings by young children after a single exposure (Spiegel & Halberda, 2011), and associative biases in the acquisition of linguistic meanings (Markman, 1991). While associative cognition almost certainly pre-dates the emergence of language, the co-ordination of these abilities to address the task of language is a process of adaptation (see also §2.3.1).

Taken together, these facts suggest that there is at least the wherewithal for further adaptation of the human comprehension system in response to language. This point is broadly in agreement with the standard ostensive-inferential model (Sperber & Wilson, 2002; Scott-Phillips, 2014), which suggests that linguistic decoding has become an additional input to comprehension. But there is potential for an alternative suggestion that extends beyond this. A robust cultural system of linguistic meaning associations, which functions to coordinate the representations of a population, constitutes an environment in which the comprehension system may reasonably and efficiently exploit associative meanings as a 'good-enough' interpretation in and of themselves, at least some of the time.

This picture presents an alternative contribution to the pragmatic conversation in evolutionary linguistics. Model C in Fig. 1.3 illustrates this departure from the standard ostensive-inferential model. The inferential foundations of communication remain in place in this model of comprehension. However, it also assumes and exploits the stable associations that constitute an individual's linguistic knowledge, to the degree that linguistic meaning alone may support an interpretation. The suggestion is that this linguistic specialisation can, in principle, support the entire process of comprehension, at least on some occasions. This departs from the standard ostensive-inferential model in two key respects. Firstly, linguistic input is of specialised interest to the comprehension system relative to other sources of information, e.g. metapsychological

representations about the mind-states of speakers. This may be the result of access to associative representations that are more stable and available to the comprehension system than parallel inferred representations. Secondly, the content an individual associates with a linguistic utterance is a possible candidate output for the comprehension process, without necessarily being modified by contextual or metapsychological integration.

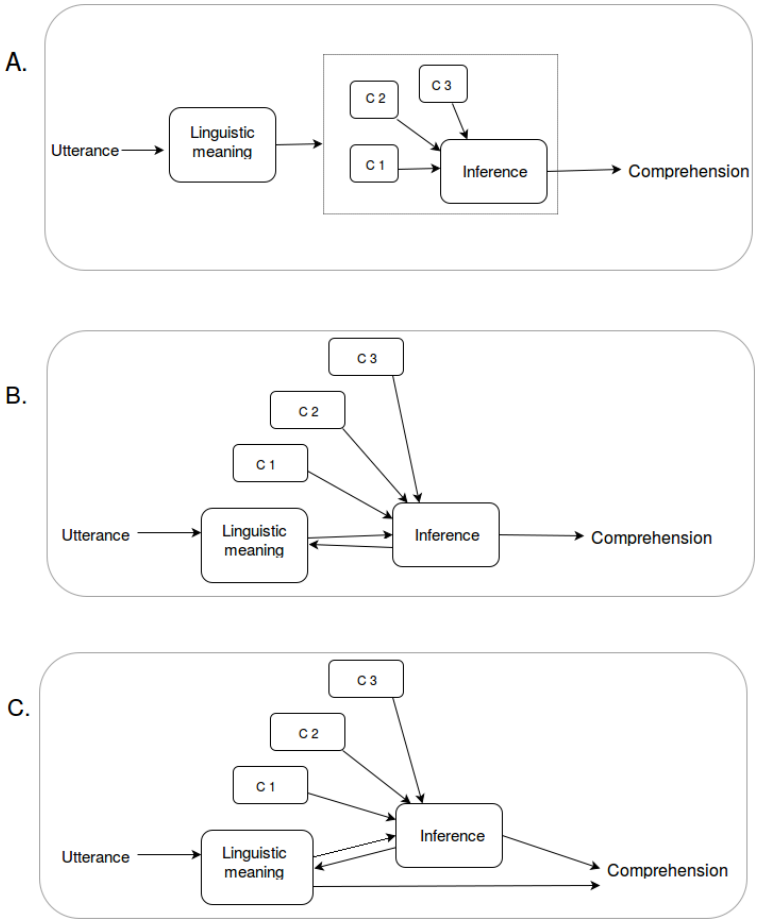


Figure 1.3: Three models of inferential communication; an extension of Fig. 1.2. Model **A** represents a *'two step' model* of inferential communication whereby the associative linguistic meanings activated by a linguistic utterance are then supplemented by additional contextual information. Model **B** represents a reformulation of A, whereby linguistic meaning is one of many stimuli that feeds into a central inferential system; the linguistic meaning in this model is never a candidate output. This model is known as the *ostensive-inferential* model. Model **C** represents a modification of B, whereby linguistic meaning is one of many stimuli that feeds into a central inferential system. However, the linguistic meaning in this model is a potential output of the comprehension process, and is activated in parallel cascade to other contextual information. This model represents a schema of linguistic comprehension that is further elaborated and motivated in chapter 2.

With respect to these two key differences, the schemas for Model B and Model

C in Fig. 1.3 require some clarification. Here, a distinction has been made between representations of linguistic meaning that bypass the inferential process and those that go through it. This is a function of a qualitative difference in what is represented by 'linguistic meaning' in each model.

For the ostensive-inferential Model B, the decoded linguistic meaning yields a logical form which must be augmented by a central, inferential process that integrates contextual input. On this view, linguistic meaning is never a candidate output of comprehension. All inputs to the comprehension process are qualitatively different from the output, which must be constructed via a process of inference. This is illustrated by the procedural description given by Sperber and Wilson (2002), whereby expectations conferred by the context inputs determine the order in which pragmatically-enriched interpretations of the utterance are considered. In this way, linguistic input is augmented by the context, and the comprehension process constructs an output interpretation *on the basis of* the inputs.

By contrast, the linguistic meaning in Model C entails representations that may sufficiently constitute a comprehension output. The suggestion is that these representations may bypass the inferential process, and that this is only possible in Model C. Note that the inferential process in Model B is also available here in Model C; the difference is that linguistic meaning has also taken on properties such that accessing these representations suffices for comprehension on at least some occasions. Linguistic meaning in Model C thus offers a specialised route to comprehension, and this domain of input is of special interest to the comprehension system.

One question that arises from this linguistic specialisation is whether the process could ever, in principle, disregard the linguistic material in reaching an interpretation. To illustrate, consider the following scenario: *Ann is in a hurry while painting a mural, and needs to switch between colours often. Bob is providing assistance by handing her brushes from different pots of paint. Bob knows that the next component of the mural requires blue, and he also knows that Ann is in a hurry. Ann, quickly holding out her hand, says "red". Bob hands her a brush from the blue pot, and Ann continues*

painting. Did Bob's comprehension process disregard the linguistic input in order to reach his interpretation?

In both models, there are dedicated modules targeted on linguistic input. Though this input may be weighted against other contextual factors, and perhaps overridden by them, neither model predicts that an input simply does not feature in the comprehension process once it has been identified as an input. In the example above, Bob does indeed compute the linguistic meaning of "red", but this is outweighed by the integration of everything else he knows about the situation, such that he reaches the interpretation that Ann meant "blue". In this way, the linguistic meaning may be overridden and in some cases modified (as indicated with bi-directional arrows in the schemas above) by the contextual information, rather than discounted from the comprehension process entirely.

Though this inferred interpretation is available in both models, the potential for a direct route to comprehension in Model C means that interpretations like "blue" may be less likely to occur. Hearers endowed with Model C may indeed be more likely to reach for the red brush in this scenario (see also §2.2 for similar examples of systematic errors in referential tasks). A specialised route to comprehension supported by linguistic meaning alone predicts more sensitivity to linguistic input, relative to other sources of information. In other words, while contextual modification of linguistic meaning is possible for both models, comprehension responses based on linguistic associations alone are only predicted by Model C.

1.3 Aim of the thesis

In this thesis I aim to investigate the pragmatics of linguistic communication using evolutionary theoretical tools and psycholinguistic experimental methods. I use an adaptationist framework to formalise two mechanistic accounts of linguistic comprehension and identify contrastive features for empirical investigation. I will present a series of experiments that use referential tasks to explore how linguistic meaning, contextual

information, and metapsychology contribute to referential processing. In particular, I will use conversational precedent paradigms (Kronmüller & Barr, 2007; Kronmüller & Barr, 2015) and tangram studies (cf. H. Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989) to investigate the sensitivity of the comprehension system to linguistic input in comparison to: the knowledge-states of speakers; the broader cultural common ground of the language-speaking community; and the visual context that accompanies particular referents. The evidence presented to discriminate between both accounts of comprehension lies in two measures. Firstly, special domains of input should affect how easily and thus how quickly hearers are able to interpret a given utterance. This is measured by reaction time. Secondly, a hearer's privileged associations with linguistic input are predicted to be a possible outcome of comprehension by one account, and an impossible outcome by the other. The comprehension outcome is measured by selections made in the referential tasks. In a follow up study, I employ a visual context paradigm and attempt to investigate the production of reference for clues about which forms of information are prioritised by speakers and hearers. The overall aim is to contribute to the development of a particular interdisciplinary approach that has thus far remained overlooked, and use this to investigate whether egocentric linguistic associations can manifest a full comprehension response as outlined in model C of Fig. 1.3. This generates new questions about the ongoing impact of the linguistic environment on human cognition.

1.3.1 Thesis Road Map

The structure of the thesis is as follows: In chapter two I review a broad polarisation in the pragmatic and psycholinguistic literature concerning the relative roles of linguistic convention and contextual information in comprehension. I make explicit the theoretical approaches that reliably give rise to these polar positions across scholarly domains. I go on to map each model of comprehension to the adaptationist particulars it may entail, and in doing so illustrate two different pictures of how linguistic cognition has

developed over phylogeny. The Social Adaptation Hypothesis (SAH) holds that linguistic comprehension is performed by relevance-oriented inferential mechanisms that have been selected for by a social environment (i.e. inference-using conspecifics). In particular, the SAH holds that linguistic conventions are attended to in the same way as other ostensive stimuli and contextual information, and because of their relevance to communicative interactions. The Linguistic Adaptation Hypothesis (LAH) holds that linguistic comprehension is performed by specialised cognition that has been selected for by a linguistic environment (i.e. language-using conspecifics) that was established subsequent to, and as a consequence of, the emergence of inferential communication. In particular, the LAH holds that linguistic conventions are a privileged domain of input for the comprehension system.

In chapter three I present Experiment 1, which uses speaker-specific facts about referents as referring expressions in a conversational precedent paradigm. Participants must identify objects in an array, following instructions delivered by two different speakers who do not know one another. The referring expressions explicitly make use of mental states, such that objects are identified in terms of what they remind the speakers of. If the comprehension of participants relies on metapsychology about the speaker to generate an interpretation, any inconsistency in this information should lead to longer reaction times and random object selections. If the comprehension of participants relies on associations with linguistic labels themselves, inconsistent labels should lead to longer reaction times and incorrect responses. Inconsistent linguistic labels did indeed elicit the longest reaction times and the least successful object selections, indicating a central role for linguistic associations. Methodological issues are also discussed.

In chapter four I present Experiment 2, which introduces a further contextual variable into the conversational precedent paradigm: culturally copresent associations between labels and referents. The experiment presented in this chapter compares the relative sensitivity of processing to culturally copresent common ground, the privileged knowledge state of speakers, and the consistent use of linguistic labels. The SAH pre-

dicts that comprehension will be more sensitive to the common ground of community membership with the individual speaker; the LAH predicts a greater sensitivity to the consistency of linguistic labels. The results indicated a main effect of linguistic label on reaction time, with interactions with cultural congruence and speaker. Object selection measures suggested that hearers deploy default cultural knowledge when linguistic labels fail, rather than for new speakers. This is consistent with the LAH.

In chapter five I present two pairs of experiments: experiments 3 and 4, which explore visual context as a constraint on linguistic comprehension, and experiments 5 and 6 which employ a similar paradigm to explore the effect of visual context on the production of referring expressions. Experiments 3 and 4 use tangrams to rearrange the visual context around particular constituent shapes. These shapes are the targets in a referential task, where the congruence of either the linguistic label or the visual context were altered between trials. The reaction time and shape selection of participants were measured, with the results indicating that both visual context and linguistic label exerted comparable effects on comprehension. This is consistent with the SAH. Experiments 5 and 6 employ a similar tangram paradigm for production; if comprehension is characterised by particular sensitivities, we may expect speakers to produce utterances that lend themselves well to how hearers process them. The experiments measure the consistency of produced labels for the same referents across congruous, incongruous and abstract contexts. The results highlight some methodological limitations of the tangram paradigm for the study of repeated reference across contexts, indicating that speakers may conceive of referents as independent entities in each presentation, rather than recurrent. I discuss the implications of this along with the general results of the experiments in this chapter.

In chapter six I conclude by considering the SAH and LAH in light of the literature and empirical evidence above along with its accompanying limitations. I argue that the evidence is generally consistent with the assumptions of the LAH, and suggest future avenues of investigation.

Chapter 2

The psychological architecture of reference in context

2.1 Theoretical approaches to pragmatic competence

A central issue that historically underpins the field of pragmatics concerns the successful resolution of linguistic ambiguity. Quine (1960) illustrated the core of this problem with his canonical thought experiment on the inscrutability of reference. The story goes that a linguist's attempt to translate the meaning of the word "gavagai" in a newly discovered language can never reap an entirely certain outcome; even when the native speaker utters this word while pointing to a rabbit, there are many candidates for what the intended referent is, such as *rabbit parts*, *edible animal*, or more abstract meanings like *rabbit tropes*. Quine argued that the number of possible meanings here is in fact incalculable, which renders the task of discriminating between them through repeated observations fruitless. The conclusion is that an accurate translation of "gavagai"'s meaning is impossible to discern by any method. The most successful possible outcome can only constitute a reasoned interpretation that results from assumptions, experience and/or other such heuristics that the linguist herself brings to the table. Such extralinguistic factors that govern successful language use are the domain of the

pragmaticist, and pertain even to mundane exchanges between speakers of the same language. Consider the question "How about tea?". The speaker could be: offering someone else tea; asking for tea themselves in response to being offered coffee; suggesting an item on a shopping list; proposing to meet in a cafe; the list goes on. In this way, the intended meaning of a given utterance is not sufficiently specified by the utterance itself. Examples of this linguistic underdeterminacy abound, most obviously including common referring expressions such as anaphora (*she, that, it*). Thus, even linguistic tasks such as definite reference, which we may intuitively consider rather simple, pose a pragmatic challenge to the listener.

Gricean pragmatics provided the most influential early account of how interlocutors approach the problem of linguistic underdeterminacy (Grice, 1975, 1991). Grice described the way that language users tend to behave cooperatively in discourse, and how this behaviour appeared to facilitate the interpretation of underspecified utterances. This was summarised by The Cooperative Principle, which is: "Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged". This description of how interlocutors tend to behave comprised four Conversational Maxims: quantity (contributions should be true), quality (contributions should provide no more or less than is required), relevance (contributions should be relevant) and manner (contributions should be clear and concise). On the Gricean view, the adherence of speakers and hearers to such conversational maxims facilitates implicature and inference. Implicature refers to what is *meant* when a speaker makes an utterance, without being part of what is actually *said* in the utterance itself. Inference refers to the hearer's comprehension of an implicature. One particular subset of implicatures, called scalar implicatures, remains a lively area of inquiry in pragmatics, and makes for a good illustration of the general concept. The following example is adapted from Breheny, Katsos & Williams (2006):

John: Was the exam easy?

Mary: Some of the students failed

(a) Some and possibly all students failed

(b) Not every student failed

(c) The exam was not easy

The truth-conditional semantic content of Mary's reply is (a), but the meanings typically inferred during normal conversation tend to be (b) and (c). It is precisely because the truth-conditional content of Mary's utterance does not account for the entire intended meaning that hearers must *infer* it by using other sets of information available to them.

Gricean implicatures are inferred according to a two-stage process: first, the hearer must parse the literal meaning of the utterance (with its attendant truth-value). This is then compared against relevant cooperative maxims, discourse goals and context in order to infer the speaker's meaning. Various amendments have been made to Gricean pragmatics in order, among other things, to account for the extralinguistic dependencies that are necessary to establish truth-value (e.g. anaphora), to refine the maxims in more parsimonious terms, and to account for further discourse observations (Gazdar, 1979; Horn, 1984, 1990, 2008; Bach, 2001; Levinson, 2000). One such refinement concerns the distinction of Generalised Conversational Implicatures from Particularised Conversational Implicatures, the former being a conventional and defeasible type of 'default inference' that is automatically generated by a specialised pragmatic system as a function of parsing the utterance (Levinson, 2000; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia, 2004). In the example above, (b) would constitute a conventional (Generalised) implicature, whereas (c) may be considered a more context-dependent (Particularised) implicature and thus inferred at a later stage by incorporating more extralinguistic information. So-called 'default' inferences and propositions of what Levinson (2000) calls "pragmatic intrusion" on semantic content are the subject of many lively exchanges in the literature regarding the validity of such entities and their compatibility with Gricean pragmatics (Horn, 2005; Carston, 2005; Horn, 2006b, 2006a; Cappelen & Lepore, 2008; Borg, 2009, 2013).

However, the salient characteristic common to these different Gricean approaches is that the semantic proposition of the utterance gives rise to inferences on the basis of its content. Though care must be taken in translating Grice's top-level 'rational reconstruction' of pragmatic inference into a mechanistic procedure without oversimplification (Noveck & Spotorno, 2013), recent experimental investigations of implicatures have managed to do so fruitfully (Bezuidenhout & Cutting, 2002; Bezuidenhout & Morris, 2004). The defining points of a Gricean approach do exist such that it can be contrasted with other approaches in practice (Breheny et al., 2006; Noveck & Reboul, 2008). The important feature of Gricean inference for the current exploration is the two-stage process of inference: interlocutors first parse the linguistic utterance for conventional meanings - and by some accounts, conventional implicatures (Potts, 2005) - which serves as input to a pragmatic system that reconciles this against the context and goals of the discourse in order to yield a speaker meaning. In order to function along these lines, Gricean inference makes an important assumption: it requires that linguistic conventions themselves are stable enough to inform comprehension, in a way that is independent of the context. The linguistic form thus constitutes a privileged type of input to Gricean inference, from which a conventional meaning can be derived.

This assumption stands in contrast to Relevance Theory, an alternative to the Gricean approach (Sperber & Wilson, 1986). There are two central principles. The first is the Cognitive Principle, which holds that human cognition tends to be geared toward the maximisation of *relevance*. Relevance here is defined as that which yields the largest cognitive effects for the smallest processing effort. While Relevance Theory does not suggest a way of quantifying either cognitive effort or effects, or indeed that there can be an absolute measure (Wilson & Sperber, 2008), it proposes that the relevance of two inputs can often be compared by manipulating factors that pertain to how effortful the processing requirements of certain items are, such that a comparative difference between items can be measured (see Sperber, Cara, & Girotto, 1995, for an example). The Cognitive Principle holds that particular cognitive goals at any given time will be consistent with the more general goal of reaping the greatest cognitive effects for

the least processing effort. In other words, human cognition is automatically geared toward making this trade-off, including in the case of communication.

The Communicative Principle holds that every ostensive stimulus carries a presumption of its own optimal relevance. This second principle requires some unpacking, and offers a picture of how human communication works in the relevance theoretic model. Ostension and inference refer, respectively, to the expression and recognition of both informative and communicative intentions. An informative intention refers to a given speaker's intention to change the mental representations of the hearer; in other words, a speaker's intention to inform a hearer of something. A communicative intention is the intention that the hearer recognises *that the speaker has* an informative intention. In an ostensive-inferential model of linguistic communication, then, a speaker provides evidence for their intentions in the form of a linguistic utterance, though this is only one of many types of ostension, others including e.g. gestures. The hearer then interprets this evidence with the assumption that it will be optimally relevant to them given the conditions in which the utterance has taken place.

To illustrate this account, consider the following example adapted from Scott-Phillips (2014) (in turn adapted from Sperber (2000)): Mary is eating berries in the view of Peter, and wants to communicate that the berries are tasty. To do this, she pats her tummy and eats in an ostensive, stylised way. This reveals to Peter that the berries are tasty (her informative intention), and the manner of her action reveals that Mary *wants to communicate this* to Peter (her communicative intention). The latter would not have applied if Mary had simply eaten enthusiastically in a usual way, and so would not have constituted an intended communication. If we were then to further propose a scenario where Peter is curious about the berries and Mary knows this, she can express the relevant informative and communicative intentions by saying "Yes". Peter will understand this if and only if *he knows that she knows about his curiosity*. This example illustrates an important point that informs the relevance theoretic account of linguistic processing: Mary produces different evidence according to her knowledge of Peter's goals, and Peter's interpretation of the evidence is similarly constrained. This

is the Communicative Principle in action; ostensive stimuli carry the presumption of their own relevance precisely *because* Mary produces her ostensive utterance for Peter according to the constraints of their relevant knowledge. The appropriate production and treatment of utterances, then, depends on a range of contextual factors, including the mutual knowledge of the interlocutors (in relevance-theoretic terms, their *mutual cognitive environment*, in more psycholinguistic terms, their *common ground*). That is to say, the *linguistic content of the utterance itself* is produced and comprehended according to these contextual factors.

The differences between the relevance theoretic account and the (neo-)Gricean account may not be immediately obvious, and indeed the early formulation of relevance theory was presented in terms similar to the Gricean two-stage process (Sperber & Wilson, 1987; Wilson & Sperber, 1993). There is, however, a subtle and important difference. Recall that, on a relevance theoretic view, a given hearer's inferences are governed by the Cognitive Principle that we tend to maximise relevance, which is to say we attend to that which confers the largest cognitive effect for the least effort. In the example above, Peter's expectations of relevance are a function of the context, which includes (among other things) the knowledge he mutually shares with Mary. It is the constraint of these contextual factors that determines the least effortful route of interpretation for Peter. Sperber and Wilson (2002) formalise the comprehension procedure thus: (a) follow a path of least effort in computing cognitive effects; in particular, test interpretations in order of how manifest and accessible they have been made by the context, and (b) stop when your expectations of relevance are satisfied. In this way, Peter's processed interpretation of the linguistic utterance *in the first instance* is determined by the context in which he hears it. The relevance theoretic account is thus context-driven, such that the hearer arrives at the relevant interpretation of the utterance without ever considering more conventional or literal alternatives that aren't warranted by the context (Wilson, 1994; Sperber et al., 1995; Carston, 1998a). As Sperber and Wilson put it: "The hearer does not *first* decode the sentence meaning [...] *then* derive a range of implicated conclusions. [explicatures and implicatures] are developed in

parallel, against a background of expectations that may be revised or elaborated as the utterance unfolds." (Wilson & Sperber, 2012, p. 14, emphasis in original). The role of language in the relevance-theoretic model is thus not conceived as a two-stage process, and some hearer expectations are in fact suggested to modify the identification of explicit meaning through "backwards inference" (Wilson & Sperber, 2012, p. 14). This can be conceived, then, as the inverse scenario to Gricean inference, where the linguistic utterance is first conventionally interpreted and *then* reconciled with contextual information.

While both accounts agree that the task of interlocutors is to produce and interpret evidence for intended meanings, a principal difference concerns the primacy of linguistic conventions in approaching this task. The nature of the distinction drawn between explicit and implicit meaning is a lively point of disagreement between relevance theoretic and Gricean pragmaticists, and is part of what Horn (2006a) terms the "semantics/pragmatics border wars". The traditional pragmatic model which, though modified, still persists in neo-Gricean formulations, holds that "non-truth-conditional aspects of meaning are read off the output of semantically interpreted logical form" (Horn, 2006a, p.1). While there is disagreement on the extent to which some implicatures (that is, parts of the speaker's intended meaning) intrude on semantic content, these details do not modify the general assumptions of a two-step model. The contextual enrichment of word meaning ostensibly provides a challenge to the notion of literal or semantic meaning. For example, it is argued that the compound noun *"swan boat"* must be understood with the contextual provision of a relevant relation between both nouns to derive 'boat that looks like a swan', even though neither noun is typically context-sensitive (in the way that, say, pronouns are) (Recanati, 2010). Enrichments like these do not challenge the Gricean notion of conventional meaning, as they are simply not considered part of semantic content by definition, under the Gricean rubric (Bach, 2001, 2010; Horn, 2006a).

It remains common to the various Gricean positions on this issue that there does exist a semantic propositional form, that it arises from the sentence uttered, and that this

can be considered in some way independent from the context. The stability that this view affords to the conventional content of utterances leads to the relatively conservative treatment of ambiguity in general. Grice's Modified Occam's Razor states that "senses are not to be multiplied beyond necessity" (Grice, 1991), and this sentiment is echoed elsewhere in the contemporary literature: "Do not posit an ambiguity unless you are forced to, unless there are really compelling intuitive grounds to suppose that an ambiguity really is present" (Kripke, 1977).

Semantics in these terms - that is to say, linguistically-derived meaning that is separate from contextually/pragmatically derived meaning - does not play a role in the relevance theoretic account of inferred meaning. On this view, there is no such division between conventional linguistic meaning and contextually-derived pragmatic meaning. A distinction is instead made between explicit speaker meaning (*explicature*) and implicit speaker meaning (*implicature*), both of which are context-dependent (Carston, 2009). For illustration, consider the following example (from Carston, 1998b):

Alex: How was the party? Did it go well?

Brit: There wasn't enough drink and everyone left early.

(a) There wasn't enough *alcoholic* drink *to satisfy the people at [the party]*$_i$ and *so* everyone *who came to [the party]*$_i$ left *[the party]*$_i$ early.

In the example above, the explicit content of Brit's reply is the conjunction of two propositions: (i) there was not enough drink, and (ii) everyone left. However, the propositional scope of 'drink' and 'everyone' is far beyond that which Brit intended to denote; consider all the different types of drinks denoted by 'drink' and all of the people denoted by 'everyone'. It is considered more likely that Alex would take Brit to have directly expressed a proposition like (a), which in turn is a basis for the reasoned inference of the implicit meaning (that the party did not go well). (a) can be characterised here as a meaning that *directly develops on the logical form of the utterance*, and this constitutes an explicature (Sperber & Wilson, 1986). Where, in the Gricean schema, the formal proposition is narrowed by further inference, a relevance-theoretic account proposes that Alex's context-determined expectations of the exchange form

part of the proposition he gleans from the linguistic utterance in the first instance. Taken to its extreme, this approach poses a challenge to the notion of any literal or 'semantic' meaning, a debate that has taken various forms in the literature (see Ariel, 2002, for a concise review). In light of Relevance Theory's context-bound characterisation of linguistic meaning (Carston, 1998b, 2009; Wilson & Carston, 2007), and a context-driven comprehension procedure that may never consider a literal meaning at any point (Sperber & Wilson, 2002), it appears that any reliable associations that constitute 'knowing a language' primarily function to constrain one's expectations of relevance in probabilistic ways. Such factors become more or less relevant depending on the given situation. In other words, our sense that "cat" reliably denotes feline mammals in most situations may be somewhat comparable to our probabilistic expectation that a falling ball will bounce back up: the latter expectations are active in a game of basketball, while the former will be brought to bear on a dialogue. On this view, the role of linguistic knowledge is to aid a fundamentally pragmatic process of deriving meaning. That is to say, linguistic knowledge confers some expectations of relevance (about how people typically use words), and these govern the inferred interpretation alongside other relevant factors.

The status of linguistic conventions differs between these accounts, and this gives rise to different strategies of interpreting speaker meaning. For the relevance theorist, the semantic content of language is not afforded the stability that could warrant its primacy in linguistic comprehension. Rather, every utterance is considered radically underdetermined, such that the number of possible interpretations on any given occasion is infinite: "Linguistic codes [...] have an *unlimited* range of things they might refer to" (Scott-Phillips, 2014, p. 18, emphasis added). As such, linguistic conventions are not fundamentally different to any ostensive stimulus (such as a stylised cough), and semantic meaning does not necessarily exist as such at all, and not in a form that could warrant the treatment of linguistic utterances as a special domain of input.

Griceans may well agree that various combinations of linguistic meaning and context derive infinite possibilities, but the central assumption in this case is that the con-

ventional associations acquired by the hearer are able to limit these possible meanings *enough* that they are worth consulting as a matter of course. An illustration of this difference can be seen in the case of scalar implicatures. Consider the examples in table 2.1, taken from (Horn, 2006a).

Scalar utterances	*Lower bounded meaning*	*Double bounded understanding*
a. Pat has **3** children	'… at least 3...'	'… exactly 3...'
b. You ate **some** of the cake	'… some, if not all...'	'… some, but not all...'
c. He's a knave **or** a fool	'… and perhaps both'	'… but not both'

Table 2.1: Scalar implicatures, with lower bounded propositional forms and double bounded interpretations.

On the Gricean view, the semantic content of a scalar provides a lower bound to the proposition. This leaves pragmatic ambiguity as to the upper bound, which is resolved by inference from the context. On a relevance theoretic view of such scalars, however, "the ambiguity has been relocated to the propositional level" (Horn, 2006a, p. 4), whereby the double-bounded understandings constitute an *explicature*; that is, a pragmatically-determined component of what is explicitly said (as opposed to something that is implicated) (Carston, 2008). In other words, relevance theory regards the semantic content as fundamentally uninformative on its own, and the solution to this problem is a pragmatically-determined propositional form. Gricean pragmatics, on the other hand, affords the semantic content some reliable value, even though it may be underdetermined. The key difference between approaches, then, concerns whether or not the conventional associations of language, once acquired, are informative enough that we may rely on them in the absence of other inputs to inferential processing in linguistic communication.

The question, in other words, is this: given the task of recovering intended meaning, are linguistic conventions useful enough to this end that we attend to them in a special way? Or does metapsychological inference based on the context and the speaker recover the meaning of fundamentally ambiguous utterances? The following section

consults the psycholinguistic literature on the integration of linguistic and contextual sources of information in comprehension.

2.2 Psycholinguistic accounts of reference in context

In the psycholinguistic literature, as in the pragmatic literature, there is little question about *whether* language users produce and interpret their utterances with sensitivity to the context and other interlocutors. Language production in particular has a long history of observations to this end. Developmental psychologists have long noted that children's speech becomes more oriented to their listeners as they mature (Piaget, 1959; Flavell, 1968), and the phenomenon of child-directed speech or "motherese" is well-established (Snow, 1972). People produce descriptions that appear to cater to the knowledge-state of their audience (Danks, 1970; Fussell & Krauss, 1989a), and they are also able to communicate a message to their intended audience while concealing information from overhearers (H. Clark & Schaefer, 1987). Observations from social psychology have also established that the knowledge-states of others appear to augment many behaviours, including language production for reference (see Krauss & Fussell, 1991, for a review). The central area of debate, then, concerns *how* and *when* language users integrate contextual information in the course of processing, and for comprehension in particular.

An early account of how relevant contextual information is identified and brought to bear on language processing proposed the central role of common ground (H. Clark & Carlson, 1981; H. Clark, 1992). Common ground is a special kind of mutually shared knowledge that is *known to be mutually shared* by interlocutors, and was proposed to constitute the critical background against which utterances are formed and comprehended (H. Clark & Murphy, 1982; H. Clark, Schreuder, & Buttrick, 1983; H. Clark & Wilkes-Gibbs, 1986). As common ground requires that mutual knowledge

is known to be mutually known, it constitutes a form of meta-knowledge; it requires some metapsychology about the mental states of others (in other words, 'theory of mind'). Though the completely accurate calculation of common ground is an infinite, computationally implausible spiral of 'I know that you know that I know...' (Schiffer, 1972), H. Clark and colleagues proposed a number of copresence heuristics by which common ground could be inferred (H. Clark & Marshall, 1981).

Physical copresence concerns items that are perceptually available to both interlocutors, and the fact that the items are perceptually available to both parties *is itself also* perceptually available - for example, an object on a table between two people who can also see each other and both parties can see that both parties can see the object. *Linguistic copresence* refers to things said in a dialogue that then constitute a subset of common ground about the current or previous discourse. Parties to the discourse can then reliably assume this knowledge of each other - for example, pronouns that refer to people established earlier in the conversation. Finally, *Cultural copresence* (also called 'community membership') concerns mutual knowledge that can be inferred on the basis of belonging to the same community.

A common ground account of reference holds that hearers assume utterances are optimally designed for them, and that speakers endeavour to make sure this is the case: "The speaker designs his utterance in such a way 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" (H. Clark et al., 1983, p. 246). If hearers assume this optimal design of the utterance, they are justified in only making recourse to the information that is mutually shared with the speaker in their common ground. It has been argued that common ground is used in the comprehension of conventional expressions, definite reference, and speech acts in general (H. Clark & Marshall, 1981). In fact, the strongest reading of this account suggests that common ground fully constrains the comprehension process: "Our proposal is straightforward: *The intrinsic context for understanding what a speaker means on some occasion is the common ground that the listener believes holds at that moment between the speaker and the listeners he or she*

is speaking to" (H. Clark & Marshall, 1981, p. 258, emphasis in original).

While the ability to assess mutual knowledge with theory of mind is not generally disputed between psycholinguists, the routine use of this ability in processing is more so. Children are able to distinguish between their own beliefs and those of others before the age of three, which is most strongly evidenced by spontaneous references to the emotional and psychological states of others (Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986). Young children are sensitive to the fact that others can lack the knowledge they themselves have (Chandler & Greenspan, 1972; Mossler, Marvin, & Greenberg, 1976; O'Neill, 1996), and that others can hold beliefs that are false (Wimmer & Perner, 1983; Zaitchik, 1991). A meta-analysis of false-belief tasks indicates that this ability emerges after the age of 3 across cultures, settings and tasks (Wellman, Cross, & Watson, 2001). Some have taken the ease with which we seem to employ common ground in conversation as evidence that "use of theory of mind is also rapid... automatic, requiring no effortful attention... and universal." (Stone, Baron-Cohen, & Knight, 1998, p. 640).

A study by Keysar et al. (2000) examined whether this ability is routinely employed in comprehension by manipulating the physical copresence of referents. A referential communication game, where a director gave instructions to participants, centred on an array of objects in a free-standing grid. Some of the compartments of the grid were obscured on one side; the director had no view of what was inside, but the participant did, and this was clearly evident to the participant. In one such trial from this task, the director instructed the participant to "move the small candle one slot down". In the compartments visible to both people were a two-inch candle and a three-inch candle. In a compartment visible only to the participant was a one-inch candle. Although the participants were aware of which items were mutually visible, the participants often considered the one-inch candle as a referent; they sometimes reached for this item, and were delayed in identifying the correct referent (the two-inch candle).

This study of 'perceptual perspective taking' (that is, where the referents were all directly perceptible to the participants) was followed up by a 'conceptual perspective

taking' task (Keysar, Lin, & Barr, 2003): in this study, participants put a roll of tape in a paper bag, so that they knew what was inside the bag but the tape itself was hidden from view. The bag was one of multiple items on a table. A director then joined the task without knowing what was inside the bag, and asked for the "tape". Instead of privileging the mutually-visible video tape, participants attempted to move the bag in a third of all cases, and 71% of all participants attempted to move it at least once. In addition to this off-line outcome measure, eye-tracking data revealed that participants fixated on the bag six times more often when it contained a plausible item like tape compared to a neutral item not indicated by the referring expression. This could not have been perceptual interference; the object itself was not visible. Only the knowledge that it was present could have driven this behaviour. Similar results have been found in tasks that distinguish between common ground and privately available information for pronoun resolution and demonstrative reference in conversation (Keysar, Barr, Balin, & Paek, 1998). These results support an account of comprehension that assigns common ground to a more peripheral role than the optimal design model proposed by H. Clark and colleagues.

The Perspective Adjustment model of comprehension proposes that common ground plays a more peripheral, corrective role in linguistic processing (Keysar et al., 1998). On this view, egocentric linguistic associations provide a means for the fast retrieval of possible interpretations. Common ground assessment is proposed as a slower process that requires more effortful attention to access more complex metapsychological structures (Lin, Keysar, & Epley, 2010). Perspective adjustment suggests that both processes operate continuously in cascades (see McClelland, 1979), but that common ground assessments are rarely able to constrain the early egocentric search for referents precisely because of their differential speeds. In other words, a hearer's knowledge of linguistic conventions provides a fast and economical heuristic for comprehending the intended meaning of an utterance, which may be modified by slower metapsychological processes. Neurolinguistic data on the various neural substrates of conversational interaction supports this view, showing that neural circuits associated with

episodic memory and language are recruited in anticipation of upcoming referring expressions, while circuits associated with interpersonal mentalizing are recruited 'on-demand' when those anticipations are not successful (Bögels, Barr, Garrod, & Kessler, 2015).

Barr (2008) disambiguates two subtly distinct positions that may be conflated in the Perspective Adjustment account: *strategic egocentrism* and *automatic activation* (also called the anticipation-integration model). On the former view, egocentrism constitutes an adaptive strategy for reducing ambiguity (Barr & Keysar, 2005). The use of one's own associative representations in interpretation is a good default process, on this view, because it is less cognitively costly than accessing the common ground, and the representations of individuals within a given language community are sufficiently similar. Strategic egocentrism allows that common ground can override this default when circumstances require it, but that default reliance on one's own linguistic associations reaps a processing benefit.

Automatic activation, on the other hand, is a dual-process model that distinguishes between different types of cognition: automatic, associative processing, and more controlled, rule-based processes. On the automatic activation view, some low-level language processes may operate in an automatic way that is insensitive to common ground, while higher-level processes may take common ground into account in a more controlled way. The deployment of these parallel processes may produce interference or errors of common ground that arise as simultaneous contradictory beliefs (Sloman, 1996). This can be understood by analogy to optical illusions, where the objective reality that is known by the viewer is incongruent with their perceptual experience. Hence, common ground information can be *anticipated* but may not be *integrated* on this view. In short, the automatic activation view treats a limited sensitivity to common ground as an artifact of modular parallel processing. By contrast, the perspective-adjustment model proposes that the routine use of linguistic knowledge is an adaptive strategy that confers some benefit over other methods of interpretation, rather than simply being an incidental feature of processing.

Support for automatic activation comes from a series of experiments by Barr (2008), where participants received pre-recorded instructions that they were told were from a speaker in another room. Before the critical referring expression of their instructions, participants were much more likely to consider candidate referents in the common ground. However, when the onset of the referring expression was compatible with a competitor in the participants' privileged ground, the participants were just as likely to fixate on it. In other words, participants' early fixations on competitor referents were not guided by the shared or privileged status of those objects. This effect also held in an explicitly non-interactive version of the task, and in a further task where instructions were heard from two different pre-recorded speakers, each with independently shared ground with the participant. However, this result is also compatible with the perspective-adjustment view that linguistic comprehension first begins with the linguistic utterance, which is later reconciled with relevant contextual inference regardless of any expectations before the utterance occurs. The lack of integration in the automatic activation account is also markedly disanalogous with optical illusions in that participants (and language users in general) eventually *do* resolve such ambiguities according to common ground knowledge in order to perform the task, whereas unintegrated perceptual systems remain so. Contextual and common ground considerations demonstrably affect the behaviour of language users, rather than remaining a separate domain from language use.

A competing psycholinguistic account of common ground in comprehension called the Constraint-Based model takes this observation as a key starting point. The Constraint-Based approach contends that common ground and the perspective of interlocutors is one of many partial constraints on linguistic comprehension (Brown-Schmidt & Hanna, 2011). On this view, referential ambiguity is resolved by the combination of multiple constraints that influence the hearer's arrival at an interpretation. The strength of the evidence a particular constraint provides relative to other constraints determines the extent to which it will bias an interpretation. The effect of common ground can therefore vary according to the goals and context of the discourse (Brown-Schmidt &

Hanna, 2011). Thus, common ground is incrementally acquired through interactive grounding, and this process contributes to the strength of a common ground item's constraint on processing (Metzing & Brennan, 2003).

An experiment by Hanna et al (2003) demonstrated effects compatible with a constraint-based account of comprehension. Participants were asked to move shapes around a virtual environment by a live confederate speaker, in a task somewhat comparable to Keysar et al's (2000) free standing grid. Instructions such as "Put the blue triangle on the red one" were given in visual contexts that contained two red triangles and one blue triangle. When both red triangles appeared in the common ground, participants fixated on both equally. When one of the red triangles was in privileged ground it did interfere, but participants fixated more on the common ground referent in the initial stages. The privileged ground item was significantly less likely to produce interference when it was yellow. Early fixation on common ground items is incompatible with the strong version of the perspective-adjustment model, where perspective information does not appear at all in the intial stages of processing. The constraint-based account allows for this kind of lexical competition effect, as it predicts that common ground will provide one of many simultaneous (and sometimes competing) early constraints on the search for referents (Nadig & Sedivy, 2002; Hanna & Tanenhaus, 2004; Novick, Thompson-Schill, & Trueswell, 2008; Snedeker & Yuan, 2008).

Similar early effects of common ground on reference resolution have also been found when participants are following instructions with lexical ambiguities that are only temporary, and resolved linguistically by the end of the phrase (Heller, Grodner, & Tanenhaus, 2008; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Brown-Schmidt, 2009b). Other studies have found that the eye-gaze of discourse partners confers early effects on comprehension (Hanna & Brennan, 2007). Such results support the view that common ground is routinely used as one of multiple probabilistic constraints on comprehension. This raises an interesting conceptual point about the degree to which linguistic interpretation can be pragmatically determined; if the linguistic expression is considered against the viable referents in the environment, then this would appear

to require at least some linguistic comprehension to occur *before* contextual constraint takes place. The gradient nature of local, contextual and common ground constraints is held to be predictable according to which source provides the strongest evidence (i.e. is most salient and relevant) for the given task; if linguistic constraints favour a particular interpretation very strongly, then contextual effects will be weaker, and vice versa (Hanna et al., 2003; Brown-Schmidt, 2012).

A key difference between the constraint-based and perspective-adjustment accounts of contextual integration lies in their conception of comprehension processes as either optimal, or adequate. The constraint-based approach represents a model whereby hearers are always calibrated to make use of the parameters most relevant to a given task of comprehension, including their shared knowledge with the speaker: "common ground, or more generally, perspective information, is always potentially available to language processing decisions, but the likelihood an addressee will adopt [it] will *depend on the strength and relevance of the perspective representation*, and whether perspective is in conflict with other constraints" (Brown-Schmidt & Hanna, 2011, p.27, emphasis added). This follows from a tradition that operates from a *principle of optimal design* (H. Clark et al., 1983), an approach that first identifies the components that best meet the criteria of the task, then proposes a means of establishing these components (see also Seidenberg & MacDonald, 1999).

Though this originally conceived a much stronger role for common ground, the assumption that hearers always prioritise the most relevant information follows this same principle of optimal design. In other words, the *quality* of the resources, as defined by the task, is what determines their place in computation. One benefit of this approach is that the gradient, flexible deployment of multiple constraints is able to account for a lot of results in the experimental literature, even if those results appear to be compatible with an egocentric account (Brown-Schmidt & Hanna, 2011). However, this is also a cause for caution. As pointed out by Barr (2008), a constraint-based model is potentially unfalsifiable if the weights on various constraints are left free to vary; egocentric patterns of data can be made to fit a constraint-based account in an ad

hoc way, by positing that various items were not made salient enough for the task.

For the perspective-adjustment model, the assumption that comprehension is *optimally designed* for the task of recovering speaker meaning is not justified (Keysar et al., 1998). The perspective-adjustment conception of linguistic processing may be better-termed an *adequate* design that reaps the right outcome often enough to persist despite any limitations. This model uses an approach that first identifies the resources most readily available to the hearer, and then asks how these can be brought to bear on the task at hand. In other words, the *availability* of resources in a given task is what determines its place in computation. For perspective adjustment, linguistic associations constitute readily available resources, the more stable features of which enable the routine use of egocentric knowledge to retrieve interpretations that are often reliable enough in normal conversation (Barr & Keysar, 2005). For Barr and Keysar (2005), the availability of feedback and clarification in dialogue compensates for any systematic errors that may occur as a result of egocentric processing, and the stable shared inventory afforded by the language community as a whole provides a reliable overlap of users' representations. Interlocutors' internal representations can also become more aligned by virtue of low-level processes during interaction (Pickering & Garrod, 2004), which may bolster those representations as a less-costly source of default interpretation. In other words, "these tricks work because they are specifically designed to exploit the structure of the [linguistic] environment" (Barr & Keysar, 2005, p. 24).

The constraint-based model's primary focus on task requirements can be thought of as a 'top-down' approach to comprehension, and the perspective-adjustment's focus on resource availability may be contrasted as 'bottom-up'. My characterisation here invokes the three-level hierarchy of explanation set forth by Marr in his work on vision (Marr, 1982), which has been widely adopted in cognitive science. The 'top' level of *computational* description refers to why a system does what it does; that is to say, it defines the task or problem that the system solves. The next level down is *algorithmic* or *implementational*, which concerns how this function is carried out and what kind of representations/resources are employed in order to do so. The *physical* level of

description refers to the physical instantiation of the system.

Questions asked at each level logically entail different sets of possible answers. Computational cognitivism contends that top-level considerations necessarily constrain the particulars at lower levels (Marr & Nishihara, 1978; Marr, 1982; Pylyshyn, 1984; Gallistel, 1990). According to this dominant paradigm, which Piccinini (2006) terms "computational chauvinism", features at the implementational level can only be identified and described in terms of the computational level; in other words, the task determines the necessary units of computation. This approach is not without challenge, however, and a top-level description of the task may lead to conceptions of lower levels that do not hold up. One prominent example is the 'physical symbol systems' that were predicted by early computational theorists (Newell, 1980) to exist in the nervous system, and appear to be increasingly unlikely (Fodor & Pylyshyn, 1988; Harnad, 1992).

It has been argued that computational level descriptions need to be at least partially responsive to the limitations and specifications of lower levels in order to provide plausible guidance. Limits on top-level theorising tend to mean that all internal states must correspond with (or represent) external, observable (or worse, abstractly conceived) properties, and so accounts guided only by these considerations may be blind to components that pertain only to the internal system (Poldrack, 2006; Egan & Matthews, 2006). In other words, that a system *looks from the outside* like it is performing one way does not necessarily mean that such a characterisation will be represented in the mechanism. In light of concerns like these, others in cognitive science have adopted a more dynamic approach that is grounded in requirements at multiple levels (Churchland, 1986; Keeley, 2000; Feest, 2003; Egan & Matthews, 2006; Poldrack, 2006).

Note here that my use of 'bottom up' to describe such an approach is perhaps something of a misnomer; it does not entail using lower levels as a means by which to derive task-level descriptions. Such an approach would indeed be, as Marr put it, "like trying to understand bird flight by studying only feathers" (Marr, 1982, p.27), and represents a project that has long since failed to deliver (Barlow, 1972). The use of 'bottom-up'

here is intended only to contrast with the top-down, unidirectional constraint of computational cognitivism, and to represent an approach that takes constraints at the level of implementation as part of its starting point in addressing how the top-level task can be solved. The top-down model of cognition starts by asking: "what does this task require of the cognitive system?" which is likely to yield an *optimal* answer. The bottom-up cognitive model starts with: "what [implementational/physical] resources are most available to the system [for this task]?", which is more likely to yield an *adequate* answer. The identification of a computational task is integral to both questions; the difference is that the range of possible answers is determined by different standpoints. For the former question, the scope of possible answers is bounded by the requirements of the task. In the 'bottom-up' case, the scope of possible answers is bounded by the capabilities of the system.

Though psycholinguistic models are closely tied to empirical data, it is interesting to note that the 'top-down' and 'bottom-up' approaches of constraint-based comprehension and perspective-adjustment, respectively, correspond with the approaches taken in the broader theoretical debate outlined in §2.1. The constraint-based approach contends that the resources used in computation are defined by the parameters of the task. This model considers the role of shared knowledge as one of many parallel parameters that have "immediate and probabilistic effects on interpretation, depending upon the strength and salience of the speaker's perspective, and its relevance to the addressee" (Hanna et al., 2003, p. 60). This view holds that the computed strength of these parameters must be continually updated because "the relevancy of constraints changes moment to moment" (Hanna et al., 2003, p. 60), and so the calibration of the optimal cognitive system changes accordingly. In fact, the task requirements directly predict what the cognitive system does, such that "in conversational interactions in which the participants have behavioral goals, the state of the context must be based upon the speakers' and addressees' intentions and actions" (Hanna et al., 2003, p. 60). For the language-as-action tradition from which the constraint-based approach is derived, all conversations entail behavioural goals (H. Clark, 1992; Tanenhaus, Cham-

bers, Hanna, & Hall, 2004), and so at any given time in conversation the addressees must recruit intention-based units in service of the computational task.

This approach mirrors that of the relevance-theoretic framework outlined in §2.1. While relevance theorists acknowledge that cognition may not always successfully maximise relevance (Sperber et al., 1995, p.50), the approach remains motivated by the premise that it is always geared toward doing so, and that the *requirements of the task* are the primary determinants of which cognitive representations are employed in processing. Hence, a person's expectations of relevance adjust in order to best meet the demands of different tasks moment-to-moment. More fundamentally to relevance theory's account of comprehension, the optimal solution is specified at the level of implementation. On this view, the mechanisms involved in the task of interpreting a speaker's intended meaning "are seen as rational processes geared to the recognition of speakers' intentions" (Sperber & Wilson, 2002, p. 4). In other words, the best quality resource for the task of retrieving intended meaning (that is, metapsychological inference of intended meaning) is afforded the status of an obligatory component in processing, in a 'top-down' fashion.

Recall that this is in contrast with the neo-Gricean approach, which takes the stable semantic content of linguistic utterances to be a useful resource that can contribute to the task of interpreting a speaker's meaning. When presented with an utterance, the Gricean hearer must make recourse to the utterance (that is, the linguistic structures that have been made available). The hearer can supplement this with whatever indirect information may also be available. As laid out by Grice: "the hearer will reply on the following data: (1) the conventional meaning of the words used [...] (2) the cooperative principle, (3) the context, linguistic or otherwise, of the utterance, (4) other items of background knowledge" (Grice, 1975, p. 50). Though the cooperative maxims have been revised considerably, this basic schema persists. One way of paraphrasing this schema is: given that (i) you must interpret the speaker's intended meaning, (ii) they have provided you with an utterance, and (iii) other sources of information may be variable or less direct, it is of primary importance to attend to the content of the

utterance.

This view treats the stable associations (i.e. the semantic content) of words used in utterances as a resource that is always present in cases where one is presented with an utterance, and so considers this content a sensible basis for interpretation. In this way, the Gricean reliance on stable linguistic information mirrors the 'bottom-up' approach of egocentric perspective-adjustment. For perspective-adjustment, low-level alignment processes and the similarity of associations across speakers of the same language mean that the hearer's linguistic knowledge is a useful and readily available resource for the task of dialogue (Barr & Keysar, 2005). The same authors also raise concerns similar to those raised about computational cognitivism in neuroscience, regarding the possible differences between computational appearance and implementational reality: "the mere observation that a speaker produces an utterance that is in alignment with mutual knowledge does not warrant the inference that he or she computed that knowledge as mutual at any time" (Barr & Keysar, 2005, p. 38). The perspective-adjustment view that linguistic knowledge is contentful enough to support the interpretation of utterances has a natural affinity with the neo-Gricean view, whereby interpretation is especially informed by the linguistic utterance itself. The consideration of linguistic knowledge as a useful and readily available resource proceeds from a 'bottom up' approach to utterance comprehension.

To sum up, proposed strategies for the interpretation of a speaker's intended meaning differ in consistent ways, according to the theoretical approach to the question. A top-down approach presumes an optimal solution to the computational task of interpreting a speaker's intended meaning. On this view, the constraint of multiple relevant parameters is brought to bear on interpretation, including metapsychological inference about the speaker and the common ground they share with the speaker. A more bottom-up approach considers the most stable and easily accessible resources available to the comprehension system for the task of interpreting a speaker's intended meaning. These will always include the utterance itself, and the linguistic conventions known to the hearer constitute an *adequate* resource for interpretation, given the similarity

of representations across speaker communities as well as the interlocutors of a given discourse. The linguistic input is thus a special source of information for the task of interpreting utterances on this view.

While the psycholinguistic literature is primarily concerned with the time-course of processing in comprehension, the relative status of linguistic and contextual input that results from each account is notable in itself, and implies different evolutionary scenarios. The following section observes a similar theoretical contrast between approaches in the evolutionary literature, integrates each account of linguistic comprehension with the adaptationist particulars it may imply, and derives hypotheses from these two broad camps.

2.3 Adaptationist psycholinguistics of comprehension

This thesis takes the position that evolution by natural selection is the only scientific explanation for the appearance of functional design in living organisms (Dawkins, 1983, 1986), including human cognitive architecture (Cosmides & Tooby, 1987). As such, the current exploration does not ask whether linguistic cognition has been produced by natural selection *or* learning; however, there are proposals elsewhere in the evolutionary literature that learning may constitute an alternative to adaptation by natural selection (Buller, 2005), and that proximate processes challenge the Modern Synthesis more broadly (see Laland et al., 2014, for discussion). It has also been proposed elsewhere in the language evolution literature that the use of domain-general abilities for language challenges the idea of innate specification (Culbertson & Kirby, 2015; Chater, Clark, Goldsmith, & Perfors, 2015). Most studies of this type are explicitly concerned with the idea that particular the linguistic structures we produce, such as syntactic form, result from innate cognitive architecture (Christiansen & Chater, 2008; N. Evans & Levinson, 2009; V. Evans, 2014) - i.e. the question concerns where the

form of languages comes from. The dichotomy of "nature versus nurture" or "learned versus innate" is a popular and useful framing for investigations where the language structure itself is the object of study. I depart from this traditional focus on the origins of linguistic structure to instead focus on psycholinguistic processing. Adaptationism is compatible with the presence of learning, and learning and processing abilities are themselves subject to natural selection (Cosmides & Tooby, 1987; Symons, 1990; Barrett & Kurzban, 2006; Barrett, Frankenhuis, & Wilke, 2008; Frankenhuis & Barrett, 2013; Panchanathan & Frankenhuis, 2016). This analysis therefore does not use the oppositional framing that asks whether cognition for linguistic comprehension has been produced by natural selection *or* processing. The aim is to identify which features of processing would be consistent with different evolutionary scenarios. Thus, the question is: what processing features for linguistic comprehension could natural selection have produced?

One common approach to the evolutionary explanation of a given trait is to identify its function; in other words, to identify the problem that the trait is a solution to. This follows from the adaptationist program popularised by Ernst Mayr, which promotes one key research question in particular: "The adaptationist question *'What is the function of a given structure or organ?'* has been for centuries the basis of every advance in physiology. If it had not been for the adaptationist program, we probably would still not yet know the functions of the thymus, spleen, pituitary, and pineal "(Mayr, 1983, p. 328, emphasis added). It is this question that led Pinker and Bloom, in a landmark paper that is widely held to have galvinised the evolutionary study of language, to identify that language use is a solution to the task of communication (Pinker & Bloom, 1990). Identifying the function of language as *communication* is a way to generate further investigation of how, exactly, language might have arisen to facilitate this.

This is reminiscent of the 'top-down' model of cognitive science outlined in §2.2. The corresponding adaptationist approach encounters similar challenges to computational cognitivism when this route is adopted, because establishing an objective criterion for function itself is an ongoing project (Williams, 2008). This has led to critiques

that emphasise the consideration of other lower-level constraints that can account for the form of a given trait (Gould & Lewontin, 1979; Pigliucci & Müller, 2010; Futuyma, 2013). In particular, the top-down motivating question 'what is the function of this trait?' may only allow answers of the form *'the function of T is x/y/z...'*. In order to take lower order factors into account, a more bottom-up question like 'what can account for the form and distribution of this trait?' (Lloyd, 2015) may shed more light. The distribution of linguistic processing is such that language use is a species-typical trait in humans that develops reliably, and without any instruction, across all cultures; failure to perform this function is commonly approached as a disorder of development that requires intervention (L. Bloom & Lahey, 1978). Given the adaptive value of communication, adaptation is a strong first candidate for explanation (cf. saltational accounts as in Chomsky & McGilvray, 2012).

However, the current point of debate concerns differing accounts of *how* hearers comprehend linguistic utterances in order to retrieve a speaker's meaning. That is to say, the task this system solves (the function) is common to both accounts. Simply asking what the function of comprehension processing is, then, cannot help frame further investigation into how this is achieved. Rather, it is necessary to establish the *form* this processing takes. Establishing what form the mechanism takes can then lend itself to an explanation of why it looks the way it does. A dominant focus of the language evolution literature concerns the form and distribution of *language structure*, which is perhaps why psycholinguistics has not, historically, been a method of investigation in much of the evolutionary linguistics literature (see Scott-Phillips, 2010a, for a brief discussion).

The following section proposes that, if the adaptive task is common to evolutionary accounts of language processing, using a 'top-down' adaptationist approach for theoretical insight into *how* this task is solved provides little disambiguation. Psycholinguistic investigation, on the other hand, is a useful source of evidence.

2.3.1 Psycholinguistic models in an adaptationist framework

The features of psycholinguistic processing, in general and for the particular topic of contextual integration, lend themselves well to adaptationist inquiry. This requires further elaboration of what an adaptation is. Natural selection coordinates (1) a system of reliably developing properties in an organism with (2) a set of structural properties outside of the adaptation that recur across generations (such as the environment, or other features of the organism), so that (3) the interaction of both produces a functional outcome that solves a problem for the organism, which ultimately contributes to the organism's reproduction (Tooby & Cosmides, 1990). Put another way: *"to function, adaptations are selected to assume the presence of, to rely on, and to exploit stable and enduring structural and statistical regularities, both in the environment and in other aspects of the phenotype"* (Tooby & Cosmides, 1990, p. 761). Note here that, though "innate vs learned" is a popular framing of evolutionary debates, adaptationism is compatible with the presence of learning (Cosmides & Tooby, 1987; Barrett et al., 2008; Frankenhuis & Barrett, 2013). Many adaptations are indeed best characterised as learning capabilities. So-called 'facultative modules' assume and depend on external input in order to solve a functional problem, and such learning abilities are themselves subject to natural selection. Thus, the adaptationist question in the case of language concerns what form of linguistic processing may have resulted from natural selection.

Both accounts of linguistic comprehension outlined in this chapter can be clarified using this adaptationist rubric. The context-driven account of linguistic processing proposes that (1) a hearer's ability to identify and integrate ostensive linguistic stimuli with multiple relevant sources of information, including metapsychological knowledge about the speaker, is coordinated with (2) the attendant context and goals of the discourse, such that (3) the hearer arrives at an optimal interpretation of the speaker's meaning. It is important to note that the principal importance of linguistic input to the inferential process here is that it is an *ostensive* stimulus. Ostensive stimuli may be linguistic or nonlinguistic, and constitute inputs to the inferential process alongside

other parameters of the context. Ostensive-inferential comprehension is fundamentally socio-cognitive, and targeted on the mental states of others. Ostensive stimuli, in this process, provide some evidence for the speaker's mental state, rather than any other kind of representation. The integration of linguistic information is thus only of use as an input to this metapsychological process of inference, and the course of this processing is determined by the context. The context-driven account, under this rubric, suggests that successful linguistic comprehension is achieved by fundamentally inferential means.

This can be contrasted with the more egocentric language-driven account of comprehension, in which linguistic input can be subject to a more specialised access in order to retrieve the speaker's meaning. This view proposes that (1) the retrieval of the hearer's own linguistic associations is coordinated with (2) the linguistic associations of others by the linguistic environment they share, as well as with the feedback of the discourse, and the integration of contextual information, so that (3) the hearer arrives at an interpretation of the speaker's meaning with less cognitive expenditure. The crucial difference here is that the ability to exploit linguistic conventions is the pertinent set of processing properties selected for in the hearer. This set of properties presumes contextual and metapsychological information (i.e. it is coordinated with *other aspects of the phenotype*), as well as shared regularities in linguistic associations and in the discourse environment itself. This set of differences is delineated in table 2.2 (p.49).
The interpretation of a speaker's intended meaning allows the successful navigation of the social environment. The means by which this is achieved differs between these adaptationist accounts, according to which salient parameters of the environment are assumed by the comprehension process. The context-driven comprehension system integrates ostensive stimuli (including utterances) with contextual information in order to make metapsychological inferences about the speaker's meaning. The language-driven comprehension system exploits the hearer's own linguistic representations as a reasonable proxy for the speaker's meaning, at least some of the time. This relies on the linguistic environment which emerged subsequent to (and as a consequence of)

	(1) Properties in the organism	(2) Structural properties outside of the adaptation	(3) Functional outcome produced by the coordination of 1 & 2
Context-Driven Comprehension	· *Ability to integrate ostensive linguistic stimuli with relevant context***	· *Relevant contextual information & discourse goals* · *Central inferential reasoning*	· *Optimally relevant interpretation of the speaker's meaning*
Language-Driven Comprehension	· *Ability to retrieve conventional/ associative linguistic representations*	· *Linguistic associations of others* · *Ability to integrate ostensive stimuli, including linguistic utterances, with relevant context***	· *Cognitively efficient interpretation of the speaker's meaning*

Table 2.2: The similarities and differences of Context-Driven Comprehension and Language-Driven Comprehension in an adaptationist framework: for CDC, the task of interpreting speaker meaning selects for the hearer's ability to identify and integrate the ostensive stimuli with multiple relevant sources of information, which relies on the goals of the discourse and inferential processing; for LDC, the task of interpreting speaker meaning selects for the hearer's ability to exploit their own linguistic representations, and this ability relies on the associations held by others, the feedback of discourse, and the attendant relevant context. ** note here that these refer to the same ability; in other words, Language-Driven comprehension presumes that context-driven inferential processing of ostensive stimuli already exists, and specialised exploitation of linguistic representations is coordinated with this ability.

context-driven, inferential communication. Here, the linguistic environment coordinates the linguistic representations of the population to an extent that is reliable and robust enough to make this strategy viable.

For both accounts, the task of successfully interpreting a speaker's meaning - and the task of communication in general - is ultimately driven by a pressure for successful social navigation. The difference lies in which aspects of the environment are exploited by the comprehension system in order to achieve this. Context-driven comprehension is principally concerned with the mental states of conspecifics that make up the social environment; Language-driven comprehension is principally concerned with the linguistic associations made available by the cultural environment that coordinate the representations of the population.

Context-driven comprehension is consistent with an evolutionary account that positions linguistic comprehension as a social adaptation (i.e. an adaptation to inference-using conspecifics) (Scott-Phillips, 2014). This Social Adaptation account draws on the hypothesis that primate intelligence is a specifically social adaptation which evolved in response to the demands of living in ever-increasing group sizes (Dunbar, 1998a, 2003). The *social brain hypothesis* observes that primate brains are large, relative to their body size, and suggests that this is a consequence of relying on large, interdependent groups of conspecifics. An important requirement of the social environment is the ability to keep track of the social network, reason about the intentions and motives of conspecifics, and to behave in a way that maintains one's own relationships to ensure group cohesion. Tasks like these require fairly sophisticated cognitive abilities of the kind exhibited by primates, and so the social brain hypothesis is a dominant explanation of primate intelligence (Byrne & Whiten, 1989; Brothers, 2002; Dunbar, 2003). This is supported by studies revealing that relative brain size in primates correlates with many measures of social complexity, such as group size, grooming clique size, the prevalence of social play and tactical deception, and the frequency of social learning (see Dunbar & Shultz, 2007; Dunbar, 2012, for reviews). This also extends to humans, who live in very large and complex groups in comparison to other primates,

and have particularly advanced social cognition (Dunbar, 2003; Frith & Frith, 2010; Whiten & Erdal, 2012). There is a large body of evidence detailing the advanced social cognition of humans, which includes the ability to identify false and absent beliefs in others (Chandler & Greenspan, 1972; Wimmer & Perner, 1983), to track the beliefs of others automatically in the course of our everyday awareness (Kovács, Téglás, & Endress, 2010; van der Wel, Sebanz, & Knoblich, 2014), to show a greater sensitivity to cues of being observed by others than is exhibited by chimpanzees (Nettle, Cronin, & Bateson, 2013), and to represent the mental states of others recursively (O'Grady, Kliesch, Smith, & Scott-Phillips, 2015). Human children, while comparable to chimpanzees in terms of physical cognitive skills like tool use, are significantly more sophisticated than chimpanzees in terms of their social cognition (Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007). The incentive and ability to track the mental states of others in order to navigate the social environment is suggested to have set the stage for the emergence of ostensive-inferential communication (Sperber, 2000; Scott-Phillips, 2014). The example outlined in §2.1 noted that Mary and Peter were able to communicate about the edibility of berries with ostensive, stylised behaviour, precisely because they possessed knowledge about each others' intentions. The original example by Sperber (2000) set out incrementally accumulated layers of metapsychology that may illustrate how the ability to communicate with ostension and inference may have emerged in the course of human evolution within the social cognitive niche. As concisely summarised by Scott-Phillips:

"In scenario one, Mary is picking and eating berries, simply because the berries are edible. In scenario two, Peter watches her do this, and so forms a belief that the berries are edible. In scenario three, Mary wants Peter to believe that the berries are edible, and so she picks them precisely because this will cause Peter to believe that the berries are edible. In scenario four, Peter is aware that this is what Mary is doing. In all the scenarios up to here, Mary is picking berries in the same, standard way - she has no reason not to - but in scenarios five and six an important change occurs. In five, Mary wants Peter to believe that she wants him to believe that the berries are edible. In order

to satisfy this intention, she picks the berries not as normal, but in a somewhat stylised or exaggerated way. [...] In six, Peter correctly interprets this evidence, and consequently recognises Mary's intention. This (and only this) is ostensive communication" (Scott-Phillips, 2014, p. 101)

This example demonstrates how complex metapsychology may be incrementally acquired with increasing layers of behaviours that begin as first-order attempts to manipulate attention and reason about the mental states of conspecifics. Scott-Phillips (2014) argues that linguistic conventions are made possible by this metapsychology, and so these inferential abilities must logically precede linguistic conventions in the emergence of human communication. In other words, the emergence of linguistic conventions served to make inferential, relevance-oriented social cognition more efficient and powerful. Language, on this view, is preceded by pragmatic cognition in all cases, and so provides a benefit to the species only in so far as it relates to improving the performance of inferential ability. For the social adaptation account, language is important for comprehension primarily because it is ostensive, and therefore relevant enough to attend to along with other relevant contextual factors of the interaction. This closely resembles the specifications of Context-Driven comprehension detailed in table 2.2, whereby the pressure for successful comprehension selects for the hearer's ability to identify and integrate multiple relevant sources of information, including metapsychological knowledge about the discourse partner. For the Social Adapation view, then, the cognition that performs our linguistic comprehension is an adaptation to the social environment.

A language-first, inference-second account of language emergence has been argued to be impossible (Sperber et al., 2010). Ontogenetically, there is a considerable body of evidence that suggests theory of mind abilities enable many crucial aspects of language acquisition (Baldwin, 1995; Akhtar & Tomasello, 2000; Sabbagh & Baldwin, 2001; P. Bloom, 2002; Golinkoff & Hirsh-Pasek, 2006; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007; Vazquez, Delisle, & Saylor, 2013). The role of metapsychological abilities in making linguistic conventions *possible* is

difficult to dispute, and forms the crux of the claim that language comprehension is fundamentally inferential by design (Scott-Phillips, 2014). On its face, this may seem to render language-specialised comprehension implausible. But language-driven comprehension does not necessarily entail that language emerged before inference in phylogeny, and so the role of metapsychology in making conventions *possible* does not speak to the question of whether subsequent specialisation for language has occurred in humans.

Language-driven comprehension is predicted by an evolutionary account that specifies a linguistic adaptation. As outlined in table 2.2, the language-driven comprehension account presumes that inferential abilities already exist in the phenotype. The role of metapsychology in enabling the emergence of language is not challenged by specification for linguistic conventions; on the contrary, the linguistic adaptation view requires that this is the case. Recall that the psycholinguistic perspective-adjustment model proposes that inference about the context is automatically in operation, as well as automatic access to linguistic associations (Keysar et al., 1998; Barr & Keysar, 2005). This linguistic knowledge is more accessible than inference about the context because it is egocentrically available. If linguistic information is able to resolve an ambiguity, slower inferential process need not be brought to bear on the task. If linguistic information does not resolve the task in time, inference may contribute relevant information that can ensure a successful interpretation. This account does not oppose inference's role in making linguistic conventions possible in acquisition; it proposes that, once these linguistic conventions are available to the individual, they constitute a valuable means of interpreting the utterances of others.

A similar view can be taken on the emergence of linguistic specialisation; once linguistic conventions have become available to the population and constitute a stable part of the cultural environment, they bring with them a new set of parameters that it becomes adaptive to exploit. Human evolution has long taken place against a backdrop of previous human activity that exerts selective pressures (Kingdon, 1993). A canonical example is the stable, cross-generational domestication of cattle and associ-

ated dairying activities, which may have selected for greater lactose tolerance in some human populations (Aoki, 1986; Feldman & Cavalli-Sforza, 1985; Durham, 1991). In a similar way, the stable system of culturally transmitted linguistic conventions that constitutes the linguistic environment, which emerged subsequently to and as a consequence of prior inferential abilities, may have provided a setting for the emergence of a language-first comprehension strategy in humans. The convergent behaviour of a language-group - that is, the similarity of representations held by different individuals that result from adopting the same conventions - means that an individual's linguistic inventory may become a reasonable proxy for that of another speaker. If this proxy proves reliable enough, then interlocutors stand to gain from the increased cognitive efficiency of exploiting their egocentric knowledge. The similarity of semantic representations among language community members is a necessary component of the perspective-adjustment model (Barr & Keysar, 2005), as it is this environment that gives egocentric associations their value in interpreting the utterances of others. This cognitive model and its external dependencies are predicted by a Linguistic Adaptation view, on which the cognition that performs our linguistic comprehension is an adaptation to the linguistic environment.

The Linguistic Adaptation and Social Adaptation hypotheses of comprehension predict different forms of linguistic processing. Standard evolutionary psychology suggests that adaptations in cognitive architecture must be organised, as other biological systems are, in a modular way, comprising multiple dissociable subparts (Simon, 1991; Barrett & Kurzban, 2006). Broadly, this means that subparts of cognitive architecture are particularised to a certain range of tasks in order to make the given tasks tractable. This is because the system must define the necessary processing rules for each scenario in order to be sensitive to the context (and such context-sensitivity is, presumably, the motivation behind proposing domain-general functions). Given this, the more variable the input and task procedures a given system allows, the more complex its algorithms must be. Symons (1990) likened the proposal of generalised 'deep structure rules' that manifest across multiple domains, such as *"maximise the number*

of offspring you rear to maturity", to a computer chess program consisting only of the rule ''win"; a generalised description across distinct situations cannot constitute a rule of operation, because it fails to account for the particular, on-the-ground means by which the adaptive outcome is actually achieved. Such a mode of operation would require an arbitrarily complex algorithm. In order to be tractable, then, there must be limits placed upon the functions of a given algorithm (Fodor, 1983, 2001). Though there is nothing that requires content-specificity in the evolutionary psychology approach, many task-specific modules are also content-specific in nature (for example, a folk-psychology module is targeted on mental states, while a folk-physics module is about the movement of objects) (Carruthers, 2006; Sperber & Hirschfeld, 2004).

We might expect, then, that sensitivity to a particular type of input indicates the specialisation of a system. The sensitivity of comprehension processing to conventional linguistic information and contextual information, respectively, can provide evidence about which aspects of the cultural environment our linguistic processing is adapted to. If the emergence of language in humans changed the fitness landscape such that comprehension output processes became *specialised* for linguistic communication, we could expect the comprehension system to be particularly sensitive to linguistic utterances as a special domain of input. If the emergence of language simply contributed further to the efficiency of a socially-adapted inferential system, we may expect that comprehension is specialised to attend to the most relevant contextual parameters, including the knowledge states of speakers.

In short, we are presented with two different accounts of how linguistic comprehension has evolved. Establishing evidence for one account over another is important for our evolutionary understanding, as they each concern different aspects of the cultural environment. Both accounts address the adaptationist task of interpreting a speaker's intended meaning, and so further clarification of this task in a 'top-down' fashion does not provide contrastive differences between them. One way to resolve this is a 'bottom-up' line of inquiry that establishes the material *form* that this trait (comprehension processing) takes. This approach requires psycholinguistic inquiry.

2.3.2 Hypotheses

In this section, I will consider some conceptual issues that arise from deriving hypotheses from the SAH and LAH, before presenting the pairs of contrastive hypotheses addressed in the remainder of the thesis.

For the LAH model, linguistic meaning is a special domain of input to the comprehension system, relative to other inputs of contextual information. Testing this raises a problem: without a quantifiable value that can be applied to both sets of input, a comparison of linguistic and contextual effects is difficult to make. This is a recurrent problem for the empirical investigation of comprehension models that incorporate different domains of information. Psycholinguistic constraint-based models do not give task-independent measures of salience or constraint from different sources (such as referential context vs lexical constraint) (Brown-Schmidt & Hanna, 2011). Constraint-based interpretations can thus be given to results that are predicted by rival models of processing, and models where constraint weights are left free to vary may be unfalsifiable (Barr, 2008). This issue is of particular concern for the investigation of Relevance Theoretic comprehension, where relevance is characterised as a trade-off between cognitive effects and cognitive effort. Relevance "can be achieved in any conceptual domain" (Sperber et al., 1995, p.36), but there is no method for quantifying either cognitive effects or cognitive effort for a given domain of operation, and an absolute measure may be impossible (Wilson & Sperber, 2008).

There is no straightforward measure of contextual and linguistic effects on comprehension to use for comparison. However, we ought to expect that comprehension outcomes can be informative about which domains of information the process has relied on. For Sperber and Girotto (1995), the empirical investigation of relevance effects required manipulation of the type of information made available to the comprehension process such that multiple performances of the task could be compared to each other, if not to an absolute standard. In a similar fashion, the following experiments manipulate the extent to which various grades of contextual information and linguistic input

are available for completion of the experimental task. In short, the manipulation of available inputs is predicted to elicit differences in the interpretations that hearers ultimately reach, and how long it takes to reach them. By conducting such tasks across several domains of contextual information, a broad indication of the comprehension system's processing sensitivities to them - at least in comparison to linguistic stimuli - may emerge. Indeed, the existing empirical literature indicates that an appraisal of comprehension performances across a range of conditions may, at present, be the only way to establish such a picture.

Another feature of the LAH model is that the representations associated with linguistic utterances are a possible candidate output for the comprehension process. Recall that, for the relevance-oriented SAH model, the most relevant parameters of the discourse will determine the hearer's interpretation. Supposing that the most relevant contextual parameters happen to be linguistic utterances, the predicted comprehension outcomes of both models are potentially indistinguishable.

However, the different *treatment* of linguistic meaning in each model ought to provide some means of disambiguation. In particular, linguistic meaning in the SAH model is an input to a process of contextual integration, such that even if linguistic utterances were encoded as the most relevant parameter over phylogeny or in a given task, the output interpretation for a given task could never consist of a linguistic representation alone. Under conditions where the linguistic meaning derives an incorrect interpretation and the context provides a correct interpretation according to the goals of the task, only the LAH predicts that incorrect interpretations based on linguistic meaning alone are possible. Experiments 1 and 2 below are of particular importance to this line of evidence, as interpretations based on linguistic labels alone would derive incorrect responses under certain conditions of the task that are incompatible with the SAH.

On the basis of the accounts delineated in §2.3.1, the remainder of the thesis investigates contrastive hypotheses across several contextual domains. If linguistic comprehension is primarily a social adaptation, it should be sensitive to the knowledge states

of others. We may especially expect this to be the case if the knowledge of others is highly relevant to yielding the correct interpretation. If, however, linguistic comprehension entails a specialised linguistic adaptation as in the LAH model, we may expect processing to exhibit much more sensitivity to the linguistic stimulus itself. Experiment 1, presented in chapter 3, tests these hypotheses by using speaker-specific facts about referents as precedent labels in a referring task. The experiment investigated whether participants experience slower response latencies and select different referents when speaker-specific referring expressions were used by different speakers, and linguistic labels were used inconsistently.

If linguistic comprehension is part of a primarily social adaptation, we might expect the contribution of linguistic associations to be comparable to other stable associations that exist in cultural knowledge (or in the cultural copresence as outlined in 2.2), which would affect the interpretation of referring expressions by constraining expectations of relevance. However, if linguistic stimuli are privileged input to the comprehension process, we might expect that the linguistic label itself confers a greater effect on processing relative to other conventional cultural associations, or to speaker-specific information. If comprehension processing considers linguistic assocations to be different from wider stable associations made available in cultural knowledge, this classifies language as a special kind of cultural system from the point of view of human cognition. Experiment 2, presented in chapter 4, tests these hypotheses by using speaker-specific labels for objects with strong conventional assocations (e.g. the Eiffel Tower is associated with Paris). The experiment investigates whether participants experience slower response latencies and yield different interpretations when cultural associations with referents are violated by the referring expressions; when they are delivered by different speakers; and when the linguistic labels themselves are inconsistent.

If linguistic comprehension makes use of the most relevant information, we might expect visual context, which is a more directly perceptual source of relevant information than speaker-specific knowledge-states, to determine the interpretation of an utterance. If linguistic comprehension is an explicitly linguistic adaptation, however,

we might expect linguistic labels themselves to exert a larger effect on comprehension than the accompanying visual context. Experiments 3 and 4, presented in chapter 5, test these hypotheses by using tangrams to change the visual context around a given referent, which has been assigned a referring expression in the task. The experiments investigate whether a change in linguistic label confers a larger delay in comprehension, and a change in referent selection, compared to a change in the visual context.

If comprehension is characterised by particular sensitivities, we might expect speakers to produce linguistic utterances that lend themselves well to the way that hearers process them. If linguistic cognition is fundamentally inferential, we may expect people to produce labels that make recourse to the contextual information available to the hearer. If linguistic cognition is specialised to linguistic conventions, we may expect speakers to produce labels that are consistent, even if they are not a perfect match to the context. Experiments 5 and 6, also presented in chapter 5, test these hypotheses by using tangrams to change the visual context around a given referent that participants must label for others in a task. The experiment tests whether mismatched and abstract arrangements of shapes lead participants to produce different referring expressions compared to when the context remains constant.

2.3.3 Summary

In this chapter I have reviewed a long-standing polarisation regarding the role of linguistic input in comprehension that has manifested in both theoretical pragmatics and empirical psycholinguistics. Similar theoretical methods employed in each case have given rise to a similar point of disagreement. If one uses the requirements of the computational task itself to fully define the mechanism of interpreting a speaker's intended meaning, the mechanism must make recourse to inferred representations of interlocutor knowledge and contextual considerations in the first instance. If one uses considerations about which resources are most readily available to perform this same task, the answer is more likely to derive a mechanism that depends on linguistic conventions.

The stability afforded to the conventional semantic content of language differs accordingly between a context-driven and language-driven view. On the former, the linguistic conventions held by individuals are insufficient to provide stable recourse in the task of comprehension, precisely because the task is to recover intended speaker meaning. On the latter, these linguistic conventions are a fast and reasonable heuristic that can enable more efficient interpretation by virtue of their similarity among speakers. Each proposal of linguistic comprehension implies a different evolutionary account of how human cognition has evolved, and this chapter has presented an adaptationist account of both models. The massively modular nature of adaptation means that we can expect a system to exhibit sensitivity to the particular domain it is adapted to. The evolutionary literature on language has hitherto focussed on the adaptive structure of languages, rather than the psycholinguistic features that different evolutionary scenarios may predict. The debate regarding context integration that has arisen in pragmatics and psycholinguistics has significant import for the evolutionary picture, which stands to gain some clarity from the investigation of these accounts. The 'top down', theoretical adaptationist approach is unable to discriminate between these rival hypotheses, as both mechanisms serve the same function of interpreting a speaker's intended meaning. Rather, a 'bottom up', empirical approach must be used, whereby the *form* of linguistic processing is established in order to provide evidence for the adaptationist picture.

The rest of this thesis will present a series of experiments that test hypotheses derived from these accounts. Four contrastive predictions made by the Social Adaptation and Linguistic Adaptation accounts of comprehension are tested by the experiments presented in Chapters 3-5. These are:

- SAH: Referring expressions that depend on speaker-specificity will produce a slower response latency when used by a different speaker, *vs* LAH: Referring expressions that depend on speaker-specificity will produce a slower response latency when the linguistic label itself is not used consistently. This was investi-

gated in Experiment 1, Chapter 3, which found results that were consistent with the LAH.

- SAH: Strong associations in cultural knowledge will contribute to linguistic interpretation and produce a slower response latency when that knowledge is in conflict with the linguistic label; cultural knowledge will be deployed with sensitivity to knowledge state of the speaker *vs* LAH: Referring expressions will produce a slower response latency when the linguistic label itself is not used consistently, regardless of cultural associations or speaker specificity. This was investigated in Experiment 2, Chapter 4, which found results that were consistent with the LAH.

- SAH: The most relevant contextual information will determine linguistic interpretation, and incongruent contexts will produce the slowest latency in comprehension, *vs* LAH: Linguistic labels are a privileged input for comprehension, and inconsistent label use will produce the slowest response latencies. This was investigated in Experiments 3 and 4, Chapter 5. The results were mixed, but broadly consistent with the SAH.

- SAH: Speakers may produce referring expressions that make most recourse to the hearer's visual context *vs* LAH: Speakers may produce referring expressions that are consistent with their past productions, rather than with the visual context of the hearer. This was investigated in Experiments 5 and 6, Chapter 5. The results were broadly consistent with the SAH. However, they also raised some methodological questions which are discussed later in the chapter.

The broad question that covers all of these experiments is: how sensitive is the comprehension process to linguistic input qua linguistic input, relative to various other grades of relevant contextual information? The next four chapters will attempt to shed light on this question.

Chapter 3

Facts about referents as precedent labels

3.1 Introduction

One way of exploring the relative effects of linguistic and extralinguistic input on comprehension is the study of conversational precedents. The repeated use of referring expressions in dialogue enables hearers to reduce ambiguity as to what the intended referent is (H. Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Barr & Keysar, 2002; Brennan & Clark, 1996), and hearers expect speakers to be consistent in naming (Graham, Sedivy, & Khu, 2014). Evidence for this can be seen in lexical entrainment (Garrod & Anderson, 1987), during which the association between a referring expression and its referent becomes increasingly rigid through repeated use. Brennan and Clark (1996) suggested that the referential precedents set by interlocutors in the same discourse constitute partner-specific *conceptual pacts*. A conceptual pact is an agreement between interlocutors to view a referent in a particular way, and hence is sometimes referred to in the literature as entraining on a shared perspective. During the course of an interaction, discourse partners come to share a particular referring expression for a particular referent through repeated use, and expect that if someone has used a particular expression for a particular referent once, they will do so again. This

pact between discourse partners gives rise to the expectation that, if a speaker suddenly uses a different label to the precedent they have set, it must refer to a different item. This pattern of speaker-specific expectations is in line with the broader Principle of Contrast (E. V. Clark, 1987), which simply holds that "different words mean different things". In the case of conceptual pacts, this principle pertains specifically to the referring expressions that discourse partners come to agree upon with each other; unless a speaker wants their partner to attend to a different referent, they ought to use the referring expression they have already established with their partner.

A study by Brennan and H. Clark (1996) demonstrated this effect with a matching task, in which participants referred to a number of picture cards presented in different arrays. For example, in an array where more than one shoe is present (an array such as: a sandal, a loafer and a flower), the partners may establish a disambiguating label such as "pennyloafer" to identify one of the shoes. Speakers continued to use these labels in subsequent trials with partners they had established the precedent with, even when there was only one shoe in the array, which would ordinarily render a label like "pennyloafer" overinformative. However, when the same speakers spoke to new partners about the same referents, they were more likely to modify their referring expressions to be only as informative as they needed to be (e.g. "shoe"). These results were presented in support of the view that discourse partners are sensitive to the partner with whom they have made a conceptual pact through the repeated use of referring expressions.

A body of research has since grown around interrogating whether this consistent referential behaviour we observe in language users is indicative of their sensitivity to the knowledge of discourse partners, or an effect of their reliance on the linguistic associations we adopt in the course of conversation. This line of inquiry is particularly relevant to the contrastive foci of the SAH and the LAH. The SAH predicts that comprehension is fundamentally geared toward the knowledge-state of speakers, and thus predicts that comprehension is especially sensitive to this parameter of the discourse. Specialisation toward linguistic associations that are made available during an interaction is a prediction of the LAH.

The partner-specificity of conversational precedents was investigated by Barr and Keysar (2002), who argued that the entrainment produced by repeated reference may emerge from memory associations with linguistic labels, rather than conceptual pacts with discourse partners. In their study, participants were entrained on particular labels for objects in a grid. The objects were then referred to with the label that had a precedent in the discourse, either by the same speaker or a new speaker. If entrainment is based on partner-specific representations, they reasoned that a precedent label entrained with one speaker would be inhibited when it is used by a new speaker. The responses of hearers was measured with eye-tracking, as well as the referent identified by the participants. The results showed that participants fixated on the same objects (and reached for them) just as quickly when a precedent label was delivered by a new speaker as when it was delivered by the original speaker. These results were held in support of the view that a partner-independent processing benefit from linguistic precedents is driven by individual memory associations with linguistic labels.

A similar effect was found in a later study by Metzing and Brennan (2003), which demonstrated a speaker-independent benefit associated with maintained precedent labels. However, they also found that listeners experience an interference in comprehension when speakers depart from their own precedent labels, which does not occur when new speakers use a new label. In a referential coordination task, participants interacted with a confederate speaker about objects in the task, where they entrained on particular referring expressions such as *"the shiny cylinder"* with repeated references to a given object. Then, either the original speaker or a new speaker referred to the same object with either the same label or a different one (like *"the silver pipe"*). The latency of participants' first touches and fixations on the target objects was measured. When the original label was used, participants were equally quick to look at and touch the referent regardless of who the speaker was. New referring expressions, however, revealed an effect of speaker: when a speaker who had already established a precedent expression then used a new one, the hearer was slower to look at the referent than when a new speaker used a new expression.

This pattern of effects - a strong partner-independent benefit for processing when a label is used consistently, and a strong cost to processing when a speaker who has set a precedent then uses a new term - was supported by Kronmüller and Barr (2007) in two eye-tracking studies. The primary point of debate addressed by this study is the time-course of processing in the case of broken precedents: do hearers deploy their knowledge of a linguistic term first, or are partner-specific associations deployed first in the interpretation of a new label in the discourse? The time-course of eye fixations indicated that the speaker-specific effect for broken precedents resulted from the late deployment of speaker-specific information; participants first fixated on old referents associated with a label, and then corrected this pre-emption in line with information about the speaker.

The question of which sources of information hearers attend to *first* may be relevant to the question of whether comprehension is specialised toward the individual's linguistic associations, or whether the comprehension of a linguistic utterance is determined by inferences about information that is mutually shared with the speaker. It is consistent with the LAH, for example, to find that hearers fixate on the referents associated with a label regardless of the speaker in the first instance. On-line measures such as these are especially well-suited to the investigation of predictions about what a hearer will attend to at the earliest point of processing. It is worth noting here that the specialisation of the comprehension system primarily concerns sensitivity to particular domains of information in terms of the role it ultimately plays in the comprehension process. We can also expect the manipulation of these domains to manifest in outcome measures, namely: the interpretation a hearer ultimately makes, and how long it takes them to do so when these domains of input are altered.

One example of how off-line measures and on-line measures can contribute differently to this investigation is a study by Keysar et al. (2000) (also briefly outlined in §2.2), which found that hearers do not limit their search for referents according to what is copresent with the speaker, but rather consider all information available in service of reference resolution, even if it is privileged information. The experiment manipu-

lated the physical copresence of referents in order to test whether the knowledge state of speakers is represented by hearers when they comprehend labels. A director gave instructions to participants who had to move objects between compartments in a free-standing grid. Some of the compartments were obscured on one side, such that the participant was aware that only they could see some of the objects and they were in 'privileged' ground. Other objects in the grid were visible to both, and therefore in shared or common ground. In a typical trial, the director instructed a participant to "move the small candle one slot down". A two-inch candle and a three-inch candle were in the common ground of both. In privileged ground of the participant was a one-inch candle.

Although the participants were aware of which items were mutually visible, the participants often considered the privileged-ground one-inch candle as a referent, as reflected in their on-line fixations on the object array. This was apparent in the measures that indicated how hearers ultimately interpreted the referring expressions, and how long it took them to do so: participants sometimes reached for the privileged-ground object, and were delayed in identifying the correct referent (the two-inch candle) when a privileged-ground candidate was present. These results also held in a conceptual version of the task in which the privileged-ground object was obscured by a paper bag, while the common-ground object remained in full view. The knowledge that a candidate object was present, even though concealed from both the director and participant, was enough to interfere with their interpretation of director's instructions to the extent that 71% of all participants reached for the paper bag (Keysar et al., 2003). Taken together, this is evidence in support of the view that metapsychological information about the speakers, including their knowledge of previous referents for labels and knowledge about the candidate referents available to the hearer, is not a fundamental constraint on the interpretation of referring expressions. Rather, hearers may have general expectations that labels will be used consistently for the same referents, and rely on their own associations with those labels.

One criticism of these studies is that in each critical trial where the participants

had to choose between shared or privileged objects, the privileged object was always the best match for the referring expression (Brown-Schmidt & Hanna, 2011). In the example above, this means that the privileged-ground one-inch candle was the most appropriate referent for the label "small candle" in comparison to the common-ground two-inch and three-inch candles. The 'fit' of the privileged ground referent with the referring expression could therefore confer a strong competing lexical constraint on comprehension, on a constraint-based view. In other words, a model based on multiple probabilistic constraints suggests that, if other sources of information appear more *relevant* to the task than common ground, this may give rise to apparently egocentric effects in the grid experiments. This could mean that participants were not presented with an appropriate alternative referent and that, if there were two candidates of equal *plausibility* in privileged and copresent ground, the copresent object may confer a stronger constraint on interpretation.

This effect was found in a task where competitor items were in both privileged and common ground; participants were more likely to fixate on a common ground competitor than a privileged ground competitor (Hanna et al., 2003). The task required participants to play a referential game with an on-screen grid of shapes while eye-tracking measured their fixations. In a typical trial, participants followed the instruction "Put the blue triangle on the red one" with a virtual grid that contained two red triangles and one blue triangle. When both red triangles appeared in the common ground, participants fixated on both equally, and the privileged ground item was significantly less likely to produce interference when it was yellow. This supports the criticism that the plausibility of referents according to their 'fit' with the referring expression determines the amount of interference they provide in referential tasks.

A related concern is that the use of non-interactive paradigms does not provide participants with strong enough evidence for common ground. By using an interactive paradigm for the study of conversational precedents, Brown-Schmidt (2009a) found partner-specific effects for maintained precedents using an interactive conversation methodology, where participants were instructed to move shapes around an on-

screen environment while their eye movements were monitored. Confederates and participants jointly entrained on labels for images in the arrays. When a new speaker maintained a precedent set by the original confederate, this produced more competitor fixations compared to precedents maintained by the original speaker. Note that this may not necessarily indicate the use of metapsychological common ground, however; partner specific effects in reference resolution may also result from simple memory associations with a discourse partner, rather than a representation of that partner's knowledge (Horton & Slaten, 2012).

More recently, all of the studies reviewed here featured in a meta-analysis of 10 different eye-tracking studies of referential precedents (Kronmüller & Barr, 2015). The analysis examined time-course profiles of interpretation, and was principally concerned with which effects occurred earliest in the course of hearer's interpretations; specifically, which sources of information hearers fixated on at the earliest points of processing. The meta-analysis revealed a robust pattern of effects. Maintained precedents delivered by the same speaker facilitated earlier fixations on a referent than maintained precedents delivered by different speakers. However, this effect differed from others in its rapid decay; rather than contributing to the accumulation of evidence over the time-course of interpretation, the effects dissipated quickly over the time-course of interpretation. This can be interpreted as the result of episodic effects, whereby the repeated cues give a very temporary priming effect in the early fixations of the time-course data. When precedents were broken by a different speaker, hearers fixated on a target referent earlier than when precedents were broken by the same speaker. The strongest effect across all studies was a main effect of label; interpretation was facilitated earlier overall for maintained precedent labels than for broken precedent labels, regardless of the speaker (Kronmüller & Barr, 2015).

If early fixations represent a specialisation for particular domains of input for comprehension, that input should be preferentially weighted during processing and this should be manifest in outcome measures. For example, if linguistic precedents are a special domain for comprehension, broken precedents should affect the interpretation

that hearers make, and how long it takes them to make it. If, however, the knowledge state of speakers confers a fundamental constraint on interpretation, we would expect to see a considerable *cost* to processing in cases where precedents are maintained by a different speaker. Such a cost could result either from the inhibition of private knowledge, or from a new referent search necessitated by the unexpected event of a new speaker using a label for the same object established by a previous speaker. This processing cost affects the overall time taken to reach an interpretation, particularly if recovery measures are employed during the course of processing. The absence of a processing cost in maintained precedent conditions with a new speaker is contrary to a model like the SAH, in which the knowledge states of speakers are an obligatory constraint or source of information for referential interpretation.

One reason this effect is absent could be that the lexical meanings of straightforward labels like "the tent" or "the shiny cylinder" confer stronger constraints on comprehension that obscure speaker effects when labels are maintained by new speakers. This is related to earlier criticisms of precedent studies in which privileged ground objects are a better "fit" to precedent labels than common ground objects; if the lexical meaning itself exerts a strong constraint, this may obscure speaker effects with maintained precedents. Following this criticism, we might expect that if the knowledge state of the speaker is the most relevant parameter for the successful resolution of reference (and therefore the *most relevant constraint* on interpretation), this may reveal an effect of speaker that underlies the successful use of labels in context. In other words, would a processing benefit for maintained precedent labels regardless of speaker hold in cases where the knowledge of the speaker is *required* for the successful comprehension of a referring expression?

One way of bringing out speaker-specific effects of this type may lie in using speaker-specific referring expressions in place of the simple descriptive noun phrases used in previous precedent studies. This approach was used in a study by Diesendruck and Markson (2001) that investigated children's avoidance of lexical overlap, or 'mutual exclusivity bias'. In the classic mutual exclusivity experiment, 3 year old children

were presented with names for corresponding pictures of unfamiliar objects. Children rejected second names for objects, and instead attributed new names to new objects. For example, when children were taught that one unfamiliar object was a *wug* and then were asked for the *mef*, the children would choose a different object. In order to demonstrate that this mutual exclusivity bias extends beyond simple word learning into other domains of language use and was underpinned by more pragmatic processes, Diesendruck and Markson's study used facts about the referents as referring expressions as well as the canonical nonsense words. Children were given two unfamiliar objects, taught a name for one of the objects ("This is a *mef*), and then asked to pick out an object using a new name ("Can you pass me the *wug*?") . In other trials, children were given two unfamiliar objects, taught a fact about one of the objects (*"My uncle gave this to me"*) and then asked to pick out an object using a new fact as a referring expression (*"Can you give me the one my cat likes to play with?"*). The fact-based referring expressions related specifically to the speaker (i.e. "the one *my* cat likes to play with"). Children chose a different object when prompted by a different referring expression in both the label-based and fact-based conditions. This demonstrated that children's avoidance of overlap between referring expressions wasn't restricted to the acquisition of lexical items, but was argued to represent a strategy based on the communicative intentions of speakers similar to that of conceptual pacts: rather than being a simple bias specific to the acquisition of words, mutual exclusivity could be indicative of inferred expectations about how speakers use referring expressions.

3.2 Motivation

The experiment presented in this chapter investigates the relative effects of prior linguistic precedents and speaker-specific knowledge in the comprehension of referring expressions. We can expect the comprehension system to exhibit sensitivity to the particular types of input for which it is specialised. As established in §2.3, the organisation of cognitive systems (as with all biological structures) tends toward specialisation of

function, which often results in sensitivity to particular domains of input. More specifically, this input will exert an effect on the ease with which an interpretation is reached, and thus the time taken to do so, as well as the interpretation itself. We may then expect slower response latencies, and less successful identifications of a referent, when these special domains of input are disrupted in some way. Most studies on conversational precedents have examined how comprehension is affected when a speaker who is either familiar with or new to the discourse maintains a precedent label compared to when she breaks a precedent label (Kronmüller & Barr, 2015). For example, when a speaker refers to an object as "the black vase", then later refers to it as "the upside-down funnel", this typically incurs a processing cost for the hearer; when a speaker refers to an object as "the black vase" but a *different* speaker calls it "the upside-down funnel", this cost does not occur (Kronmüller & Barr, 2007). This paradigm lends itself well to investigating the relative effects of speakers' knowledge and linguistic labels on the comprehension process.

The Social Adaptation Hypothesis (SAH) of linguistic comprehension suggests that linguistic processing is constrained and determined by the relevant parameters of the discourse, and in particular by metapsychological knowledge about the speaker. This follows from an account of linguistic comprehension that is rooted in social cognition, and therefore fundamentally inferential at all levels of processing (Sperber, 2000; Scott-Phillips, 2014). Recall that if comprehension is primarily shaped by adaptation to the social environment, we would expect comprehension to be most sensitive to the knowledge state of the speaker when it comes to interpreting utterances. The meanings conventionally associated with words, or "knowing a language", can be characterised in this case as a set of knowledge that constrains one's expectations from the discourse. On this view, interlocutors process referring expressions as *evidence* for what the speaker intends, and in this sense the utterance functions as an index for the speaker's mental state.

The Linguistic Adaptation Hypothesis (LAH) of linguistic comprehension suggests that linguistic processing is specialised to exploit linguistic labels as a source of con-

ventional information in interpreting utterances. This account allows that there may be sensitivity to the knowledge state of the hearer in comprehension, but holds that there is specialisation for linguistic input in particular that should be observable in processing. It follows from this account that a linguistic association, once made, will be more readily accessible to the hearer regardless of whether or not it is delivered by another speaker. In other words, the comprehension system has a stronger reliance on linguistic conventions, and so once a referring expression has been used for a particular referent, future uses of this expression will yield the same interpretation. On this view, hearers process referring expressions according to their own meaning associations with the label, and in this sense the utterance functions as an index for a given meaning.

Speaker-specific facts about referents as referring expressions are *explicit* indexes of speaker's mental states, and this may bring out the effect of speakers' knowledge states on the comprehension of conversational precedents. Phrases such as *"Can you give me the one that makes me think of my sister?"* and *"Can you give me the one that reminds me of my neighbour?"* require hearers to make recourse to speaker-specific knowledge if they are to make a correct interpretation. Speaker-specific information is considered obligatory for successful interpretation of phrases like this, where the semantic content of possessive determiners and pronouns (i.e. it reminds *me* of my neighbour, it makes *me* think of *my* sister) necessarily depend on extralinguistic information, and explicitly pertain to the knowledge of speaker.

The SAH view proposes that common ground with the speaker is the special domain of input upon which hearers depend for interpretation. The utterance is treated as evidence for the speaker's mental state, and thus requires the hearer to make recourse to common ground. In the case of speaker-specific referring expressions, the relevance of the speaker's mental state is made explicit, and therefore confers a strong constraint. In contrast, the LAH view proposes that hearers bring the referential meaning they have associated with the utterance itself to bear on their interpretation. The use of speaker-specific referring expressions as precedent labels may thus provide insight into whether hearers preferentially anchor their comprehension of such precedents in

the most relevant speaker-specific knowledge, or the most convenient linguistic association.

The experiment in this chapter compares the effect of the speaker and precedent label on the reaction time and referent selections of participants. In a simple matching game, participants were presented with an array of pictures of everyday objects. One of the objects was referred to by a speaker-specific precedent label such as "This one reminds me of my uncle". In a subsequent phase, the same speaker may then ask the participant to select an object from the same array by maintaining this label ("Which one reminds me of my uncle?"), or they may use a different label (for example, "Which one reminds me of my sister?"). In other trials, a different speaker may deliver a prompt that maintains or breaks the precedent label in this same way.

All of the objects are equally compatible with these idiosyncratic referring expressions, and thus the constraint of lexical effects conferred by more straightforwardly descriptive phrases (like "the shiny cylinder" in previous experiments) may be minimised. The sources of information available to the participants are their own associations with the linguistic utterance, and their knowledge about whether a speaker has used the expression before - that is to say, whether a speaker has told them that an object "reminds me of my uncle". The time it takes participants to select a referent, and the particular referent they choose, may give evidence as to which domains of information are being brought to bear on referential comprehension. If the comprehension of precedents grants more weight to the knowledge of the speaker, an effect of speaker should be exerted on a subjects' choice of referent and the length of time that choice takes. Similarly, if the subject's own association with the linguistic precedent plays a larger role in comprehension, an effect of label should be exerted on the referent choice and reaction time.

3.2.1 Predictions

The Social Adaptation Hypothesis (SAH) of comprehension holds that referential interpretation is based on the most relevant contextual information for the task, including knowledge state of the speaker. In the following experiment, the identity of the speaker forms a component of the referring expression itself, and so would confer a strong constraint on comprehension on this view. A comprehension process that is crucially informed by relevant common ground with the speaker predicts a main effect of Speaker on the reaction time of participants, and an interaction with Label. When the same speaker maintains a precedent label, this ought to elicit the quickest response from participants.

A key prediction of the SAH is that, when a speaker-specific precedent is used by another speaker (i.e. a Different speaker Maintains a precedent label), it will elicit a slower reaction time that is comparable to when precedent labels are broken. This is because the maintained label refers explicitly to an object established by the first speaker; an object that reminds Speaker A of their uncle has no reason to also remind Speaker B of *their* uncle, as all of the objects could be equally plausible (items are cycled through all conditions across the whole experiment in a latin square design). Reaching an interpretation in this situation ought to be comparable to when a speaker breaks their own precedent, and when a different speaker breaks a precedent; the precedent object itself ought to be inhibited, and other referents ought to be considered equally plausible. This equivocal situation in all cases ought to produce a considerable delay in response latency, if interpretation is based on the knowledge state of the speaker.

The SAH of comprehension also predicts that the object referred to by a precedent label will not be selected more often by participants when an established label is delivered by a new speaker; while all of the referents are plausible candidates for reminding a speaker of something, we should expect (on this view) that when a new speaker uses an established precedent, all of the available objects have an equal chance of being selected. In other words, a maintained precedent for an object should *not*

elicit more selections of that object if the label is delivered by a new speaker. When the same Speaker breaks their own precedent, we would expect that the referent of the previous label will not be selected. This model therefore predicts a significant effect of Speaker on whether their object selection matches the object to which the precedent label earlier referred.

By contrast, the Linguistic Adaptation Hypothesis (LAH) of comprehension holds that hearers will rely on the association they have made between a referent and the linguistic label, with less regard for the identity or knowledge state of the speaker. Though the semantic content of possessive determiners and pronouns is context dependent, we might expect (on this view) that the linguistic input *itself*, which has already been associated with a particular referent, will form the basis of interpretation when the label occurs again. This model therefore predicts a main effect of Label on the reaction time of participants, and an interaction with speaker. As with the SAH, reaction times ought to be quickest when the same speaker maintains their own precedent, as a result of repetition and episodic priming effects. The LAH predicts that broken labels will incur a processing cost and therefore elicit a slower reaction time. While the current body of research on precedents indicates an effect of speaker in the case of broken precedents, this is not necessarily entailed by the LAH. There may be episodic effects whereby the same speaker delivering the same label means there are more cues for retrieval; while these manifest early in time-course data, as in Kronmüller and Barr (2015), the rapid decay of these effects over the course of processing means they may not appear in the off-line behavioural measures of the present experiment.

A key prediction of the LAH is that, when a speaker-specific precedent is used by another speaker (i.e. a Different speaker Maintains a precedent label), it will elicit a faster reaction time than when precedent labels are broken by either speaker. This is because the maintained label has been previously associated with a referent, and this representation will be brought to bear on the interpretation of that same label, regardless of whether the speaker is in possession of appropriate knowledge. Unlike cases where the label is broken and all referents are plausible, thus eliciting significantly

slower reaction times, the LAH predicts that when a different speaker maintains a label, a referent will be selected with relative ease.

The LAH also predicts that the objects referred to by a precedent label may be selected equally as often by participants when a precedent is maintained, regardless of whether it is delivered by the same or a different speaker. While the linguistic adaptation account does suggest that speaker specific information can be integrated in comprehension, specialisation for individual associations with linguistic input may mean that this is more often elicited from the feedback available in a typical discourse environment (Barr & Keysar, 2005, see also §2.2). This is unavailable to the participants in the experimental setting, where there is a time limit for responding to each trial and there is no feedback offered. We would expect this to give rise to more systematic errors than would arise in more naturalistic settings. If this is the case, we may therefore expect the main effect of Label on reaction time to also manifest in the coarser measure of referent selection.

3.3 Experiment 1

3.3.1 Participants

40 native English speakers (12 male, median age 19.5) were recruited from the Edinburgh University campus via mailing lists and word-of-mouth. Participation was voluntary, with no monetary incentive. A temporary software issue resulted in a failure to log the responses of one participant.

3.3.2 Design

The experiment was a within-subjects and items design, with two factors of two levels: Speaker [Same, Different] where the speakers in the precedent and prompt phase of each trial were either the same speaker or a different speaker, and Label [Maintain, Break] whereby the referring expression used in the precedent and prompt phase of

each trial was either maintained or broken. All participants were therefore exposed to a total of four conditions, by the factorial combination of all levels of Speaker and Label.

3.3.3 Materials

Visual Arrays:

Twenty-four arrays of images were used. Each array comprised four images, and each image was a photograph of an everyday object. The photographs were converted to 400 x 300 pixel black and white jpeg files, corrected for consistent contrast. An example array is presented in Fig. 3.1, as shown in the establishment phase and prompt phase of a critical trial.

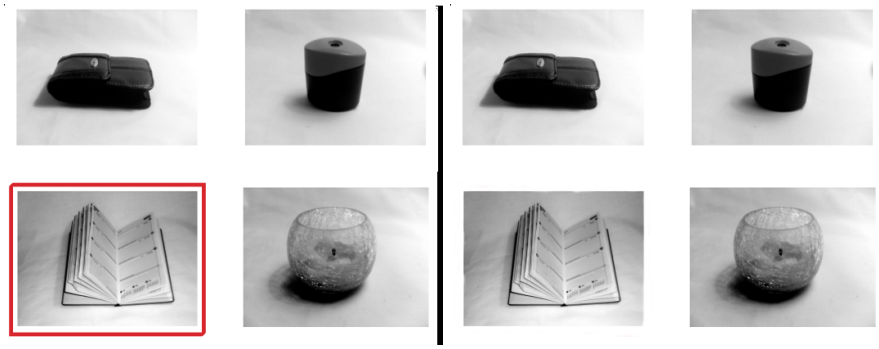


Figure 3.1: One of twenty-four visual arrays used in Experiment 1, shown here twice. Establishment phase: One object in the array would be highlighted by a red box around its cell, accompanied by a referring expression. Prompt phase: The highlight disappeared, and the array remained on screen while participants received an audio prompt to select one of the objects.

Audio:

The audio stimuli were recorded in two sessions by two volunteer voice actors. To help ensure that participants would differentiate between speakers, the volunteers were one male and one female. Each audio set consisted of an idiosyncratic memory association. The format of referring expressions was 'This makes me think of' or 'This reminds me of', such that all referring expressions explicitly depended on the knowledge of the speaker. The audio sets contained a phrase that established the precedent (e.g.

"This makes me think of Thursdays"), and a corresponding prompt (e.g. "Click on the thing that reminds me of Thursdays"). The memory associations were restricted to interpersonal relations (such as "my sister" or "my nephew"), particular times or events (such as "my birthday" or "my job"), and personal places (such as "my hallway"). This was to further ensure that the referring expression explicitly depended on some relation to the speaker, rather than entities with more salient relations (such as "This reminds me of Christmas"). Audio sets were edited and counterbalanced according to the order of male and female prompts in each condition. All audio files were edited using the Audacity®software package.

Following the 2 x 2 experiment design, there were four lists of stimuli items. Each item, consisting of audio prompts paired with visual arrays, was rotated through all four experimental conditions in the manner of a latin square. The arrangement by condition is illustrated in table 3.1 below.

Condition	*Establishment phase*	*Critical prompt*
Same Speaker **M**aintain Label	*Speaker* ***A:*** This one reminds me of **my uncle.**	*Speaker* ***A:*** Which one reminds me of **my uncle**?
Same Speaker **B**reak Label	*Speaker* ***A:*** This one reminds me of **my sister.**	*Speaker* ***A:*** Which one reminds me of **my uncle**?
Different Speaker **M**aintain Label	*Speaker* ***B:*** This one reminds me of **my uncle.**	*Speaker* ***A:*** Which one reminds me of **my uncle**?
Different Speaker **B**reak Label	*Speaker* ***B:*** This one reminds me of **my sister.**	*Speaker* ***A:*** Which one reminds me of **my uncle**?

Table 3.1: The composition of audio prompts for each experimental condition, according to manipulations of *Label* and *Speaker*. Note that the prompt phase audio always uses the same referring expression across conditions; this ensures that reaction times are comparable.

3.3.4 Procedure

All experimental stimuli were administered using E-Prime software (Psychology Software Tools Inc., 2014).

Briefing and instructions:

Participants were tested individually, in acoustically deadened booths. They received a scripted verbal briefing from the experimenter, which contained a cover story about

what the experiment entailed. The task was presented as a memory game, in which they would be asked to identify objects that had particular memory associations for the instructors. The briefing was that previous participants had brought in everyday objects from home, for which they had an associated memory. These previous participants had used the objects to play a memory game with each other, where objects had to be identified according to what they reminded the speakers of. Participants were told that they were about to hear the recordings of two such players, in the interest of seeing how they performed without the players being present. It was impressed upon participants that the two players they were about to hear had never met each other, that had in fact played the game on different days, and that their recordings had merely been brought together to construct the current game. Participants were then told how to use the Serial Response Box.

The trial instructions were delivered verbally, and repeated by on-screen instructions that participants could go through at their own pace. Participants were told they would see four objects on the screen in front of them. It was explained that a red frame would highlight an object and at that time the participant would hear the owner's memory association with it. The red frame would then disappear. It was explained that the participant would then be asked to select an object from the array using the Serial Response Box buttons in front of them, and that there was a time limit of 4 seconds. Participants were told that if they did not know which selection to make, they could simply guess. The on-screen instructions were followed by examples, whereby a photographed object (e.g. a tiny string doll) was given a memory association (e.g. a lucky charm for long-jumping). The purpose of this illustrative example was to impress upon participants the idiosyncratic nature of the associations.

Practice and Comprehension check:

In order for participants to become familiar with the button-screen configuration, participants were given a practice round with counterbalanced male and female instructions. The practice round consisted of four trials that followed the critical trial pro-

cedure. After the practice round, participants were asked three comprehension check questions (target answers given in italics): Can you tell me what the red frame is for? *It highlights the object being talked about*; Did you hear an instruction to select an object before you did so? *Yes*; Did the two players you were listening to know each other? *No*. If the comprehension questions were answered incorrectly, the instructions, practice and checks were repeated to ensure the participant understood before proceeding to the test trials.

Critical trials:

Each trial consisted of a precedent phase, immediately followed by a prompt phase.

Precedent phase: Participants were presented with a new array of four objects for each new trial. A red frame then highlighted one object in the array, for which a precedent label was established by accompanying audio (e.g. "This makes me think of my niece"). After this audio was delivered, the frame disappeared and the object array remained.

Prompt phase: The participant would then hear either the same speaker, or a different speaker, prompt the participant to select an object. This prompt would be delivered using either the same label expression (in this example, "Give me the thing that reminds me of my niece") or a different referring expression (e.g. "Give me the thing that reminds me of my Mum"). Upon hearing the prompt, the participant was then to use one of the four Serial Response Box buttons to select one of the four objects in the array. The participants were not given feedback about whether they had selected the correct object; the game simply proceeded after each selection. This lack of feedback and the time pressure of the task is more likely to pressure participants to rely on their most reliable, least effortful strategy of comprehension.

After completing the experiment, participants were invited to fill in an exit questionnaire in order to gauge what they thought the experiment was about, and what their answering strategy was.

3.3.5 Dependent variables

Reaction Time

Participants' responses were made via a Serial Response Box, the reaction time (RT) of which was recorded by E-Prime software (Psychology Software Tools Inc., 2014). For each trial, the audio file for the prompt phase played over a single slide that displayed the object array. The slide terminated upon the participant's response, or upon the exhaustion of the imposed time limit. When participants did not respond before the time limit, no data was logged. When participants did select an object, the RT for the entire slide was logged. Reaction time was therefore calculated by subtracting the duration of the audio prompt (up to the onset of the referring expression) from the total slide RT. Note that the referring expression in the prompt phase was the same across all four conditions, as illustrated in table 3.1. This method allowed an objective criterion for the identification of noncompliant trials; in cases where the calculated RT was negative, a response had been given before the critical referring expression in the prompt could have been heard by the participant. Trials that resulted in a negative reaction time during data processing were therefore omitted from analysis on the grounds of task noncompliance.

Precedent Match

The red highlight around the 'precedent object' in the establishment phase of each trial corresponded to a cell in each array. The position of the highlight was random for each trial, but was pre-set in the experiment code. The participant responses provided in the prompt phase of each trial also corresponded to a cell in each array.

Both values were logged for each trial, and it was therefore possible to calculate whether or not the object selected in the prompt phase matched the object highlighted in the establishment phase.

3.4 Results

3.4.1 Reaction Time

I used R (R Core Team, 2013) and LME4 (Bates, Mächler, Bolker, & Walker, 2015) for a linear mixed effects analysis of the relationship between reaction time (RT), Speaker and Label. As fixed effects, I entered Speaker and Label (with interaction) into the model, using Same Speaker, Maintained Label as the reference level. As random effects, I had a maximal structure (Barr, Levy, Scheepers, & Tily, 2013) of random intercepts for subjects and items, as well as by-subject and by-item random slopes for Speaker and Label. In some cases these models failed to converge, in which case the structure was simplified by removing interactions between random slopes. In these cases, the non-converging models indicated the same pattern of effects and significance as the reduced random effects models. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.

Label affected RT ($\chi^2(1) = 20.64$, $p < 0.001$), slowing response time by 448.7ms $\pm$40.6 (SE) when a label precedent was broken. There was no main effect of Speaker ($\chi^2(1) = 2.79$, $p = 0.09$). There was a significant interaction between Label and Speaker, ($\chi^2(1)=13.99$, $p < 0.001$), whereby responses for broken labels were quicker by 266.78ms $\pm$67.9 (SE) when they were delivered by a different speaker. Reaction times per condition are illustrated in Fig. 3.2.

3.4.2 Object Selection

The cells selected in the critical trial were automatically coded by whether they failed to match the cell highlighted in the precedent trial (Nonmatches = 0) or matched their response to the precedent (Matches = 1). I used R (R Core Team, 2013) and LME4 (Bates et al., 2015) for logistic mixed-effects regressions with a maximal random effects structure (Barr et al., 2013) of random intercepts for subjects and items, and

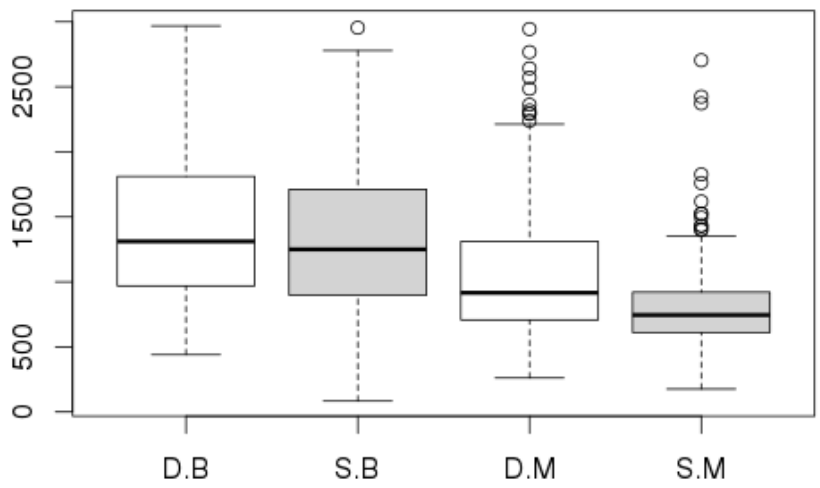


Figure 3.2: Reaction Time (ms) by condition (Speaker, Label): DB = Different speaker and Broken precedent; SB = Same speaker and Broken precedent; DM = Different speaker and Maintained precedent; SM = Same speaker and Maintained precedent.

by-item random slopes for fixed effect manipulations, using Non-matches as the reference level. In some cases these models failed to converge, in which case the structure was simplified by removing interactions between random slopes. In these cases, the non-converging models indicated the same pattern of effects and significance as the reduced random effects models.

A logistic regression explored the effect of Speaker (taking Same Speaker trials as the reference level) and Label (with Maintained Label trials as the reference level), with each trial performance coded as matching the precedent object (1) or not (0). There was a significant main effect of Label (β = -0.65, SE = 0.33, p = 0.049), whereby broken labels led to fewer selections of the precedent object. A change of Speaker made no significant change in the likelihood of matches (β = -0.54, SE = 0.29, p = 0.069), with no significant interactions (β = -0.54, SE = 0.45, p = 0.23). Fig. 3.3 illustrates matches vs nonmatches per condition.

3.5 Discussion

The reaction time data exhibited a main effect of Label, whereby Broken precedents led to much slower responses from participants. There was also a significant interaction

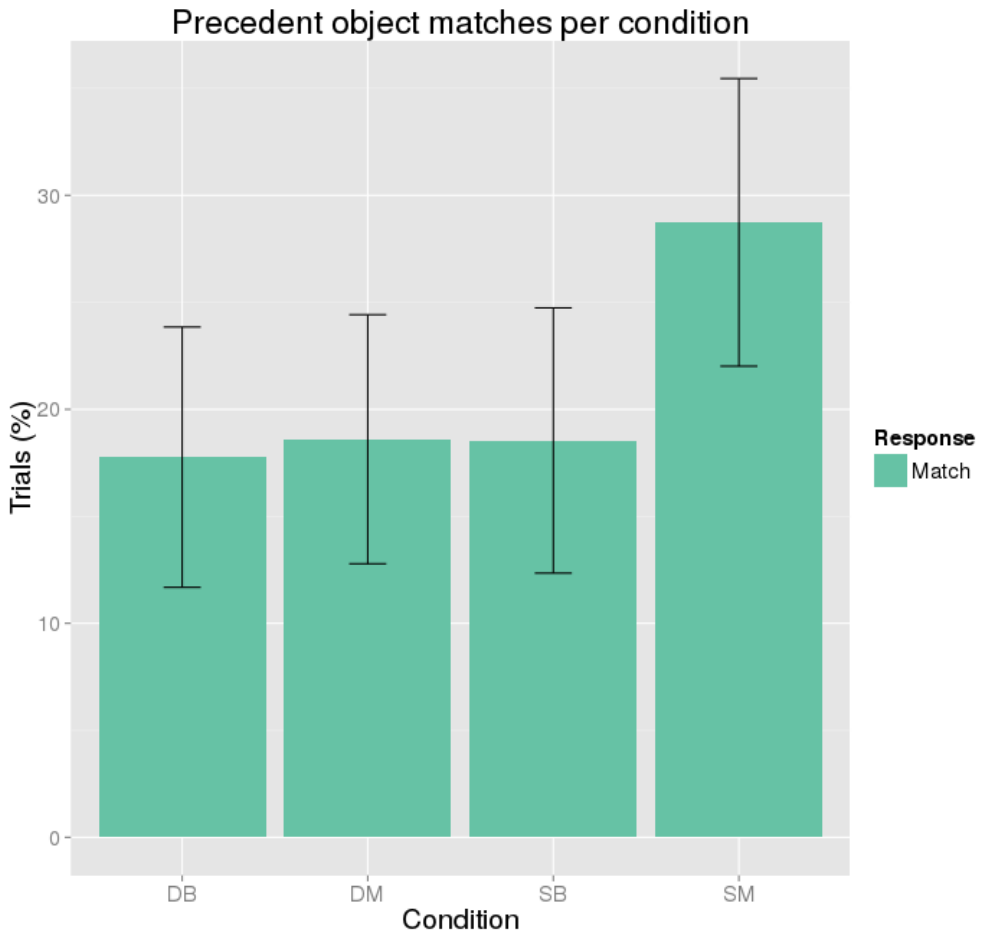


Figure 3.3: Proportions of precedent matches by condition (Speaker, Label): DB = Different speaker and Broken precedent; SB = Same speaker and Broken precedent; DM = Different speaker and Maintained precedent; SM = Same speaker and Maintained precedent. Error bars show 95% confidence interval.

with Speaker, whereby maintained labels delivered by the same speaker elicited the fastest reaction times from participants. The reaction time data exhibits a main effect of Label and an interaction with speaker. This is consistent with the wider body of data on conversational precedents, and is predicted by the Linguistic Adaptation hypothesis (LAH) of comprehension.

There was no significant main effect of Speaker in the reaction time data, and maintained labels led to faster reaction times than broken labels, regardless of speaker. Recall that all of the referring expressions explicitly called attention to the speaker by the use of personal possessive pronouns and determiners, and explicitly rested on the knowledge state of the speaker. Despite this, participants appeared to have associated

only *the linguistic label* with the referent in the establishment phase, rather than *the knowledge of the speaker* as indicated by that label. This result is inconsistent with the Social Adaptation hypothesis (SAH).

An interaction between speaker and label in the reaction time data confirmed a same speaker advantage for maintained precedents, and a smaller different speaker advantage for broken labels. This may be more clearly illustrated in Fig. 3.4 below, which depicts a higher density of responses around the 1000ms mark for different speaker, break label trials; note that same speaker, break trials elicited more responses toward the 2000ms mark. This pattern is consistent with what we may expect from the eye-tracking data, whereby fixations are facilitated earlier in these same conditions (Kronmüller & Barr, 2015). Unlike previous eye-tracking studies, the current experiment cannot lend insight to the question of *when* these effects occur during processing. One interpretation of this interaction, which is consistent with the main effect of label exhibited here, as well as patterns in the wider literature, is that hearers recover from broken precedent labels quicker when the new label is delivered by a new speaker.

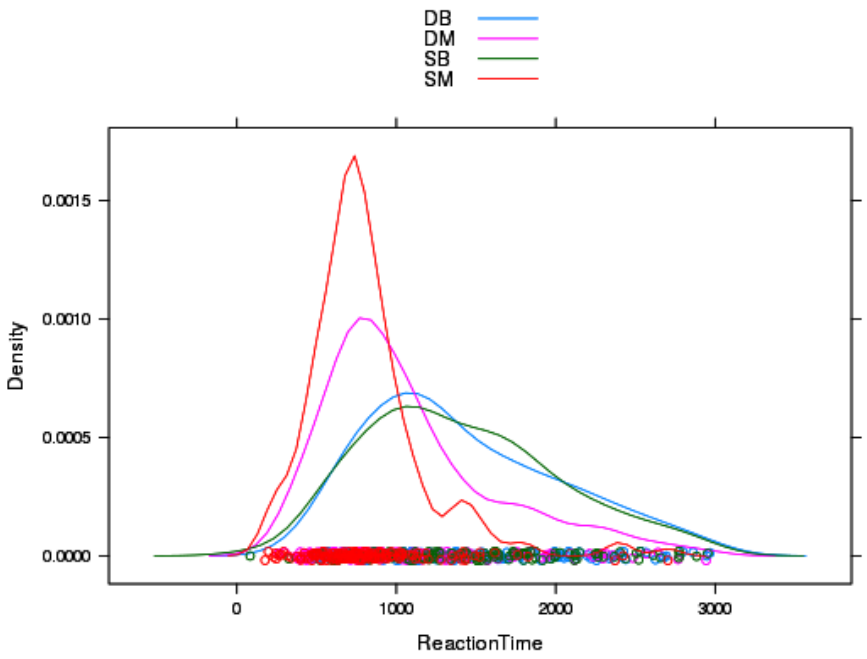


Figure 3.4: Density plot of reaction times by condition (Speaker, Label). DB = Different speaker and Broken precedent; SB = Same speaker and Broken precedent; DM = Different speaker and Maintained precedent; SM = Same speaker and Maintained precedent

Trials where Different speakers delivered Maintained precedent labels were a key condition for the contrastive predictions of the LAH and the SAH. The SAH predicts that the knowledge state of the speaker forms the critical foundation of utterance interpretation, and so when a Different speaker uses a speaker-knowledge-dependent label that has been established by someone else, the result should be as equivocal and delayed as when a speaker uses a new expression with no precedent in the discourse. The LAH, on the other hand, predicts that maintained precedent labels will confer a faster reaction time than broken labels, regardless of the knowledge state of the speaker. The main effect of Label on reaction time indicates that maintained precedent labels elicited faster responses across both levels of speaker, and is consistent with the LAH.

The pattern of effects in the object selection data was also consistent with the LAH. Participants preferentially matched their selections to the precedent object maintained label trials delivered by the same speaker, and there was a main effect of Label on object selection. In addition to this, the precedent object was selected about as often in trials where the same speaker broke their own precedent as when the precedents were broken by a different speaker, which is inconsistent with the kind of mutual exclusivity we may have expected from a SAH comprehension system; if a precedent has been broken by the same speaker, the SAH would not expect this object to be selected when the same speaker used a new label.

A key distinction between the language-driven and context-driven approaches to comprehension outlined in Chapter 2 (and §2.1 in particular) was the stability of semantic information as part of the inferential process; where the language-driven view accepts the classical delineation between semantic content and extralinguistic contextual information, a more fundamentally inferential view holds that this content itself is the result of multiple relevant constraints that guide the hearer's interpretation of the utterance in the first instance. It would seem, then, that a disregard for the particularised semantic meaning of items like *me* and *my* in phrases such as "Which thing reminds *me* of *my* niece?" may undermine a language-driven interpretation of this experiment's main effect of Label. If this semantic content is as stable as the language-driven view

suggests, why wouldn't it constrain comprehension toward some speaker-specificity? It could be argued that the experimental situation itself led participants to treat language in an atypical way. However, an equally plausible explanation may be that the semantic content of "my" is simply an index of the speaker *whoever the speaker may be*, rather than a category that takes on more specific interpretation on the basis of the particular speaker's identity. In other words, the results of this experiment appear to show that hearers (under time pressure) will recognise the meaning of these personal pronouns/determiners to be a nonspecific reference to any given speaker. This entire proposition is assigned to the referent in a temporary convention, and the recency/availability of this convention lends itself to repeated access throughout the discourse. That the linguistic stimulus lends itself to a stronger associative effect than the speakers do, even when that stimulus ostensibly *relies* on information about the speaker, is consistent with the kind of input sensitivity we would expect of Linguistic Adaptation, rather than Social Adaptation alone.

3.5.1 Limitations and future directions

While measures such as reaction time and object selection can give an indication of which domains may be preferentially attended to by participants for the task of comprehension, they cannot indicate the time-course of how that process is taking place, or the order in which participants are considering particular referents.

The use of explicitly subjective referring expressions such as "Can you give me the thing that reminds me of my sister?" meant that all objects in the grid were compatible potential referents in all conditions, with the exception of the reference level (Maintained precedent labels delivered by the Same speaker). This meant that participants were not coerced into choosing a particular referent by virtue of its match to a descriptive referring expression, a criticism that has been applied to previous studies on precedents (Brown-Schmidt & Hanna, 2011). However, this level of ambiguity is likely to have had the effect of leaving participants equally perplexed about which ob-

ject to choose in the event of a new label, regardless of whether it was delivered by a new or a different speaker. To some extent this radical ambiguity is not representative of typical conversations between language users.

In this light, all of the effects present - faster reaction times for broken precedents with a different speaker than by the same speaker, and faster referent selection in maintained precedent trials across levels of speaker - may be best understood in terms of *recovery* in all conditions other than the reference level. In one sense, this highlights a conceptual point in favour of the LAH: the broken label trials in particular were more difficult for participants to complete. Making recourse to the available linguistic convention can be a successful strategy for reducing extreme ambiguity. And further: without this crutch, comprehension takes longer, and may run into difficulty. This interpretation is consistent with the rates at which participants failed to make a selection across trials (N = 346 successful trials for maintained label trials, N = 303 successful trials for broken label trials; for comparison, N = 325 successful trials for same speaker trials, N = 324 successful trials for different speaker trials).

However, the lack of alternative common ground from which to infer speaker information may also mean that the recovery enabled by the linguistic precedent was not in clear opposition to any inferred information (except in the reference condition). The lack of alternative information in the form of common ground could conceivably be responsible for the main effect of label; if a rival source of information in the form of common ground were available across all conditions, perhaps the results would show a pattern of 'recovery from different speaker' rather than 'recovery from broken label'.

This does, however, raise an interesting question about why there was a small, but statistically significant, benefit to reaction time in cases where different speakers delivered broken labels. It is not clear how a different speaker could have constrained interpretation to aid recovery in the absence of lexical constraint from a descriptive phrase, or inferred information from common ground. Perhaps this effect (which was very slight, as shown in Fig. 3.4) would be larger if such other possibilities were available to participants.

While the object selection data was consistent with the LAH, it is worth noting the small scale of these effects; the main effect of Label approached the conventional threshold for statistical significance ($p = 0.049$). While a main effect of Label may give the impression that participants were equally likely to match their selection to the precedent object regardless of speaker, this is not quite the case. Fig. 3.3 shows that this effect appears to be carried by the same speaker, maintained label condition; a significant interaction with speaker fell slightly short of the conventional threshold for statistical significance ($p = 0.069$), and reflects that the only condition to show markedly increased matches was SM. The number of non-matches in all other conditions was stable at around 82%. This proportion was higher than expected, as the precedent object was one of four objects in every array, and so chance-level selection would have put non-matches at 75%. This indicates that the precedent object was preferentially excluded in all conditions except SM.

The most forthcoming explanation for this may be related to the ambiguity of the trials, in that participants simply found all other conditions much more difficult, and the repeated exposures of SM provided instances where they were most sure of how to proceed. The number of non-matches in SM trials were still much higher than expected at 71.5% of SM trials; while this means participants matched their responses to the precedent object more than chance, and more than in the other conditions, the effect was still notably small. Together, these indications from the object selection data may mean that the overall difficulty of the task was simply too high for participants to perform reliably. The high level of difficulty is also indicated by the number of trials that were omitted from analysis on the grounds of noncompliance with the task; of a total 936 trials, 287 recorded reaction times of less than 0, which indicated a response that was made before the target instruction could have been heard. In other words, just under a third of all trials were noncompliant with the task, which may have been the result of confusion or 'giving up' on trying to give a correct answer. The feedback in the exit questionnaires supports this, with many indicating that they found the task very challenging, and a few citing confusion.

A lack of feedback, and time pressure for the task, may have contributed to its overall difficulty. These features meant that participants were encouraged to rely on more default processing strategies rather than amend their performance in response to a score or an interactive partner. This was intended to reveal any systematic errors that may result from preferentially weighting individual associations with the linguistic precedents). However, one cost to this approach is that it may lack some ecological validity, and may set the level of difficulty too high to reap reliable results. A lack of feedback may have also equated to a lack of incentive for participants to try to comprehend despite the difficulty of the task, and may have instead encouraged a strategy of simply guessing.

The nature of the referring expressions in this task explicitly rested on the speaker's knowledge, in order to put a stronger constraint on interpretation and potentially reveal speaker-specific effects in the condition of Maintained precedent labels delivered by a Different speaker. It may be that the speaker-specifity of the input still did not achieve the strength or relevance that a more natural discourse environment would confer. The challenge for proposing amendments that address this criticism is to define the parameters that confer a strong enough constraint of common ground on interpretation. The use of linguistic-level constructions, conceptual-level expressions and task goals that explicitly require speaker-specific contextual information for correct interpretation would ostensibly provide such a constraint. Any claim that they do not must entail more than simply observing that the expected effect did not occur; it must detail why these constraints were insufficiently salient, and the conditions that must be met in order for them to become so (see also Barr, 2008). One answer may be that non-interactive paradigms attenuate effects of speaker that would be present in more natural conversational settings.

In light of this, a future direction of study may look at using speaker-specific facts as conversational referents in a task that uses live confederates in place of recorded speakers. An interactive paradigm may also improve the difficulty of the task, as a briefing story about the recorded speakers would not be necessary. This would relieve

participants of the cognitive burden that may accompany understanding the speakers in more abstract, rather than interactive, terms. Another way of improving the task difficulty would be to ensure that there is some degree of plausibility that can narrow the search for referents, while also ensuring that the plausibility did not coerce participants to privilege one independent variable over another.

3.6 Conclusion

This chapter has presented an overview of the psycholinguistic literature on conversational precedents, and considered the utility of this vein of study for examining the relative effects of linguistic input and shared knowledge on referential comprehension. A reaction time study of precedent effects was presented to interrogate the contrastive hypotheses of the Linguistic Adaptation Hypothesis (LAH), and the Social Adaptation Hypothesis (SAH) of comprehension. The manipulations were Label and Speaker. During the task, an object would be referred to by a particular label, and then referred to again either by the same label or a different one, by the same speaker or a different speaker. Participants then had to select the correct referent. All referring expressions used in the tasks depended on the knowledge of the speaker. The aim of this method was to strengthen the constraint that common ground may exert on referential comprehension by rendering all objects in the referential task equally plausible unless participants made recourse to shared knowledge with the speaker. The LAH predicted a main effect of Label, and allows for an interaction with Speaker. Contrastively, the SAH predicted a main effect of Speaker, and allows an interaction with Label. A condition in which Maintained precedent labels were delivered by Different speakers was a key contrast between both accounts, whereby the LAH predicted faster reaction times than broken label trials, and the SAH predicted comparably slow reaction times. The reaction time results exhibited a main effect of Label, whereby reaction times were faster for maintained labels than broken labels, and an interaction with Speaker. Participants responded to the referring expressions quickest when the same speaker maintained a

label, and responded quicker overall when labels were maintained regardless of the speaker. This is consistent with the existing literature on conversational precedents, and with the predictions of the LAH.

However, limitations of the study included a level of referential ambiguity in the case of broken labels that was, to a considerable extent, unlike normal discourse situations. The off-line dependent measures mean that real-time information about how participants recovered in this situation was not available. In addition, the object selection data revealed that participants tended, in all cases, not to match their selection to the precedent. This was true of the reference condition, where the same speaker maintained a precedent label; precedent objects were matched in only 28.5% of these trials, which indicates that that the task may have been simply too difficult for participants. While the overall pattern of data is consistent with a comprehension system that is more sensitive to linguistic input *as linguistic input*, rather than as an indicator about the knowledge state of the speaker, it may be the case that a source of common ground in broken label trials would give rise to a different pattern of effects. In light of this consideration, this experiment is followed by a paradigm that incorporates a source of common ground that has not been explored in previous studies on conversational precedents: cultural copresence. The following chapter explores the relative extents to which speaker knowledge, linguistic precedent, and the copresence of culturally salient information are brought to bear on referential processing.

Chapter 4

Speaker-specific and cultural associations in reference

4.1 Introduction

According to the Social Adaptation Hypothesis (SAH) of linguistic comprehension, contextual knowledge constrains speakers' expectations of relevance and thus their interpretation of utterances. This contextual knowledge includes a hearer's individual knowledge about the world, and knowledge that is shared by a given community. The stable cultural knowledge particular to a group constrains the members' expectations of relevance in a similar way. The knowledge of what is typical for their group allows them to communicate according to what is optimally relevant. For example, if a parent in a postnatal group asks another if they have a "bottle", this will likely be interpreted to mean a baby's milk bottle on account of their shared community knowledge. However, if the same request is made at an all-adult monthly barbecue social group, the attendees are likely to interpret this to mean a beer. It would be surprising if the intended meanings in both of these scenarios were reversed; the information common to members of these groups makes one interpretation more likely than another. In this way, cultural knowledge is deployed according to the knowledge shared with our interlocutors, and constrains our interpretation of what speakers mean.

As briefly mentioned in §2.2, Clark proposed a set of *copresence heuristics* that hearers could use to establish common ground in order to comprehend and produce utterances (H. Clark & Marshall, 1978, 1981). These heuristics pertain to different sets of knowledge that may be shared with a given interlocutor. These sets of knowledge were categorised in terms of the strength of the evidence they could provide to hearers, with *physical copresence* constituting the strongest evidence for mutually shared knowledge. Physical copresence refers to items that are perceptually available to both interlocutors, such as a candle on a table between two people. That this is the case is *also* perceptually available to both interlocutors, and so both speaker and hearer can directly perceive that something is mutually known to them. *Linguistic copresence* concerns a similar set of mutual knowledge, but in a conceptual capacity; that which is established between the interlocutors during the discourse. In other words, if a candle has featured in your conversation, you can both have evidence that the candle is mutually known. This is weaker evidence for mutual knowledge than physical copresence, because it relies on the memory, attention, cooperation and rationality of the other interlocutor (H. Clark & Marshall, 1978). An even weaker set of evidence for mutual knowledge, a subset of linguistic copresence, is *indirect copresence*, whereby if a candle is established in the discourse, another related referent such as the wick is also indirectly copresent. The last set of knowledge that can be used to establish common ground is *cultural copresence*, whereby "some particulars are assumed to be universally known in a cultural milieu" (H. Clark & Marshall, 1978, p. 62), and thus certain sets of information are culturally copresent for everyone in a given community. This knowledge constitutes a critical backdrop for the comprehension of utterances between members of a cultural group.

Among the copresence heuristics, *cultural copresence* is not considered to belong to the same hierarchy of strength. Cultural copresence is relatively permanent, and the particulars take more time to become familiar and lose familiarity. This is in contrast with the other three sets of knowledge, which can be very fleeting and may only ever be present for one conversation. There are different scopes of cultural copresence. As

noted by Clark and Marshall: "the trick is to judge cultural milieus. Ann may think that she and Bob mutually realise that they are both high school graduates, or drug dealers, or ninteenth century history buffs [...] *her assumptions will change accordingly*. If her assessment is correct, her definite reference is likely to succeed, and if not, it isn't" (H. Clark & Marshall, 1978, p.62, emphasis added). In other words, the knowledge shared with a particular person, on the basis of cultural copresence with them, will set the assumptions that determine comprehension of their utterances. Culturally copresent knowledge is thus a more enduring source of constraint on interpretation than other forms of common ground, which may be more fleeting or pertain only to a single interaction. Culturally copresent knowledge is also deployed according to the knowledge state of the speaker on this view, as opposed to a set of egocentric assumptions that are deployed regardless of the knowledge that interlocutors share.

Language itself constitutes a cultural system, and Clark (1996) situates linguistic conventions alongside other forms of community knowledge. Language users are able to take a large amount of information for granted about other speakers of the same language, such as syntax, phonology, semantics, idioms, politeness, and so on. This can be considered as part of the communal common ground along with other sets of information. Cultural communities can comprise speakers of the same language, enthusiasts of the same hobbies, residents of the same areas, and so on. With each attribution comes a set of information that can be reasonably assumed of the individual (H. Clark, 1996). Community networks are complex and nested. For example, Ann could belong to a particular street, in a particular area of a particular town in a particular country, and at each level this designation confers different sets of specific knowledge. Other sets of knowledge may be correlated with one another, such as place of residence and language. The central point for Clark (1996) is that knowledge is not distributed equally across the population, that communities can be defined according to their access to particular sets of knowledge, and that a given pair of speakers can infer their common ground on the basis of belonging to this community.

Clark's view of linguistic conventions as one of many inventories of common

ground knowledge is broadly in line with the SAH view of linguistic knowledge. On this view, linguistic knowledge can be characterised as certain expectations of relevance. For example, to know the word meanings of English is to have expectations about how people typically use words, and these expectations govern linguistic interpretation alongside other relevant factors. In other words, knowing that "cat" means 'a domestic feline' in English is simply a statement about its probabilistic use among speakers of English. Some expectations may only pertain to a particular interaction; others may be stable expectations that we hold over a lifetime with all members of a speech community. These different grades of contextual information can be brought to bear on comprehension through inferential processing, and language simply constitutes a set of contextual information that is relevant for discourse.

This can be contrasted with the LAH view of language in one key respect: that linguistic conventions, and an individual's associative representations in general, constitute knowledge that can be coordinated and deployed in discourse without the explicit computation of a partner's knowledge-state. Where a context-driven model requires that an interlocutor's knowledge is inferred on the basis of what one may expect from the relevant cultural milieu, the LAH view suggests that it is possible, in principle, for an individual's knowledge to simply be deployed in response to the discourse on an associative basis. The work being done by inference and expectations of relevance in the context-driven model of the SAH is done here by the deployment of an individual's associative knowledge in response to particular sets of stimuli. For the LAH, linguistic conventions are a specialised form of this associative knowledge. The heart of the difference is this: for the SAH, linguistic conventions function as other cultural knowledge does, i.e. as a stimulus that enables inference about the speaker's knowledge; for the LAH, linguistic conventions are a stimulus that elicits a hearer's associated knowledge, and most of the time this knowledge will be similar to that of the speaker.

The different grades of common ground - knowledge from direct perception, knowledge from the discourse, and knowledge from the cultural environment - pertain to three different levels of analysis that are dealt with quite differently in a language-

driven view. Barr and Keysar (2005) delineate cognitive, interactive, and cultural "domains of language use" that lend themselves to different considerations, shown in table 4.1 below.

Domain	*Social Unit*	*Language Unit*	*Characteristic Processes*	*Theoretical Import*
Cognitive	Individual	Clauses or utterances	Judgment and decision making Attention Memory Categorization	Places constraints on computation
Interactional	Dyad or group	Conversational turns	Epistemic exchanges Grounding (Clark & Brennan, 1991) Multimodal communication	Promotes shared understanding among the dyad or group
Cultural	Community	Languages	Establishment or diffusion of conventions	Promotes commonality of semantic representation among community members

Table 4.1: Table from Barr and Keysar (2005), which delineates domains of language use; of particular interest here is the *cultural* domain, which functions by virtue of the similarity of internal representations that all members of the speech community have. This conformity means that egocentric access to meanings can be deployed successfully. Note here that this analysis does not concern more general cultural knowledge as in Clark's cultural copresence; rather, it pertains specifically to the semantic representations of language.

At the group level of analysis, the knowledge shared by a group of language users serves to enable their egocentric access to representations by making the representations of those individuals more similar to each other. Note that this analysis does not concern more general cultural knowledge, as in Clark's cultural copresence; rather, it pertains specifically to the semantic representation of language. The underlying principle of this more egocentric model is that the use of individual knowledge and conventions is favoured by hearers, and that this strategy is facilitated by the typical domains of language use outlined in Fig. 4.1. Thus, the comprehension system may routinely deploy conventions known to the individual regardless of the particular knowledge state of their interlocutor; if the strategy is unsuccessful, the typical discourse environment provides opportunity for feedback, and so any systematic error

that results from this strategy is surmountable. This is a different mechanism than the deployment of cultural associations in response to the speaker's state of knowledge; instead, the cultural knowledge shared with a speaker may be drawn on in the event that egocentrically derived interpretations anchored in linguistic convention fail.

Specialisation for language, as proposed by the LAH, would entail special attention to linguistic conventions in the course of comprehension. As proposed above, the availability of linguistic conventions in the discourse provides a means of anchoring comprehension according to the egocentric associations held by the hearer. A system that is specialised to attend to linguistic input may not preferentially rely on the knowledge state of the speaker in interpretation when the typical domains of language use provide opportunities for correction. The typical domains of language use outlined above are precisely the sort of linguistic environment that would select for a language-driven cognitive strategy over phylogeny.

There is a contrast in the sets of available information that the SAH and LAH predict hearers will be most sensitive to in the course of interpreting utterances. Specifically, the LAH predicts that linguistic conventions will exert an effect on the interpretations the hearers make, and how long it takes them to do so. By contrast, the SAH predicts that culturally copresent information with the speaker will exert an effect on the same behavioural measures. The central question in this chapter is: does the comprehension process exhibit more sensitivity to the precedents available to the hearer, or to the knowledge that is culturally copresent with a speaker?

4.2 Motivation

While studies have explored the effect of physical and linguistic copresence on precedent comprehension, the effect of cultural copresence remains to be seen. The existing body of experimental literature on conversational precedents typically manipulates the physical copresence of referents by making particular objects in an array visible to either the hearer alone, or to both hearer and speaker. Linguistic copresence is manip-

ulated by using either speakers who have used particular referring expressions with the hearer before, or new speakers who do not share the same discourse history. This basic prototypical paradigm accounts for the major psycholinguistic studies of precedent use in recent years (Kronmüller & Barr, 2015, see also Chapter 3). The pattern of results so far indicates that listeners are not as sensitive to the physical copresence of referents as they are to the consistent use of referring expressions. It is only when these referring expressions are used inconsistently that the accumulated history of the speaker appears to have a bearing on comprehension (Kronmüller & Barr, 2015). The effect of cultural copresence on comprehension has yet to be explored in this body of work.

The results of Experiment 1 in §3 were consistent with the predictions of the LAH, showing a main effect of Label on reaction time. A key condition for the contrastive predictions of the LAH and SAH was the Different Speaker Maintain Label condition; though a different (i.e. naive) speaker used a speaker-specific label that had been established by someone else, participants still used the established convention to determine their comprehension, and this consistent use conferred a faster response than conditions in which a label was broken. One reason for this may have been that the radical ambiguity of the task led participants to rely on the linguistic convention in lieu of any other evidence. In other words, all objects in the array were, in principle, equally plausible referents for any given speaker. This may have meant that the linguistic convention was the 'weightiest' factor upon which to condition a comprehension decision. In usual discourse environments, other constraints such as cultural common ground may provide another source of evidence that hearers may exploit. Culturally copresent knowledge may provide a necessary background against which hearers can comprehend speaker-specific utterances.

The experiment presented in this chapter made use of the paradigm in Experiment 1 §3, with the additional manipulation of cultural congruence. The referential matching task used arrays of objects, one of which was a culturally salient object (such as a model of the Eiffel Tower). Either a 'neutral' object in the array, or the culturally salient object, were referred to with a speaker-specific construction such as "this reminds me

of my niece", or "this reminds me of France". In the second phase of a trial, the hearer was asked to select an object with the phrase "which one reminds me of France?" by either the same speaker or a different speaker. The experimental conditions are also shown in table 4.2, §4.3.4.

For example, in trials where the same speaker broke their precedent label for a culturally salient object, the miniature Eiffel Tower would be referred to with "this reminds me of my brother" in the establishment phase. In the test phase, the same speaker would ask the participant to "Click on the thing that reminds me of France". All prompts in the experiment were culturally salient. If participants are using inference about the knowledge state of the speaker in order to comprehend referring expressions, cultural copresence will allow hearers to select the most congruous referent. However, if hearers are relying on the conventions most available in the discourse, they may select the neutral referent regardless. The relative weight placed on precedent maintenance, the knowledge shared with the speaker, and the cultural congruence of the precedent, was measured with behavioural outcome measures. In short, the manipulation of a preferentially weighted domain for comprehension will exert effects on the interpretation participants make, and how long that interpretation takes.

In other words, the established cultural relationship between some salient objects (a model of the Eiffel Tower), and a referring expression ("give me the thing that reminds me of France"), supplies hearers with culturally copresent common ground for interpreting the utterances. Whether and how they use this common ground is the primary point of contrast between the LAH and the SAH. The same behavioural outcome measures as experiment 1 (§3) were taken.

Speaker-specific referring expressions such as "This reminds me of France" are still, in principle, compatible with all objects in the array; after all, memories can be very idiosyncratic. However, if one of the objects is a model of the Eiffel Tower, the cultural milieu to which native English speakers are generally exposed will make this the most 'viable' referent when a second speaker uses the cultural referring expression. If a more neutral item, (e.g. a cup) reminds a first speaker of France, we may expect

that it would be a less viable referent when a new speaker asks for an object that reminds them of France.

In other words, participants are presented with trials where their choice of referent can either ultimately rest on assumptions of cultural copresence they make about a new English speaker, or they can rest on the precedent that has been set in the task. For this reason, a key contrastive condition for this experiment concerns trials where a precedent label (i.e. "reminds me of France") for a neutral precedent object (e.g. a ladle) is maintained by a different speaker. If culturally copresent information is employed on the basis of the new speaker, the hearer ought to select the miniature Eiffel Tower from the array. If they anchor their comprehension in their association with the linguistic precedent, however, they ought to select the ladle.

4.2.1 Predictions

The Social Adaptation Hypothesis (SAH) of comprehension holds that referential interpretation is based on the most relevant contextual information for the task, including knowledge state of the speaker. In the following experiment, the identity of the speaker forms a component of the referring expression itself, and so would confer a strong constraint on comprehension on this view. In addition, the more enduring and relatively permanent nature of culturally copresent information generates expectations of relevance that facilitate interpretation of utterances by others in the milieu - in this case, the broad cultural knowledge expected of native speakers of British English. On this view, the availability of culturally copresent information allows hearers to deploy common ground in order to resolve the kind of ambiguities that otherwise arise from speaker-specific referring expressions, as in Experiment 1.

All prompt phases in the following experiment employed culturally salient associations, such as "Which one reminds me of France?". The SAH view of comprehension holds that common ground constrains all utterance interpretations, and that this is the fundamental backdrop of comprehension in all cases. The SAH predicts a main effect

of speaker on reaction time and an interaction with label, whereby responses are faster when a Different speaker breaks a precedent label than when the Same speaker does so. This is because hearers have pragmatic expectations about the consistency of label use, and so a speaker breaking their own precedent causes more disruption to the inferential process than a different speaker using a different label. The SAH therefore predicts a different speaker advantage for reaction time in the case of broken precedents.

A key prediction of the SAH is that, when a culturally salient, speaker-specific construction like "Which one reminds me of France?" is used by another speaker (i.e. a Different speaker Maintains a precedent label for a neutral precedent object), participants will select the cultural referent in all cases. This is because a culturally salient prompt like "Which one reminds me of France?" can be interpreted using culturally copresent information when uttered by a new speaker. In other words, the SAH predicts that cultural copresence will be leveraged in cases where the speaker is new, and thus predicts a main effect of speaker on whether participants select the cultural objects.

In contrast, the Linguistic Adaptation Hypothesis (LAH) of comprehension holds that hearers will preferentially rely on the association they have made between a referent and the linguistic label, with less regard for the identity or knowledge state of the speaker. The LAH therefore predicts a main effect of Label on reaction time, whereby broken labels elicit slower reaction times than maintained labels overall.

In particular, when a different speaker maintains a culturally salient label such as "Which one reminds me of France?", we might expect hearers to select the referent that has been associated with this label earlier in the discourse with a previous speaker, rather than selecting the referent one would infer from the cultural milieu. Therefore, the LAH predicts that participants will select the neutral objects in trials where a different speaker maintains a label with a neutral precedent object. This is because the maintained label has been previously associated with a referent, and this representation will be brought to bear on the interpretation of that same label, regardless of whether the speaker is in possession of appropriate knowledge. The LAH therefore predicts a

main effect of label and an interaction with precedent object on whether participants select the cultural object in the array.

The LAH predicts that cultural common ground becomes more important for comprehension when precedent labels are broken. Specialised attention to linguistic conventions in the discourse facilitates interpretation of utterances, and the comprehension system is more sensitive to this domain of input than to the knowledge state of the speaker. When a different speaker maintains a culturally salient label such as "Which one reminds me of France?", the LAH predicts that hearers will select the referent that has been associated with this label earlier in the discourse with a previous speaker, rather than selecting the referent one would infer from the cultural milieu. Specifically, the LAH predicts that in cases where the precedent object is Neutral, and a Different person Maintains the precedent label, that participants will continue to select the Neutral object, in line with the precedent. For broken precedents, comprehension will take longer even though the culturally salient prompt is always compatible with a cultural object in the array; this is because the LAH predicts that hearers will anchor their comprehension in the linguistic precedent, rather than in inferences about the knowledge available to the speaker. The LAH therefore predicts a main effect of label and an interaction with precedent object on how long it takes participants to select the cultural object in the array.

4.3 Experiment 2

4.3.1 Participants

46 native English speakers (14 male, median age 21) were recruited from the Edinburgh University campus via undergraduate and postgraduate mailing lists, and word-of-mouth. Participants were paid £3 each.

4.3.2 Design

The experiment was a within-subjects and items 2 x 2 x 2 design. This comprised the same factors as experiment 1: Speaker [Same, Different] where the speakers in the precedent and prompt phase of each trial were either the same speaker or a different speaker; and Label [Maintain, Break] whereby the referring expression used in the precedent and prompt phase of each trial was either maintained or broken. An additional within-subjects factor Precedent Object [Cultural, Neutral] was added, whereby the object being referred to in the precedent phase of each trial was either culturally salient or neutral. Participants were therefore exposed to all levels of Speaker, Label and Precedent Object. This design is also illustrated in table 4.2.

4.3.3 Materials

Visual Arrays:

Forty-eight arrays of images were used. Each array comprised four images, three of which were photographs of everyday objects as in §3.3, and one of which was a photograph of a culturally salient object that corresponded either to a well-known location (such as a country or capital city), or a well-known public holiday (such as Christmas). Culturally salient pictures were presented to 5 native English speaking postgraduate students at the University of Edinburgh, who were asked to circle which picture matched the referring expressions used in the audio. For example, from the list of culturally salient pictures, the survey asked which picture represented Cuba, and the respondents selected the picture of a cigar. All descriptions in the informal survey elicited correct matches to the pictures. The photographs were converted to 400 x 300 pixel black and white jpeg files, corrected for consistent contrast. An example array is presented in Fig. 4.1 (p.94).

Audio:

The audio stimuli were recorded in two sessions by two volunteer voice actors. To help ensure that participants would differentiate between speakers, the volunteers were one male and one female. The audio recordings for the precedent phase of trials consisted of either an idiosyncratic memory association with which to label an object (e.g. “This makes me think of Thursdays”) or a culturally salient referring expression (e.g. “This reminds of me France”). The prompts used culturally salient referring expressions (e.g. “Click on the thing that reminds me of France”). The idiosyncratic memory associations were restricted to interpersonal relations (such as “my sister” or “my nephew”), particular times or events (such as “my birthday” or “my job”), and places (such as “my shed” or “my hallway”), as in §3.3. Audio sets were edited and counterbalanced according to the order of male and female prompts in each condition. All audio files were edited using the Audacity®software package.

The factorial combination of within-subject items resulted in four lists, as in §3.3. Each item, consisting of audio prompts paired with visual arrays, was rotated through all four within-subjects experimental conditions. These items were then manipulated according to Precedent Object, resulting in a total of eight lists. Each participant saw each item in one condition only. The audio arrangement of the experimental conditions is illustrated below in table 4.2 (p.93).

Figure 4.1: One of forty-eight visual arrays used in Experiment 2, shown here twice. Establishment phase: one object in the array would be highlighted by a red box around its cell, accompanied by a referring expression in the precedent phase of each trial. In this example, the precedent object is Cultural; in other trials, the highlighted object is Neutral. Prompt phase: The highlight disappeared, and the array remained on screen while participants received an audio prompt to select one of the objects.

Condition	*Establishment phase*	*Critical prompt*
Neutral Precedent Object ***Same Speaker*** *Maintain Label*	[**Neutral Object** Highlighted] Speaker **A**: "This makes me think of **France**"	Speaker **A**: "Which one makes me think of **France?**"
Neutral Precedent Object ***Same Speaker*** *Break Label*	[**Neutral Object** Highlighted] Speaker **A**: "This makes me think of **my brother**"	Speaker **A**: "Which one makes me think of **France?**"
Neutral Precedent Object ***Different Speaker*** *Maintain Label*	[**Neutral Object** Highlighted] Speaker **B**: "This makes me think of **France**"	Speaker **A**: "Which one makes me think of **France?**"
Neutral Precedent Object ***Different Speaker*** *Break Label*	[**Neutral Object** Highlighted] Speaker **B**: "This makes me think of **my brother**"	Speaker **A**: "Which one makes me think of **France?**"
Cultural Precedent Object ***Same Speaker*** *Maintain Label*	[**Cultural Object** Highlighted] Speaker **A**: "This makes me think of **France**"	Speaker **A**: "Which one makes me think of **France?**"
Cultural Precedent Object ***Same Speaker*** *Break Label*	[**Cultural Object** Highlighted] Speaker **A**: "This makes me think of **my brother**"	Speaker **A**: "Which one makes me think of **France?**"
Cultural Precedent Objectt ***Different Speaker*** *Maintain Label*	[**Cultural Object** Highlighted] Speaker **B**: "This makes me think of **France**"	Speaker **A**: "Which one makes me think of **France?**"
Cultural Precedent Object ***Different Speaker*** *Break Label*	[**Cultural Object** Highlighted] Speaker **B**: "This makes me think of **my brother**"	Speaker **A**: "Which one makes me think of **France?**"

Table 4.2: The composition of audio prompts for each experimental condition, according to manipulations of *Precedent Object*, *Speaker* and *Label*. Note that the prompt phase audio always uses the same referring expression across conditions; this ensures that reaction times are comparable.

4.3.4 Procedure

All experimental stimuli were administered using E-Prime software (Psychology Software Tools Inc., 2014).

Briefing and instructions:

Participants were tested individually, in acoustically deadened booths. They received a scripted verbal briefing from the experimenter, which contained a cover story about what the experiment entailed. The briefing was that previous participants of a different task had brought in everyday objects from home, for which they had an associated memory that could be summed up in one word. These previous participants had used

the objects to play a memory game with each other. Participants were then told that they were about to hear the recordings of two such players, in the interest of seeing how they performed without the players being present. It was impressed upon participants that the two players they were about to hear had never met each other, that they had in fact played the game on different days, and that their recordings had merely been brought together to construct the current task. Participants were then told how to use the Serial Response Box.

The trial instructions were delivered verbally, and repeated by on-screen instructions that participants could go through at their own pace. Participants were told they would see four objects on the screen in front of them. It was explained that a red frame would highlight an object and at that time the participant would hear the owner's memory association with it. The red frame would then disappear. It was explained that the participant would then be asked to select an object from the array using the Serial Response Box buttons in front of them, and that there was a time limit of 4 seconds. Participants were told that if they did not know which selection to make, they could simply guess.

Practice and Comprehension check:

In order for participants to become familiar with the button-screen configuration, participants were given a practice round with counterbalanced male and female instructions. The practice round consisted of four trials that followed the critical trial procedure. After the practice round, participants were asked three comprehension check questions (target answers given in italics): Can you tell me what the red frame is for? *It highlights the object being talked about*; Did you hear an instruction to select an object before you did so? *Yes*; Did the two players you were listening to know each other? *No*. If the comprehension questions were answered incorrectly, the instructions, practice and checks were repeated to ensure the participant understood before proceeding to the test trials.

Critical trials:

Each trial consisted of a precedent phase, immediately followed by a prompt phase.

Precedent phase: Participants were presented with a new array of four objects for each new trial. A red frame then highlighted one object in the array. This highlighted object was either a Culturally salient object, or a Neutral object, depending on the manipulation of Precedent Object. This was accompanied by a precedent referring expression via audio. In the Maintained Label condition, the accompanying referring expression would associate the object with a culturally salient location (e.g. "This makes me think of France"), and in the Broken Label condition this precedent would be idiosyncratic (e.g. "This reminds me of my sister"). After this audio was delivered, the frame disappeared and the object array remained.

Prompt phase: The participant would then hear either the same speaker, or a different speaker, prompt the participant to select an object using a culturally salient location as a referring expression (e.g. "This makes me think of France"). Upon hearing the prompt, the participant then used one of the four Serial Response Box buttons to select one of the four objects in the array. The participants were not given feedback about whether they had selected the correct object; the game simply proceeded after each selection.

After completing the experiment, participants were invited to fill in an exit questionnaire in order to gauge what they thought the experiment was about, and what their answering strategy was.

4.3.5 Dependent variables

Reaction Time

Participants' responses were made via a Serial Response Box, the reaction time (RT) of which was recorded by E-Prime software (Psychology Software Tools Inc., 2014). For each trial, the audio prompt played over a slide that did not allow any response input

from the participant. This ensured that no response could be made without hearing the prompt, to prevent the level of data loss incurred in §3.3. After the audio prompt, the array remained on screen during a new slide which allowed response input from the participants via the Serial Response Box. The response slide terminated upon the participant's response, or upon the exhaustion of the imposed time limit. When participants did not respond before the time limit, no data was logged. No responses were permitted as input until the audio prompt was completed.

Precedent Match

The red highlight around the precedent object in each array was random for each trial, but was pre-set in the experiment code. The participant responses also corresponded to a cell in the array. Both values were logged for each trial, and it was therefore possible to calculate whether or not the object selected in the prompt phase matched the object highlighted in the precedent phase.

Culture Object Match

The cell containing the culturally salient object for each array was random for each trial, but was pre-set in the experiment code. The participant responses also corresponded to a cell in the array. Both values were logged for each trial, and it was therefore possible to calculate whether or not the object selection in the prompt phase matched the culturally salient object or not.

4.4 Results

4.4.1 Reaction Time

I used R (R Core Team, 2013) and LME4 (Bates et al., 2015) for a linear mixed effects analysis with the dependent variable Reaction Time (RT). As fixed effects, I entered Label, Speaker and Precedent Object (with interactions) into the model, using Same

speaker, Maintained label trials with Cultural precedent objects as the reference level. I entered a maximal random effect structure (Barr et al., 2013) of random intercepts for subjects and items, and by-subject and by-item random slopes for Label, Speaker and Precedent Object. In some cases these models failed to converge, in which case the random effects structure was simplified. P-values were obtained by likelihood ratio tests of models with the effect in question against models without the effect in question.

There was a significant main effect of Label ($\chi^2(1)$= 38.08, $p < 0.001$) on reaction time, reflecting an increase of 912.49ms $\pm$115.66(SE) overall when a label was broken. There was no main effect of Speaker ($\chi^2(1)$ = 0.97, p = 0.33), and no main effect of Precedent Object ($\chi^2(1)$ = 0.9, p= 0.34).

There was a significant interaction between Label and Precedent Object ($\chi^2(2)$ = 13.66, p= 0.001) on reaction time, whereby broken labels in trials with neutral precedent objects were associated with an increase of 622.47ms $\pm$ 176.31(SE), compared to broken labels in trials with cultural precedent objects.

There was also an interaction between Label and Speaker ($\chi^2(2)$ = 6.3, p= 0.043), which reflected a different speaker processing benefit of 406.64ms $\pm$184.54(SE) when labels were broken.

A three-way interaction of Label, Precedent Object and Speaker was on the conventional threshold for statistical significance ($\chi^2(4)$ = 9.09, p= 0.059). In other words, an interaction between label and speaker varied according to whether the precedent object was neutral or cultural by 494.15ms $\pm$ 306.16(SE).

Participants gave the quickest RTs when a label for a cultural object was maintained by the same speaker. When a label for a cultural object was *broken* by the same speaker, however, participants gave the slowest RTs. This pattern can be seen in Fig. 4.2 below.

4.4.2 Object Selection

The cells selected in the critical trial were automatically coded by whether they matched the cell highlighted in the precedent trial, and by whether they matched the cultural ob-

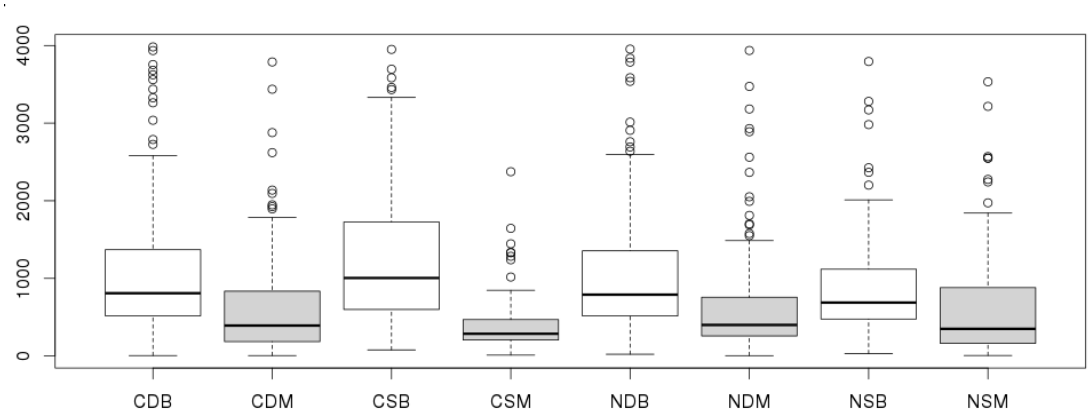


Figure 4.2: Reaction Time (ms) by condition (Speaker, Label, Precedent Object): CDB = Cultural precedent object, Different speaker, and Broken label; CDM = Cultural precedent object, Different speaker, and Maintained label; CSB = Cultural precedent object, Same speaker, and Broken label; CSM = Cultural precedent object, Same speaker, and Maintained label; NDB = Neutral precedent object, Different speaker, and Broken label; NDM = Neutral precedent object, Different speaker, and Maintained label; NSB = Neutral precedent object, Same speaker, and Broken label; NSM = Neutral precedent object, Same speaker, and Maintained label.

ject present in the array (Matches = 1) or not (Non-matches = 0). All analyses of the target selection data were carried out using R (R Core Team, 2013) and LME4 (Bates et al., 2015) for logistic mixed-effects regressions with the match data as dependent variables and Label, Speaker and Precedent Object (with interactions) as fixed effects. The models used a maximal random effect structure (Barr et al., 2013) of by-subject random intercepts, and by-subject random slopes for fixed effect manipulations, and used Non-matches as the reference level. In some cases these models failed to converge, which was resolved by removing interactions between random slopes. In these cases, the non-converging models indicated the same pattern of effects and significance as the reduced random effects models.

Two object selection measures were derived from the data, both describing different aspects of the selection made by participants during the test phase of each trial. Precedent Object Selection is a measure of whether participants selected the object that corresponded to the precedent label in the establishment phase (match) or selected a different object to the precedent highlight (nonmatch). Cultural Object Selection is a measure of whether participants selected the cultural object in the array (match) or selected a neutral object in the array (nonmatch).

Precedent Object Selection

There was a significant main effect of Label; when a label was broken, the likelihood of participants matching their response to the precedent object decreased (β = -3.61, SE = 0.29, p $<$ 0.001). There was also a significant effect of Precedent Object (β = -4.22, SE = 0.31, p $<$ 0.001), whereby the likelihood of participants matching their selection to the precedent object decreased when the object was Neutral.

The histograms below indicate notably fewer matches when labels were Broken in Neutral object trials and notably more matches when labels were Broken in Cultural objects trials, which suggests an interaction between Label and Precedent Object. Model convergence for an interaction between Label and Precedent Object was unattainable despite simplified random effects terms, with unreliable results. However, the Cultural Object Match data below overlaps with the Precedent Object Match data, and offers statistical insight about how both variables are interacting, and supports the trends illustrated in Fig. 4.3.

There was no significant effect of Speaker (β = 0.39, SE = 0.23, p = 0.091), and no interaction between Speaker and Label (β = 0.16, SE = 0.74, p = 0.83) or Speaker and Precedent Object (β = 0.47, SE = 0.75, p = 0.53).

Cultural Object Selection

As we would perhaps expect, there was a significant main effect of Precedent Object (β = -6.29, SE = 0.77, p $<$ 0.001); when the precedent object was Neutral, the likelihood that participants selected the cultural object in the array decreased. There was a significant effect of Label (β = -3.98, SE = 0.74, p $<$ 0.001), whereby the likelihood that participants selected the cultural object tended to decrease when a label was broken.

There was a significant interaction between Label and Precedent Object (β = 7.19, SE = 0.83, p $<$ 0.001). Though the likelihood of selecting the cultural object decreased overall when either a label was broken or the precedent object was neutral, when both of these conditions applied (that is, when a label for a neutral precedent object was

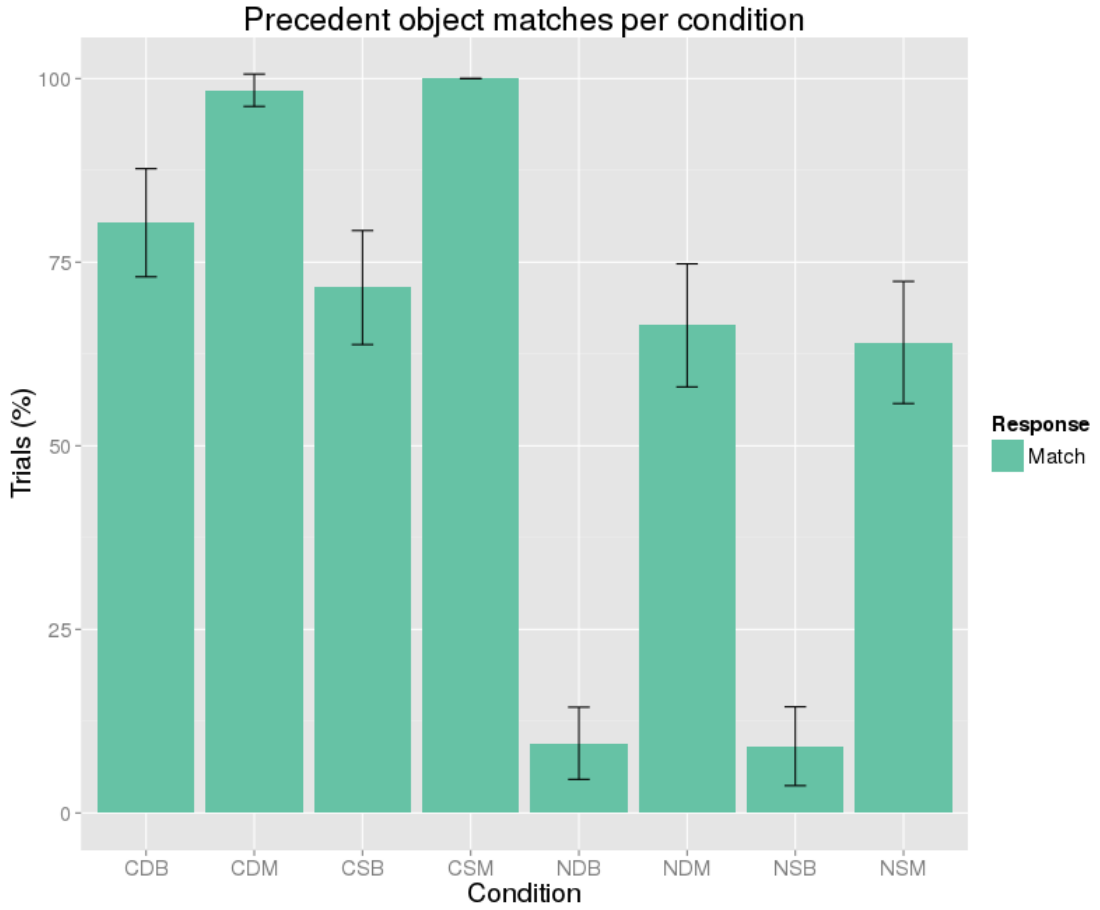


Figure 4.3: Proportions of trials where participants matched their object selection to the precedent object, according to Label, Speaker and Precedent Object: CDB = Cultural precedent object; CDM = Cultural precedent object, Different speaker, and Maintained label; CSB = Cultural precedent object, Same speaker, and Broken label; CSM = Cultural precedent object, Same speaker, and Maintained label; NDB = Neutral precedent object, Different speaker, and Broken label; NDM = Neutral precedent object, Different speaker, and Maintained label; NSB = Neutral precedent object, Same speaker, and Broken label; NSM = Neutral precedent object, Same speaker, and Maintained label. Error bars show 95% confidence interval.

broken) the likelihood of selecting the cultural object increased. Participants were most likely to match their selection to the cultural object when labels for cultural precedent objects were maintained, and did so in almost 100% of trials. This pattern is illustrated in Fig. 4.4 below.

There was no significant effect of Speaker (β = 0.58, SE = 0.56, p = 0.31), and no interactions between Speaker and Label (β = 0.18, SE = 0.74, p = 0.81) or Speaker and Precedent Object (β = 0.66, SE = 0.80, p = 0.41).

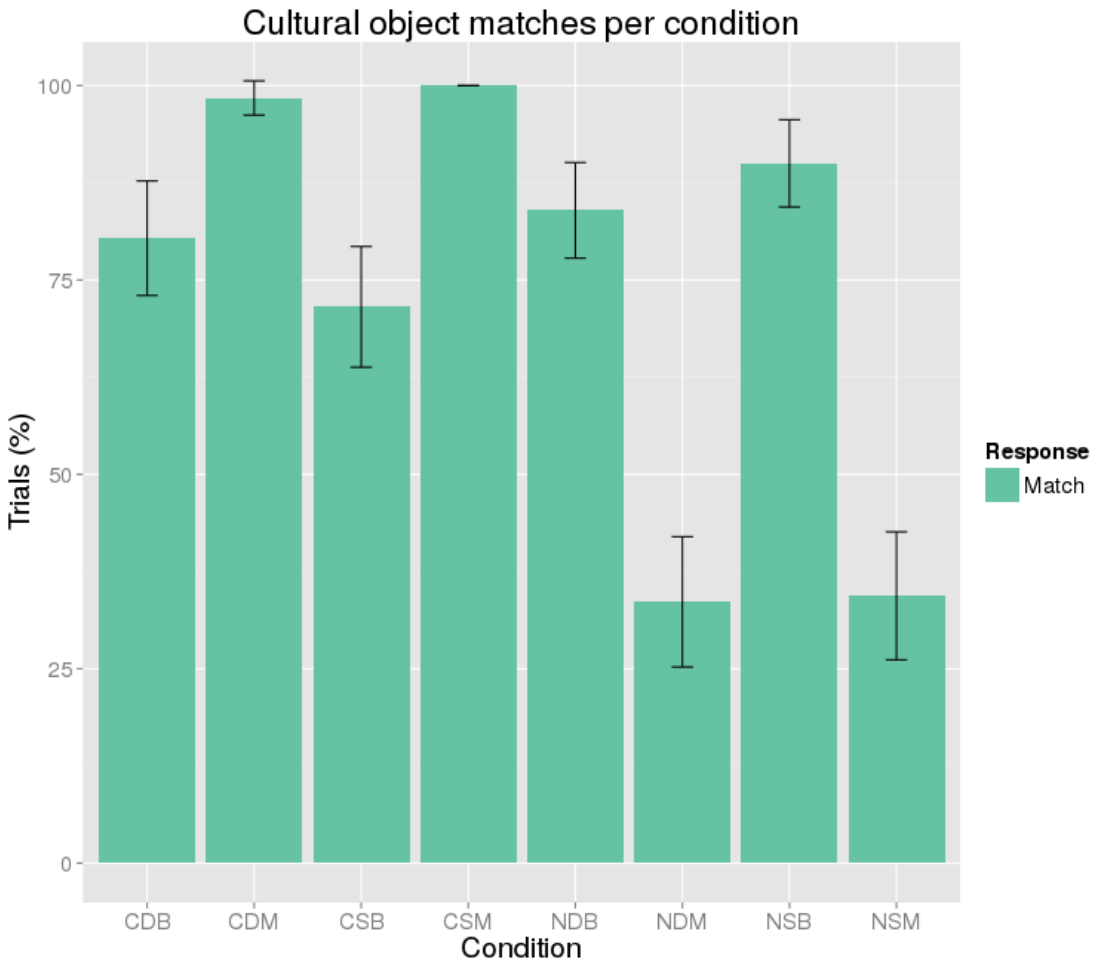


Figure 4.4: Proportions of trials where participants matched their object selection to the Cultural object, according to Label, Speaker and Precedent Object: CDB = Cultural precedent object; CDM = Cultural precedent object, Different speaker, and Maintained label; CSB = Cultural precedent object, Same speaker, and Broken label; CSM = Cultural precedent object, Same speaker, and Maintained label; NDB = Neutral precedent object, Different speaker, and Broken label; NDM = Neutral precedent object, Different speaker, and Maintained label; NSB = Neutral precedent object, Same speaker, and Broken label; NSM = Neutral precedent object, Same speaker, and Maintained label. Error bars show 95% confidence interval.

4.5 Discussion

The LAH predicted that the selection of a referent and the reaction time of response is more sensitive to the linguistic conventions in a discourse than to the knowledge state of the speaker. The reaction time data exhibited a main effect of label, and no main effect of either speaker or precedent object. This is consistent with the predictions of the LAH, and indicates that the comprehension system is more sensitive to linguistic input than to the knowledge state of the speaker and cultural copresence overall.

Maintained labels elicited faster reaction times than broken labels overall, regardless of who delivered the label or how congruous the label was with expectations conferred by cultural knowledge. Precedent labels maintained by different speakers were processed as efficiently as those maintained by the same speaker when the precedent object was neutral. However, when the precedent object was culturally salient, there was a same speaker advantage for maintained labels, and a different speaker advantage for broken labels. The effects present in the reaction time data could be concisely summarised thus: hearers prefer consistent labels and disprefer broken labels, *especially* when they're delivered by the same speaker about culturally salient referents.

One interpretation of the same speaker advantage in the case of maintained labels for culturally salient objects may be that episodic effects provide an additive benefit to processing, whether this is consistency of speaker, label or that the cultural associations being used are consistent with person's existing knowledge. That the Cultural trials in particular provided a magnifying backdrop for the effects of maintained and broken labels is perhaps consistent with a relevance-oriented explanation of comprehension, where expectations of relevance function to 'shortcut' hearers to an appropriate interpretation. A violation of these expectations may be responsible for the corresponding slow reaction times in trials where the same speaker broke their label for a cultural object. This interpretation would suggest that this effect is inconsistent with the LAH.

Experiment 1 indicated that, in the absence of any lexical constraints and copresent information, different speakers cannot provide any beneficial constraint on comprehen-

sion. Indeed lexical knowledge itself is a form of copresent information; conventions are so precisely by virtue of being shared. The experiment presented in this chapter illustrates that hearers deploy culturally copresent sources of nonlinguistic shared knowledge in the absence of consistent linguistic conventions, in much the same way that physical copresence impacts the processing of conversational precedents in the existing experimental literature. Cultural knowledge can constrain comprehension in these cases by providing hearers with a useful 'default' value for new speakers.

Despite this, however, participants did not appear to deploy this default cultural knowledge for new speakers in the case of maintained labels. Maintained labels delivered by a different speaker did not elicit slower reaction times. More surprisingly, participants also continued to choose the precedent object in the majority of trials where the cultural label ("This reminds me of France") was maintained, even though the precedent object itself was Neutral (e.g. a ladle) and there was a culturally congruous object (e.g. the Eiffel Tower) elsewhere in the same array. Both of these results are inconsistent with the predictions of the SAH.

The overall pattern of object selection effects indicated that participants matched their selections to the precedent object when labels were maintained, and were more likely to exclude the precedent object when a label was broken, regardless of the speaker. However, there was a significant interaction between label and precedent object on precedent object matches: in trials with a cultural precedent object, participants selected the precedent object when labels were broken for the majority of trials. This makes sense: all test phase prompts were culturally salient ("Which thing reminds me of France?"), and so when labels were broken, the preceding label was of a neutral type ("This reminds me of my niece"). In broken label trials across all conditions, participants mostly chose the cultural objects, and so the deployment of culturally copresent information provides a useful source of disambiguation when labels were not used consistently, even if that meant choosing an object that has already received a different precedent label. In other words, participants accepted more than one label for cultural referents when they were able to deploy cultural associations in the event of a

broken label - however, this also elicited much longer reaction times. This may suggest a more compensatory role for cultural common ground in comprehension, such that it is employed when labels are broken, and is the result of a lengthier process.

4.5.1 Limitations and future directions

The design of the experiment meant that prompts for object selection were always culturally salient (*"Which thing reminds me of France?"*). In broken label trials, the label assigned to the object was neutral (*"This reminds me of my niece"*). This increased the ease of the task, as participants found neutral prompts unresolvable in experiment 1 (Chapter 3). The use of culturally salient prompts ensured there was a 'viable' referent in each array. This prevented the level of data loss incurred by experiment 1. The results confirmed that the presence of culturally copresent information is a useful resource for resolution in the case of broken labels.

However, this design also meant that the first referring expression and the prompt referring expression referred to different objects in broken label trials when the precedent object was neutral. This presents us with an alternative explanation for the magnifying effect that cultural object trials appeared to have on the cost of broken labels: mutual exclusivity. To illustrate, consider SBC trials, where the Eiffel Tower reminds speaker A of their niece. The participant is then prompted to select an object that reminds speaker A of France, and selects the Eiffel Tower with relatively slower reaction times on account of the new label for the old referent. Compare this to SBN trials, where a neutral object, such as a glove, reminds speaker A of their niece. The participant is then prompted to select an object that reminds the speaker A of France, and selects the Eiffel Tower with faster reaction times compared to the SBC trials. The use of a different label for the same object in SBC trials violates the hearers' expectations of consistent reference and mutual exclusivity - in other words, 'a new label means a new thing'. This cost is not incurred by SBN trials, where the precedent and prompt do not obviously refer to the same target. While broken labels still elicited slower reaction

times than maintained labels across conditions, the exaggeration of this effect in trials with cultural precedent objects may be underpinned by mutual exclusivity effects, rather than relevance-oriented effects about the congruence of cultural information.

In other words, rather than indicating that hearers experience a processing benefit when references are congruous with the expectations they have because of cultural copresence, it may be the case that hearers simply prefer repeated references to the same object, and dislike inconsistent references to the same object. However, their determination that the reference *was indeed for* the same object is perhaps the conceptual point that should be emphasised here. In principle, anything could remind a person of anything (as demonstrated by the practice and critical trials of the experiment itself), but the cultural milieu provides a strong enough constraint on interpretation that can be used in lieu of linguistic precedent with the speaker.

It may be more difficult to determine whether this preferred congruence with cultural knowledge in referential processing is an effect of deploying shared knowledge that is known to be shared, or is an egocentric effect, whereby the hearer's knowledge provides a processing benefit regardless. One future direction of study that could interrogate this difference further may take the form of a large scale task in a virtual environment over longer periods of time, where participants form groups that acquire group-specific knowledge and then must communicate between groups.

Another criticism, which may address the lack of a main effect of Speaker on reaction time or object selection, is the use of a non-interactive paradigm. It could be argued that the participants in the experimental condition, where audio stimuli is prerecorded, may interpret all of the instructions as coming *'from the experiment'* rather than *'from different speakers'*. If this were the case, however, then the presence of interactions with speaker in the reaction time data would require another explanation, which is not forthcoming. An interactive replication of the experiment presented in this chapter would be able to test whether the effects reported hold up in more naturalistic discourse environments, where common ground may be a more salient and relevant parameter from the participants' point of view.

4.6 Conclusion

This chapter has presented an overview of how different domains of knowledge are conceived, in terms of common ground and in terms of representations that pertain to different levels of the discourse environment. Cultural copresence was introduced to a reaction time study of precedent effects to interrogate the contrastive hypotheses of the Linguistic Adaptation Hypothesis (LAH), and the Social Adaptation Hypothesis (SAH) of comprehension. The manipulations were Label, Speaker and Precedent Object. During the task, either a neutral or culturally salient object was referred to by a particular label, and then referred to again either by the same label or a different one, by the same speaker or a different speaker. Participants then had to select the correct referent. All referring expressions used in the tasks depended on the knowledge of the speaker, using similar constructions to experiment 1 (§3.3) of the format "This makes me think of..." and "This reminds me of...". The aim of this method was to strengthen the constraint that common ground may exert on referential comprehension by rendering all objects in the referential task equally plausible unless participants made recourse to shared knowledge with the speaker, in this case, the knowledge that is made available in the cultural milieu. The LAH predicted a main effect of Label, and an interaction with Precedent Object. Contrastively, the SAH predicted a main effect of Speaker, and an interaction with Precedent Object.

A condition in which Maintained precedent labels for Neutral objects were delivered by Different speakers was a key point of contrast between both accounts: the LAH predicted faster reaction times for maintained label trials than broken label trials, and the selection of a neutral object. The SAH predicted that a change of speaker would result in slower reaction times, and the selection of a cultural object. The reaction time results exhibited a main effect of Label, whereby reaction times were faster for maintained labels than broken labels, and an interaction with both Speaker and Precedent Object. Participants responded to the referring expressions quickest when the same speaker maintained a label for a culturally salient object, and responded quicker over-

all when labels were maintained regardless of the speaker. This is consistent with the existing literature on conversational precedents, and with the predictions of the LAH. The cultural object selection data revealed no effects of Speaker, but significant main effects of Label and Precedent Object, and a significant interaction between the two. Perhaps the most surprising result concerned the cultural object selections for the key contrastive condition for the LAH and SAH: maintained (cultural) labels for neutral objects, delivered by a different speaker. In these cases, participants heard speaker A refer to a neutral object, such as a ladle, as a thing that "reminds [me] of France". When speaker B then asked the participant to "click on the thing that reminds me of France", participants selected the ladle. This is perhaps surprising, because the ladle was alongside a culturally congruous object - i.e. the miniature Eiffel Tower - in the array, and speaker B has no ostensible reason to associate a neutral object with France. This effect on object selection suggests that the comprehension system preferentially weights the hearer's own associations with the linguistic label. The overall pattern of data is consistent with a comprehension system that is more sensitive to linguistic labels in their capacity as associative input, rather than as an indicator about the knowledge state of the speaker.

An interesting feature of the data concerned the 'magnifying' effect of cultural salience; hearers responded quicker for maintained labels and slower for broken labels, and this effect was greater when the precedents referred to culturally salient objects like a model Eiffel Tower. One interpretation, in line with the SAH of comprehension, is that the motivating context (that is, the presence of a culturally salient object) gives rise to a particular interpretation in the discourse in the first instance; when this context is congruous with the label, comprehension is thus quicker, and when this expectation is violated by the label, there is a cost to processing. If the context fundamentally underpins all interpretation of linguistic labels, we might then also expect that new labels for old referents could aid processing as long as they are accompanied by a new, congruous context. The next chapter explores this idea in a series of experiments that manipulate the visual context around referents, and the precedent labels that refer to

them.

Chapter 5

Reference across visual contexts

5.1 Introduction

According to the Social Adaptation Hypothesis, and context-driven approaches to pragmatics and psycholinguistics more broadly, linguistic comprehension is fundamentally enabled by the context. So far, this thesis has looked at different forms of common ground as a special subset of the context that we may expect to play an important role in comprehension. In §2.1, an example taken from (Sperber, 2000) illustrated the constraint of relevant contextual information on comprehension with the scenario of Mary and Peter picking berries. Recall that Mary can eat the berries in a stylised way in front of Peter, and that this will communicate (rather than merely demonstrate) that they are good to eat. If Peter is curious about whether the berries taste good, and Mary knows this, she can appropriately communicate the same proposition simply by saying "Yes". Peter will interpret this utterance correctly if and only if *he knows that she knows about his curiosity*. This example is intended to highlight the central importance of mutual knowledge in linguistic comprehension; the utterance will only be successful if it is interpreted against the backdrop of this mutually known information. However, metapsychological information about speakers is only one special subset of the context, various aspects of which also may affect interpretation. The contextual knowledge of the hearer, on the SAH view, determines the interpretation they will

consider; in relevance-theoretic terms, the context will confer certain expectations of relevance. The visual context is a more directly perceptible aspect of the situation in which discourse occurs, and this nonlinguistic information may exert direct influence over referential interpretation.

The subject of visual context effects on reference resolution has been the subject of a large number of eye-tracking studies in psycholinguistics over the last twenty years. The typical set-up of these studies is such that participants hear utterances while looking at a visual display over a series of trials, while their eye movements are measured for analysis. A seminal study by Tanenhaus et al. (1995) showed the effect of visual context on the interpretation of syntactically ambiguous sentences. Participants were given an ambiguous 'garden path' instruction such as "put the apple on the towel in the box", or an unambiguous equivalent such as "put the apple *that's* on the towel in the box". During this instruction, participants were presented with different visual arrays to interact with. One array included an apple on a towel, a box, a pen, and a towel on its own. The other array was identical, except for a second apple in place of the pen. In the first array, the fixations reflected the syntactic ambiguity of the instruction almost entirely; incorrect fixations (upon the lone towel) occurred in response to ambiguous sentences, and correct fixations (upon the box) occurred in response to unambiguous sentences, with no fixations upon the pen in either case. However, the presence of an extra potential referent (i.e., the apple) in the second array led to initial fixations on the distractor apple in response to both instructions, which prohibited later incorrect fixations on the empty towel in trials with an ambiguous instruction. In other words, listeners' online processing of syntactically ambiguous utterances differed according to what was in their visual array. This result showed no measurable stage of sentence processing that preceded the listeners' attention to visual information. Rather, visual information was being processed in tandem with the linguistic utterance. This stood in contrast to contemporary 'two-stage' structural accounts of syntactic processing (e.g. Frazier, 1987).

Related observations about visual context effects on processing use this visual

world paradigm. Among adult participants, such studies tend to produce evidence against 'two-stage' processing of syntactic ambiguity, in the form of context effects on the early fixations of listeners (e.g. Snedeker & Trueswell, 2004; Chambers, Tanenhaus, & Magnuson, 2004). In a study by Trueswell et al (1999), which used a similar syntactic ambiguity task to Tanenhaus et al (1995), five year olds exhibited fewer context effects than adults; they relied more on syntactic information to disambiguate utterances and failed to make use of relevant contextual information (Trueswell et al., 1999), possibly because the latter is more variable and children therefore take more time to acquire proficiency with its use (Trueswell & Gleitman, 2004). Studies of pronoun, demonstrative and reflexive interpretation paint a complex picture of interaction between different levels of processing, with variable semantic effects that differ across anaphoric domains, rather than proceeding after a set period of syntactic processing (e.g. Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Arnold, 2001; Kaiser, Runner, Sussman, & Tanenhaus, 2009). These results are presented in favour of a 'lexical account' of syntax processing, as high-level information (such as lexical meaning) appears to influence the fixations of participants in the earliest moments, rather than structural elements of syntax taking priority in the time-course of processing. This falls within the broader set of constraint-based accounts of linguistic processing whereby multiple simultaneous factors pertain to linguistic processing in many variable, relevant ways.

This existing body of experimental literature is particularly useful for distinguishing between parallel and serial models of processing. The nature of time-course data means that predictions about which information a language user attends to *first* can be tested under various conditions, and hence the serial 'two-stage' models of syntactic processing have been investigated against parallel 'lexical' models of syntactic processing in the visual world paradigms mentioned above. However, the question of whether a participant pays *serial* or *parallel* attention to utterances and partner perspective, or utterances and visual context, does not necessarily lend itself to questions about specialisation of function when it comes to comprehension and interpretation.

Other studies demonstrate this issue: upon hearing the preposition in the instruction "put the cube *inside* the can", participants restrict their gaze to containers large enough for a cube (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002). On this occasion, fixations were interpreted to represent the instant application of linguistic meaning to a particular visual array, rather than as an effect of the array on the derivation of meaning.

In many cases like this, it is difficult to know the extent to which fixations are the *application* of interpreted linguistic stimuli, or an *input to* the interpretation; is the search for apples evidence that the array is *constraining* the listener's interpretation of "put the apple in the box", or is this search the kind of response one would expect from a listener who had comprehended such an instruction in that particular environment? While discrete, serial stages of processing are more straightforwardly tested by eye-tracking data, there is a conceptual issue concerning the extent to which visual world eye-tracking can shed light on any core, context-invariant language processing that takes place in parallel models, particularly at the level of meaning comprehension (see Huettig, Rommers, & Meyer, 2011; Mitchell, 2004, p. 22).

This focus on the time-course of fixations is shared by the psycholinguistic literature on the effect of interlocutor knowledge (or 'perspective-taking') on the comprehension process. In particular, the study of conversational precedents has employed eye-tracking to address questions about whether a new or old speaker causes earlier disambiguation of new or old labels (see §2.2 and §3.1). These investigations have been especially concerned with the timing of speaker effects on fixations in on-line processing (e.g. Brown-Schmidt, 2009a, 2009b; Brown-Schmidt & Hanna, 2011; Kronmüller & Barr, 2015), and the interpretation of this kind of data can run into problems. For example, participants in Keysar et al's (2000) referential task fixated on items only visible to themselves at the earliest points of comprehension. Other objects were mutually visible with the participants' partner. This early fixation on privileged-ground items was interpreted to represent an early stage of egocentric processing. However, these early fixations may as plausibly be an appropriate response to being asked for objects that a partner cannot see, and represent common ground being taken into ac-

count rather than the early use of privileged information. The conversational precedent paradigm lends itself well to investigating the relative contributions of linguistic and nonlinguistic information to comprehension, but the effect of visual context on precedent comprehension has yet to be investigated.

The SAH and LAH represent parallel models of processing that make different predictions about the use of visual context and associative meaning in linguistic comprehension. For the SAH, an individual's associations with a linguistic stimulus are one of many parallel inputs to a central inferential process. This inferential mediation of all contextual and linguistic information is designed to produce an optimal interpretation with respect to the goal of understanding the speaker's intended meaning. In particular, inferential comprehension is fundamentally driven by contextual cues. The SAH predicts that the context facilitates comprehension to the degree that broken precedents should not pose a challenge to comprehension if they are compatible with the context. This is in contrast with the LAH, which holds that an individual's associations with a linguistic stimulus are, unlike other contextual inputs to comprehension, candidate outputs of the comprehension system. In particular, linguistic associations are a special domain of input for comprehension, and so broken precedents should incur a cost to processing regardless of contextual mediation. The functional outcomes of processing can be measured in order to evaluate these accounts; specifically, how participants interpret the reference, and how long it takes them to do so.

In this chapter I will investigate the effect of visual context on linguistic reference with two pairs of experiments. All four experiments use tangrams - sets of geometric shapes that can be rearranged into hundreds of different pictures - to augment the visual context in which referential precedents are used. Experiments 3 and 4 concern precedent comprehension for the same referent across visual contexts. In Experiment 3, the established label for a given referent was either maintained or broken across either the same context, or a different context. Experiment 4 followed up by either maintaining or breaking precedents across *abstract* contexts with no immediate interpretation. A statistical comparison of both experiments found that broken precedents

only produced slower reaction times for comprehension if they were not supported by an accompanying different, but congruous, context. Maintained precedents conferred a benefit to comprehension unless they were specifically incongruent with the context. This is broadly in line with the SAH, but further implications are discussed in §5.4.8. Experiments 5 and 6 concern label production for the same referent across visual contexts. In Experiment 5, participants labelled a referent under the guise of recording for a referential game. They labelled the same referent across either the same or a different context. In Experiment 6, participants had to label the same referent across an abstract context. A statistical comparison of both experiments found that speakers tended to maintain labels in the same contexts, and broke their precedent labels in different, incongruous contexts. In abstract contexts, speakers were more likely to maintain their labels than in different contexts, but less likely to maintain their labels than in the same contexts. Methodological issues with Experiments 5 and 6 accompany the discussion of results in §5.10.

5.2 Comprehension of reference across visual contexts

The SAH holds that linguistic communication is fundamentally driven by and dependent on the context in which it occurs. Inference makes use of a rich set of factors such as mutually manifest propositions and copresence, as well as broader nonlinguistic information about the situation of the discourse such as the visual context. In the relevance-theoretic account of comprehension, these factors together comprise the mutual cognitive environment - i.e. the relevant context of a given communicative exchange (Sperber & Wilson, 1987). This is the critical backdrop that enables comprehension. Recall that hearers interpret utterances as evidence for the speaker's meaning. This evidence is assumed to be optimally designed for the hearer in both a relevance-theoretic account (Sperber & Wilson, 2002; Scott-Phillips, 2014) as well

as context-driven traditions more generally (H. Clark & Carlson, 1981; H. Clark et al., 1983). In the standard relevance-theoretic model of processing, hearers converge upon the interpretation they expect to be of most relevance to them, and - crucially - these expectations are determined by the present context (Sperber & Wilson, 2002). Homonyms provide a very simple illustration of this idea: one may immediately interpret the word "bat" to concern some sporting equipment if the present context includes baseball or cricket, but in the context of animals or night-time one might assume "bat" to mean a flying mammal. Note that context of night-time could support an interpretation of 'sporting equipment' if other parameters made it so, such as the topic of a particular conversation or the hobbies of the speaker. The important point here is that the interpretation is determined from the outset by the situation in which the discourse occurs, and this includes the visual context available to the interlocutors. On this account, then, interpretations that are not supported by the context cannot occur; intended meanings that conflict with the context would require that the speaker change the context herself by providing further evidence.

By contrast, the LAH suggests that the comprehension system is specialised to make use of representations associated with a linguistic utterance, and that these associations can support linguistic comprehension without necessarily making recourse to other contextual cues. On this view, linguistic asssociations in general, and those established in a particular discourse, constitute a resource for the comprehension system. Furthermore, this comprehension system is specialised for linguistic input, such that the representations a hearer associates with the linguistic stimulus itself are able to support interpretation without contextual adjustment, at least some of the time. The LAH primarily differs from the SAH by suggesting that this is possible at all. The interpretation of linguistic precedents differs between both accounts in terms of where comprehension is anchored. For the SAH, comprehension is anchored in relevant contextual cues that are assessed by inference; for the LAH, comprehension can be anchored in conversational precedents, which provide a convenient source of interpretation for a system specialised for linguistic associations (Barr & Keysar, 2002).

One result from Experiment 2, discussed in §4.5, may suggest that contextual congruence is a significant factor in the comprehension of linguistic precedents. Participants heard a label established for a particular referent in an array, which was either culturally salient (such as an Eiffel Tower model) or neutral (such as a ladle). Participants were then instructed to select an object from the array with either the maintained precedent, or a new label (i.e. a broken precedent), delivered by either the same speaker or a different speaker. While participants' responses were slower with broken precedent labels overall, this was especially the case when the same speaker broke precedent labels that referred to culturally salient objects. This may suggest that the congruence of a linguistic label with the contextual cues of the discourse is a significant factor in comprehension, and may be preliminary evidence in favour of the SAH. In other words, this could suggest that when the expectations conferred by the context are not reliably indicated by linguistic labels, comprehension suffers.

However, this interpretation is far from conclusive. The interaction of the cultural salience of the precedent object with the label and speaker was marginal ($\chi^2(4) = 9.09$, p= 0.059), and though the magnitude of the effect was somewhat comparable to others at 494.15ms, it was much more variable (SE = $\pm$306.16ms). This may be an issue of statistical power that could be improved with more data. A more fundamental problem for this interpretation may lie in a potential confound introduced by the experimental design. Critical trials made use of cultural labels in all cases to ensure comparable reaction time calculations across conditions; for trials where precedents were broken, the establishment phase used neutral labels. When the precedent objects were cultural objects like the Eiffel Tower, this meant that the establishment phase and critical trial labels referred to the same item. However, in the comparable condition for neutral objects, these labels necessarily referred to different objects. This introduces the possibility that the 'magnifying' interaction of cultural salience for broken precedents may in fact be a mutual exclusivity effect. In other words, the disruption in comprehension may have been caused by broken labels for the same object, rather than by incongruence with cultural associations invoked by the salient cues of the object. In order to

further interrogate whether the contextual congruence is responsible, the same referent must be the target across all contexts.

Experiment 3 uses tangrams to study the effect of visual context on the comprehension of linguistic precedents. Tangrams consist of seven geometric shapes that can be rearranged to form hundreds of different pictures. The tangram format allowed for the consistent presentation of one particular shape across markedly different tangram configurations. Participants were exposed to an establishment phase, during which they identified a particular shape (e.g. a yellow diamond) with a particular label (e.g. "the flame") in a particular context (e.g. a candle configuration). During critical phases, the visual context was either repeated, or the tangrams were reconfigured into a different, incongruous arrangement (e.g. a pram configuration). Across both contexts, the target referent was presented in the same colour, size and orientation, and only the surrounding tangram configuration was changed. Participants then had to select the referent during a critical trial, after an instruction which either maintained the precedent label ("the flame") or used a new label that paired with the different context (e.g. "the wheel"). The object selections were measured, along with the reaction times for each response.

Experiment 4 follows up with a similar task. Participants were asked to select a referent shape in establishment trials and critical trials. The precedent label in critical trials was either maintained or broken using the same pairs of labels as Experiment 3, and the target referent across both contexts was presented in the same size, colour and orientation. However, the surrounding context display in each critical trial was an abstract arrangement of tangrams, with no easily discernible order. Note that this indistinguishable 'Abstract' context differs from the 'Different' contexts in Experiment 3, which had clearly indicated arrangements (such as a 'candle' versus a 'pram'). This allowed for a comparison between contexts of varying congruity with precedent labels. In the case of maintained labels, the abstract context condition also provided evidence for whether participants were indeed considering the same referent across contexts as *the same referent*, rather than two separate instantiations across displays.

The SAH holds that linguistic comprehension is fundamentally anchored in the context of a given communication. This predicts that maintained precedents can only aid comprehension when they are supported by the context. If they are supported by a different context, the SAH predicts that hearers will process broken labels as quickly as maintained labels in maintained contexts, and with similar success in target selection. If a maintained precedent is incongruous with the context, the SAH predicts that comprehension will be disrupted and therefore slower than in abstract contexts. In the case of abstract contexts, the SAH predicts that maintained linguistic precedents may determine the target selection, but that comprehension will be slower than in maintained contexts, as any anchoring in the nonlinguistic context cannot aid comprehension when this is taken away. Overall, the SAH predicts a main effect of Context on reaction time and target selection, and an interaction with Label.

The LAH view is that discourse typically provides opportunities for feedback and clarification, and so it is worthwhile to anchor comprehension in the linguistic precedents of a discourse. This predicts that maintained precedents can aid comprehension regardless of a change in context, and that broken precedent labels may incur a cost. Specifically, maintained precedents should lead to successful target selections and comparable reaction times in both repeated and different contexts, as well as in abstract contexts, leading to a main effect of Label on reaction times and target selection. The LAH also predicts an interaction with Context in the case of broken labels, where there may be a Different context benefit for reaction time.

Similarly to Experiments 1 and 2, trials where precedents are Maintained in a Different context are a key contrastive condition for the LAH and SAH. The SAH predicts that a label maintained in an incongruous context will incur a cost to processing and lead to slower reaction times and may lead to unsuccessful target selection. The LAH, on the other hand, predicts that maintained labels will be processed with similar speed across an incongruous context, and that target selection rates will be comparable to Same Label, Maintained Context trials.

5.3 Experiment 3

5.3.1 Participants

31 native English speakers (7 male, median age 23) were recruited from the Edinburgh University campus via undergraduate and postgraduate mailing lists, and word-of-mouth. Participants were paid £2.50 each.

5.3.2 Design

The experiment was a within-subjects and items design, with two factors of two levels: Context [Same, Different] whereby the context in which the referent appeared during the establishment phase would remain the same or become different for the prompt phase of each trial, and Label [Maintain, Break] whereby the referring expression used in the precedent and test phase of each trial was either maintained or broken. All participants were therefore exposed to a total of four experimental conditions, by the factorial combination of all levels of Speaker and Label. This is also illustrated in Fig. 5.2.

5.3.3 Materials

5.3.4 Visual materials:

There were 24 experimental items, each comprising a pair of tangrams. Each pair shared one target shape, whose colour and orientation was the same in both tangrams. The target shape had a different colour and shape combination for each pair, which meant that the target for each experimental item was unique (see also Appendix A). An example tangram pair is shown in Fig. 5.1.

Twenty-four additional tangram pictures were also included for 'filler' trials. This resulted in a total of 72 tangrams. Each tangram was also rendered in black to form a silhouette image. These were displayed during the first phrase of each tangram pre-

sentation, see also §5.3.6. Four shapes from each tangram, one of which was always the target shape, were rendered in a separate image file. Each of these four shapes corresponded to the four buttons of the Serial Response Box participants used to make a selection.

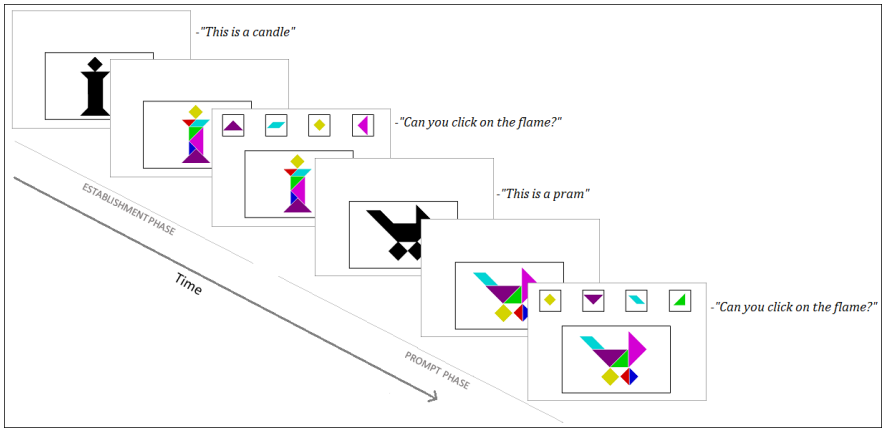


Figure 5.1: An example of one Different Context, Maintain Label trial. Pairs of tangrams were arranged according to a common target shape of the same colour and orientation, in this case the yellow diamond which is referred to by "the flame" during the establishment phase, and by "the wheel" in the test phase.

5.3.5 Audio:

The audio stimuli was recorded from one volunteer voice actor. The audio recordings consisted of a statement that established the configuration of each tangram (e.g. "This is a candle") and a prompt to select one of the shapes using a corresponding label (e.g. "Can you pick out the flame?"). A total of 144 sentences were recorded; an establishing statement and an instruction prompt for each of the seventy-two tangrams. All audio files were edited using the Audacity®software package.

Each item, consisting of audio prompts paired with visual arrays, was rotated through all four experimental conditions in the manner of a latin square, resulting in four lists of twenty-four experimental tangram pairs. These were interleaved with a filler list of twenty-four tangrams. This structure is also illustrated in Fig. 5.2.

Condition	*Establishment phase*		*Prompt phase*	
***S**ame Context* ***M**aintain Label*	"This is a **candle**"	"Can you click on the **flame**?"	"This is a **candle**"	"Can you click on the **flame**?"
***S**ame Context* ***B**reak Label*	"This is a **candle**"	"Can you click on the **flame**?"	"This is a **candle**"	"Can you click on the **wheel**?"
***D**ifferent Context* ***M**aintain Label*	"This is a **candle**"	"Can you click on the **flame**?"	"This is a **pram**"	"Can you click on the **flame**?"
***D**ifferent Context* ***B**reak Label*	"This is a **candle**"	"Can you click on the **flame**?"	"This is a **pram**"	"Can you click on the **wheel**?"

Figure 5.2: Audio phrases and prompts that accompany the presentation of tangram pairs, according to experimental conditions.

5.3.6 Procedure

All experimental stimuli were administered using E-Prime software (Psychology Software Tools Inc., 2014).

Instructions and practice:

Participants were tested individually, in acoustically deadened booths. They were shown how to use the Serial Response Box by the experimenter before beginning the practice phase. The experiment instructions were delivered on-screen, so that participants could read through at their own pace. The instructions informed participants about what tangrams are, and that this experiment used multiple sets of tangrams of different colours; this was to make it clear to participants that each coloured shape constituted an individual referent. They were given examples of what these multicolour tangrams would look like. Participants were told they would see pairs of pictures made of the same shapes, and that pairs would sometimes consist of the same picture twice, sometimes of different pictures, and that they would be asked to select a shape for each picture. Participants were told that if they did not know which shape to choose, they could simply guess. The participants were given practice rounds which consisted of selecting shapes from multicolour tangram pictures. After the practice round, the participant was given an opportunity to ask the experimenter any questions before continuing.

Critical trials:

Each trial consisted of an establishment phase, immediately followed by a test phase. These phases corresponded to the serial presentation of tangram pictures that formed a pair, using the same shapes of the same colours.

Establishment phase: Participants were presented with a silhouette of a tangram picture that appeared entirely black; this was to ensure that the intended arrangement was quickly recognisable to the participants. Accompanying audio established the intended arrangement of the tangram shapes (e.g. "This is a candle"). The picture then remained on screen, but the black silhouette was replaced with the colour version of the tangram. Each shape had a unique colour; this was to ensure that participants could easily and quickly distinguish between each shape. Four of the seven shapes in the picture, one of which was always the target referent, then appeared in a row of boxes above the tangram, with an accompanying prompt that referred to the target (e.g. "Can you pick out the flame?"). The participant was then able to select one of the four shapes using the four buttons on the Serial Response Box.

Test phase: The participant was then presented with either the same silhouette and audio again, or a different silhouette and audio (e.g. "This is a pram") followed by a full-colour view of the shape configuration. Four of the seven shapes, one of which was always the target referent, then appeared in a row of boxes above the tangram, with an accompanying prompt that referred to the target with either the same referring expression, or a different referring expression (e.g. "Can you pick out the wheel?"). The participant was then able to select one of the four shapes using the four buttons on the Serial Response Box. Participants were not given feedback about whether they had selected the correct object; the game simply proceeded. An audio tone and tangram square display appeared after each trial in order to make the 'pair' presentation more salient.

Filler trials: Filler rounds consisted of the same format as critical trials, with tangram pictures from a different list presented in repeated pairs with repeated prompts.

After completing the experiment, participants were invited to fill in an exit questionnaire in order to gauge what they thought the experiment was about, and what their answering strategy was.

5.3.7 Dependent variables

Reaction Time

Participant responses were made via a Serial Response Box, the reaction time (RT) of which was recorded by E-Prime software (Psychology Software Tools Inc., 2014). For each trial, the audio prompt played over a slide that did not allow any response input from the participant. This ensured that no response could be made without hearing the prompt, to prevent the level of data loss incurred in §3.3. After the audio prompt, the visual display remained on screen during a new slide, which allowed response input from the participants via the Serial Response Box. The response slide terminated upon the participant's response, or upon the exhaustion of the imposed time limit. When participants did not respond before the time limit, no data was logged.

Precedent Match

The cell position of the four 'answer' shapes was random for each trial, but was preset in the experiment code. The participant responses corresponded to one of the four 'answer' cells. Both the target position and response values were logged for each trial, and it was therefore possible to calculate whether or not the object selected in the test phase matched the target precedent object.

5.3.8 Results

Reaction Time

I used R (R Core Team, 2013) and LME4 (Bates et al., 2015) for a linear mixed effects analysis of the relationship between reaction time, Context and Label. As fixed effects,

I entered Context and Label (with interaction) into the model, with Maintain label and Same context as the reference level. As random effects, I used a maximal structure (Barr et al., 2013) of random intercepts for subjects and items, as well as by subject and by item random slopes for the interaction of Label and Context. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Visual inspection of residual plots revealed two experimental items that violated assumptions of homoscedasticity. Analyses that included these items resulted in the same pattern of effects reported below, but in some cases models failed to converge. After the omission of the two outlying items, models converged successfully.

Label affected reaction time ($\chi^2(1) = 5.571$, $p < 0.018$), slowing down response time by 1993.9ms $\pm$162.8 (SE) when a label was Broken. Context affected reaction time ($\chi^2(1)$=4.98, p= 0.026), slowing down response time by 1464.5ms, $\pm$194.9(SE) when the Context was Different. An interaction between Label and Context affected reaction time ($\chi^2(1)$76.46, $p < 0.001$). Response time was faster by -3176.4ms $\pm$255.4 (SE) when a Broken Label was delivered in a Different Context. See also Fig. 5.3.

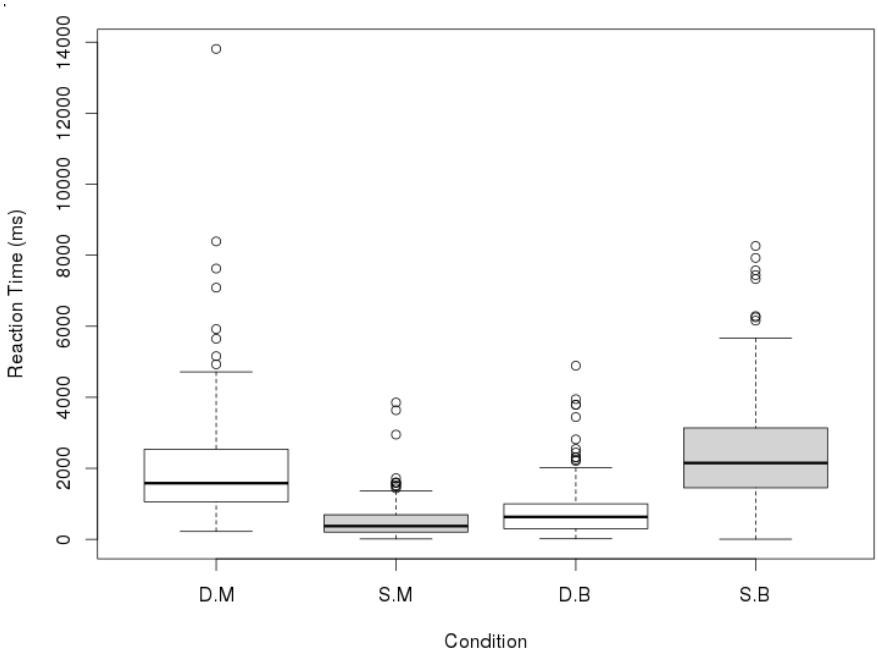


Figure 5.3: Reaction Time (ms) by condition (Context, Label): DM = Different context and Maintained label; SM = Same context and Maintained label; DB = Different context and Broken label; SB = Same context and Broken label.

Object Selection

The cells selected in the critical trial were automatically coded by whether they matched the cell highlighted in the precedent trial (Matches = 0) or not (Nonmatches = 1). All analyses of the target selection data were carried out using logistic mixed-effects regressions with the maximal random effects structure (by-participant and by-item random intercepts and random slopes for fixed effect manipulations), using Matches = 0 as the reference level. In some cases these models failed to converge, in which case the structure was simplified by removing interactions between random slopes. In these cases, the non-converging models indicated the same pattern of effects and significance as the reduced random effects models.

A logistic regression explored the effect of Context (taking Same Context trials as the reference level) and Label (with Maintained Label trials as the reference level). Participants were less likely to match their responses to the precedent shape when a Label was broken (β = -5.02, SE = 0.82, $p < 0.001$), and when the Context was different (β = -2.89, SE = 0.78, $p < 0.001$). However, the interaction of Label and Context produced the biggest change in likelihood (β = 8.54, SE = 1.22, $p < 0.001$); participants were most likely to match their response to the precedent when a different Context was accompanied by a broken Label, or when the same Context was accompanied by a maintained Label. See also Fig. 5.4.

5.3.9 Results Summary

The reaction time data reveals a main effect of Label and Context, and a significant interaction between the two; responses to same context trials were fastest when labels were maintained, and slowest when labels were broken. This pattern was reversed for different context trials, where responses were faster for broken labels and slower for maintained labels. The target selection data exhibited a similar pattern of effects: matches in same context trials were highest in number for maintained label trials, and lowest in number for broken label trials. In different contexts, however, broken la-

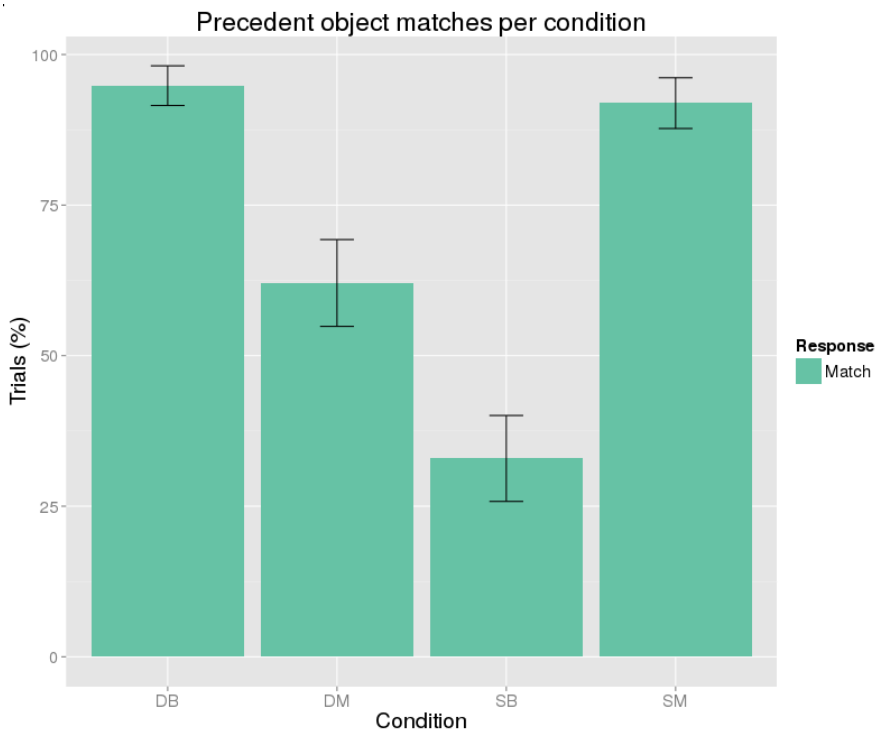


Figure 5.4: Proportion of trials where participants matched their selection to the precedent, by condition (Label, Context). D.M. = Different context and Maintained label; S.M. = Same context and Maintained label; D.B. = Different context and Broken label; S.B. = Same context and Broken label. Error bars show 95% confidence interval.

bels elicited the most matches and maintained labels markedly fewer. This is broadly consistent with the SAH, as it suggests that maintained precedents do not confer a processing benefit unless they are supported by the context, and that a change in context can facilitate a new label for an old referent with no cost to processing.

The key contrastive condition for the LAH and the SAH - maintained labels in different contexts - elicited reaction times that were slower than both [same, maintain] and [different, break] conditions, but faster than when labels were broken in the same context. This may be evidence against the SAH, as the maintained labels were unsupported by the incongruous contexts, yet elicited faster responses than we might expect under the SAH. The object selection measures painted a similar picture: participants gave more matches for maintained labels in different contexts than when labels were broken in the same context. Though in both cases the test phase label was unsupported by the visual context, when this occurred with a maintained label, reaction times decreased and target matches increased, which may indicate that participants were able

to overcome the incongruent context when they could anchor comprehension in a linguistic precedent.

One issue for the interpretation of this data concerns whether participants conceive of the target shape in both establishment and test phases as the *same referent* in each trial. It may be the case that each tangram presentation is approached as a distinct set of referents, and so the behaviour being measured by experiment 3 does not pertain to repeated reference to the same object. In experiment 4, another context accompanies maintained and broken labels in an otherwise identical comprehension task that uses abstract tangram arrangements. If participants are able to identify the target shape with the same label when it is presented as part of an abstract context - i.e. without the interference of an explicitly *incompatible* context - this would provide some evidence that participants do indeed conceive of the target shape in each pair as the same referent. Experiment 4 aims to establish whether the assumption that participants recognise the same target referent across presentations is sound.

5.4 Experiment 4

5.4.1 Participants

24 native English speakers (8 male, median age 22) were recruited from the Edinburgh University campus via undergraduate and postgraduate mailing lists, and word-of-mouth. Participants were paid £2.50 each.

5.4.2 Design

The experiment was a within-subjects and items design, with one factor of two levels: Label [Maintain, Break], whereby the referring expression used in the establishment phase and test phase of each trial was either maintained or broken. The context was abstract in all critical test phases. All participants were therefore exposed to a total of two experimental conditions. The experimental task itself was the same as experiment

3 (§5.3), but with the use of abstract configurations in the test phases of all trials.

5.4.3 Materials

Visual materials:

As in experiment 3 (§5.3), there were 24 experimental items, each comprising a pair of tangrams. Each pair shared one target shape, whose colour and orientation was the same in both tangrams.The target shape had a different colour and shape combination for each pair, which meant that the target for each experimental item was unique (see also Appendix A). However, unlike Experiment 3, one picture in each pair was an abstract reconfiguration of the tangram. These abstract pictures had no discernible arrangement. An example tangram pair for critical trials is shown in Fig. 5.5. The same list of twenty-four tangrams for filler trials as in §5.3, resulting in a total of 72 tangrams.

Each tangram was also rendered in black to form a silhouette image. These were displayed during the first phrase of each tangram presentation, see also §5.4.4. Four shapes from each tangram, one of which was always the target shape, were rendered in a separate image file. Each of these four shapes corresponded to the four buttons of the Serial Response Box participants used to make a selection.

Audio:

The same sound files as recorded for Experiment 3 in §5.3.5 were used for the present experiment. Half of the establishment phrases were replaced with standard audio 'blips' taken from the Audacity®software package (Audacity Team, 2014). These accompanied the introduction of abstract tangram arrangements in critical trials.

5.4.4 Procedure

All experimental stimuli were presented using E-Prime software (Psychology Software Tools Inc., 2014).

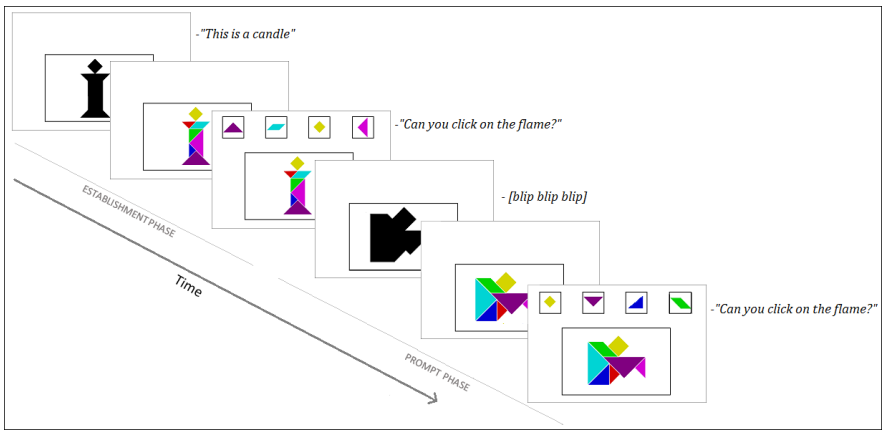


Figure 5.5: An example of one Abstract Context, Maintain Label trial. Pairs of tangrams were arranged according to a common target shape of the same colour and orientation, in this case the yellow square which is referred to by "the flame" during the establishment phase. In abstract presentations, the establishment phrase was replaced by audio 'blips' from the Audacity®package.

Instructions and practice:

Participants were tested individually, in acoustically deadened booths. They were shown how to use the Serial Response Box by the experimenter before beginning the practice phase. The experiment instructions were delivered on-screen, so that participants could read through at their own pace. The instructions informed participants about what tangrams are, and that this experiment used multiple sets of tangrams of different colours; this was to make it clear to participants that each coloured shape constituted an individual referent. They were given examples of what these multicolour tangrams would look like. Participants were told they would see pairs of pictures made of the same shapes, and that sometimes they would be presented with abstract pictures. Participants were told that if they did not know which shape to choose, they could simply guess. The participants were given practice rounds which consisted of selecting shapes from multicolour tangram pictures. After the practice round, the participant was given an opportunity to ask the experimenter any questions before continuing.

Critical trials:

Each trial consisted of an establishment phase, immediately followed by a test phase. These phases corresponded to the serial presentation of tangram pictures that formed a 1pair', using the same shapes in the same colours.

Establishment phase: Participants were presented with a silhouette of a tangram picture that appeared entirely black; this was to ensure that the intended arrangement was quickly recognisable to the participants. Accompanying audio established the configuration of the tangram shapes (e.g. "This is a candle"). The picture then remained on screen, but the black silhouette was replaced with the colour version of the tangram. Each shape had a unique colour; this was to ensure that participants could easily and quickly distinguish between each shape. Four of the seven shapes in the picture, one of which was always the target referent, then appeared in a row of boxes above the tangram, with an accompanying prompt that referred to the target (e.g. "Can you pick out the flame?"). The participant was then able to select one of the four shapes using the four buttons on the Serial Response Box.

Test phase: The participant was then presented with an abstract silhouette, accompanied by audio blips instead of an establishing statement about the picture. This was followed by a full-colour view of the abstract configuration. Four of the seven shapes in each arrangement, one of which was always the target referent, then appeared in a row of boxes above the tangram, with an accompanying prompt that referred to the target with either the same referring expression as the establishment phase in a Maintain condition (e.g. "Can you pick out the flame?"), or a different referring expression in a Break condition (e.g. "Can you pick out the wheel?"). Note that the labels used for the Break condition corresponded to those used in Experiment 3, but for the present experiment the accompanying context was always abstract in test phases. The participant was then able to select one of the four shapes using the four buttons on the Serial Response Box. Participants were not given feedback about whether they had selected the correct object; the game simply proceeded. An audio tone and tangram square dis-

play appeared after each trial in order to make the 'pair' presentation more salient.

Filler trials: Filler rounds consisted of the same format as critical trials, with tangram pictures from a different list presented in repeated pairs with repeated prompts.

After completing the experiment, participants were invited to fill in an exit questionnaire in order to gauge what they thought the experiment was about, and what their answering strategy was.

5.4.5 Dependent variables

Reaction Time

Participants' responses were made via a Serial Response Box, the reaction time (RT) of which was recorded by E-Prime software (Psychology Software Tools Inc., 2014). For each trial, the audio prompt played over a slide that did not allow any response input from the participant. This ensured that no response could be made without hearing the prompt, to prevent the level of data loss incurred in §3.3. After the audio prompt, the visual display remained on screen during a new slide, which allowed response input from the participants via the Serial Response Box. The response slide terminated upon the participant's response, or upon the exhaustion of the imposed time limit. When participants did not respond before the time limit, no data was logged.

Precedent Match

The cell position of the four 'answer' shapes was random for each trial, but was preset in the experiment code. The participant responses corresponded to one of the four 'answer' cells. Both the target position and response values were logged for each trial, and it was therefore possible to calculate whether or not the object selected in the test phase matched the target precedent object.

5.4.6 Results

Reaction Time

I used R (R Core Team, 2013) and LME4 (Bates et al., 2015) for a linear mixed effects analysis of the relationship between reaction time and Label. As fixed effects, I entered Label into the model. As random effects, I had intercepts for subjects and items, as well as by subject and by item random slopes for Label. P-values were obtained by likelihood ratio tests of the full model with Label against the model without Label. Label affected reaction time ($\chi^2(1)$=39.41, $p < 0.001$), slowing down responses by 3421.5ms $\pm$396.6 (SE) when a label was broken; this is illustrated in Fig. 5.6 below.

Object Selection

The cells selected in the critical trial were automatically coded by whether they matched the cell highlighted in the precedent trial (Matches = 0) or not (Nonmatches = 1). All analyses of the target selection data were carried out using logistic mixed-effects regressions with a maximal random effects structure (by-participant and by-item random intercepts and random slopes for fixed effect manipulation), using Matches = 0 as the reference level. A logistic regression explored the effect of Label, with Maintained Label trials as the reference level. Participants were less likely to produce matches when Labels were broken (β = -5.28, SE = 0.97, $p < 0.001$). Matches and nonmatches per condition are shown in Fig. 5.7.

5.4.7 Comparison of experiments 3 and 4

Reaction Time

I used R (R Core Team, 2013) and LME4 (Bates et al., 2015) for a linear mixed effects analysis of the relationship between reaction time, Label and Context across experiments 3 and 4. Experiment 3 trials comprised Maintained and Broken Labels across Different and Same Contexts. Experiment 4 trials comprised Maintained and Broken

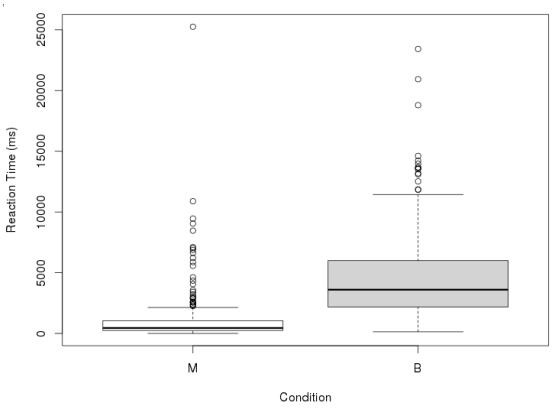


Figure 5.6: Reaction Time (ms) by condition (Label). M = Maintained label; B = Broken label.

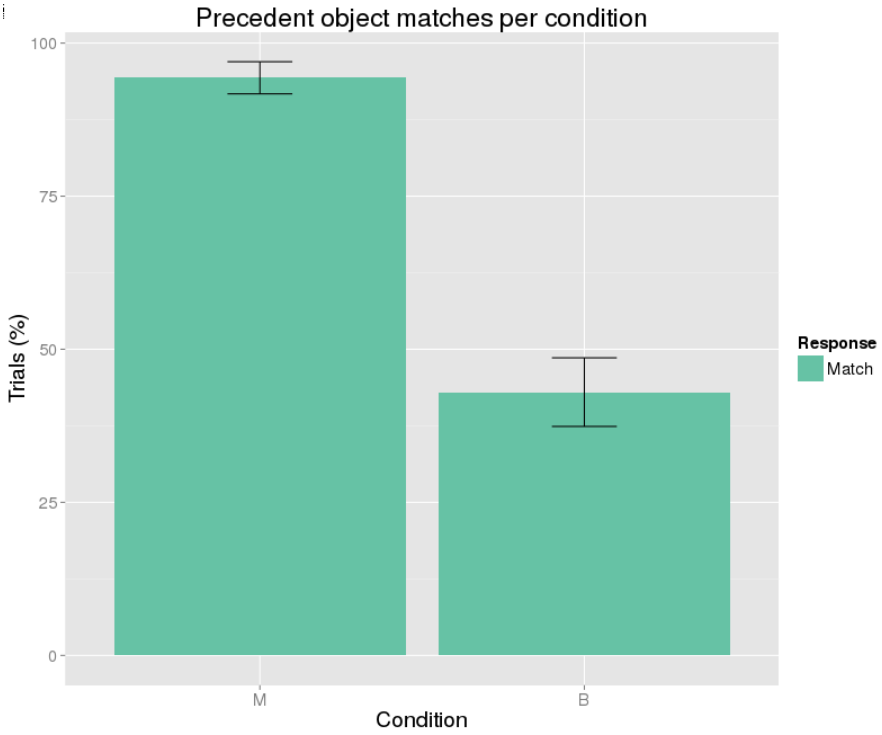


Figure 5.7: Proportion of trials where participants matched their selection to the precedent, by condition (Label). M = Maintained label; B = Broken label. Error bars show 95% confidence interval.

Labels across Abstract contexts. As fixed effects, I entered Label and Context into the model, with Maintained labels in Same contexts as the reference level. As random effects, I had intercepts for subjects and items, as well as by subject random slopes for the effects in question. In cases where models failed to converge, this was remedied by simplifying the random effects structure. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect

in question.

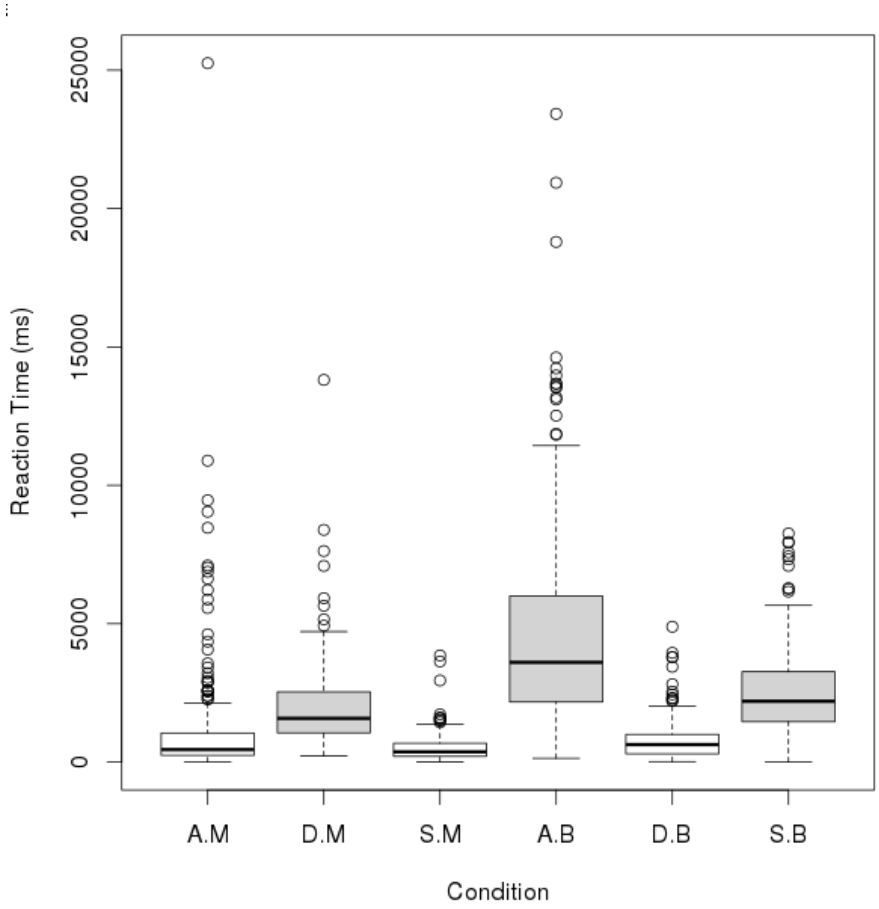


Figure 5.8: Reaction Time (ms) by condition (Context, Label): AM = Abstract Context, Maintained Label; DM = Different Context, Maintained Label; SM = Same Context, Maintained Label; AB = Abstract Context, Broken Label; DB = Different Context, Broken Label; SB = Same Context, Broken Label.

There was no main effect of Context on reaction time ($\chi^2(2) = 0.39$, $p = 0.82$). Label had a significant effect on reaction time ($\chi^2(1) = 18.37$, $p < 0.001$), slowing down participant responses by 1258.8ms, $\pm$274(SE) when a label was broken overall. Label had a significant interaction with Context, ($\chi^2(11) = 230.5$, $p < 0.001$); reaction time was shortened by -3198.6ms, $\pm$332.5(SE) when a Label was Broken in a Different context. When a Label was Broken in an Abstract context, reaction time was 1416.2ms $\pm$340.6 longer. This pattern is illustrated in Fig. 5.9 (p.132).

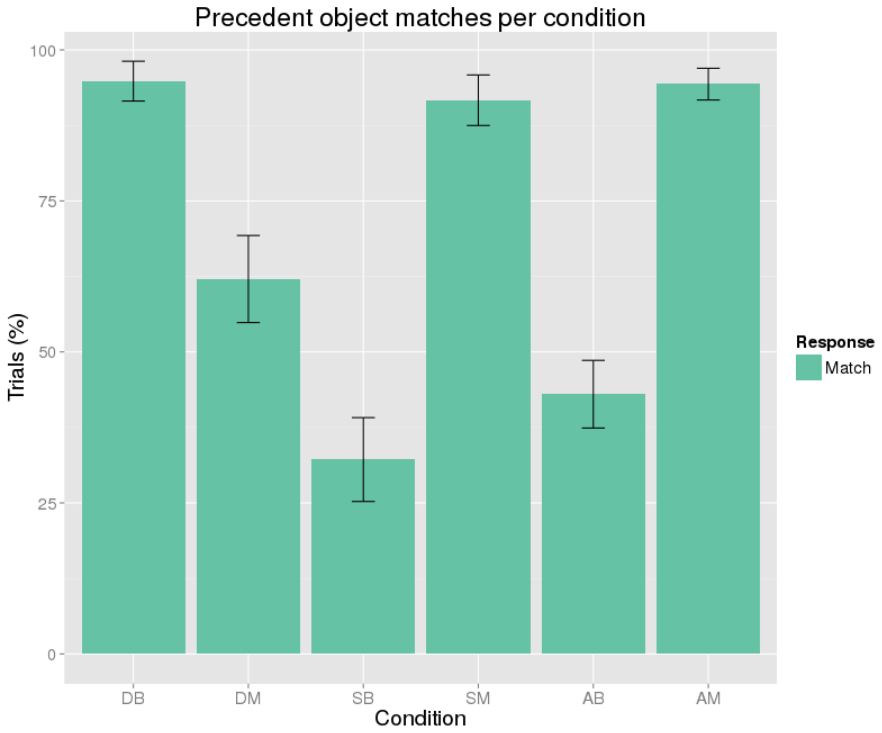


Figure 5.9: Proportion of matches by condition (Context, Label), error bars show 95% confidence interval. DB = Different Context, Broken Label; DM = Different Context, Maintained Label; SB = Same Context, Broken Label; SM = Same Context, Maintained Label; AB = Abstract Context, Broken Label; AM = Abstract Context, Maintained Label.

Object Selection

The cells selected in the critical trial were automatically coded by whether they matched the cell highlighted in the precedent trial (Matches = 0) or not (Nonmatches = 1). All analyses of the target selection data were carried out using logistic mixed-effects regressions with a maximal random effects structure (by-participant and by-item random intercepts, and random slopes for fixed effect manipulation by subject), using Matches = 0 as the reference level. A logistic regression explored the effect of Label with Maintained Label trials as the reference level, and the effect of Context with Same trials as the reference level. Participants were less likely to select the target shape in critical trials when the Label was broken (β = -4.14, SE = 0.57, p $<$ 0.001). Participants were also less likely to select the target when the context was Different (β = -2.29, SE = 0.54, p $<$ 0.001). There was an interaction with Label and Context, such that participants were significantly more likely to select the target object when labels were

Broken in a Different context (β = 7.39, SE = 0.80, p $<$ 0.001) - this effect was larger than trials where labels were Maintained in the Same context (intercept: β = 3.01, SE = 0.54). While Same Context, Break Label (SB) trials elicited shorter reaction times than Abstract Context, Break Label trials (AB), SB trials elicited the most nonmatches. A histogram of matches by condition illustrates this in Fig. 5.9 (p.133).

5.4.8 Discussion

The LAH predicted that maintained precedents should result in faster reaction times across all contexts, leading to a main effect of Label on reaction time. This was broadly consistent with the data; a main effect of Label reflected shorter reaction times when labels were maintained overall. A signficant interaction with Context was driven by different context trials, where broken labels elicited markedly faster reaction times than maintained labels. The LAH also predicted a main effect of Label on target selection, whereby maintained labels elicit more target matches across contexts. The target selection data exhibited a similar pattern to the reaction times; the most matches were elicited by maintained labels in either repeated or abstract contexts, and broken labels in different contexts. Maintained labels in different contexts resulted in fewer matches than broken labels in different contexts, but significantly more than in repeated and abstract contexts.

The SAH predicted that hearers will process broken labels in different contexts that support them as quickly as maintained labels in maintained contexts, and with similar success in target selection. The data is consistent with this, though there was no significant main effect of context. Labels broken in different contexts elicited comparable reaction times and target selection rates to the reference condition of maintained labels in repeated contexts.

The lowest target selection rates and slowest response times were elicited by broken precedents in abstract contexts. In SAH terms, new labels were unsuccessful in contexts that had no features to support them. However, this is also consistent with the

LAH, as new label trials had no precedent in which to anchor comprehension; abstract contexts also do not provide contextual information that can be brought to bear on the task in the absence of linguistic information.

A key contrastive condition for the SAH and LAH was maintained labels in different contexts. Overall, maintained labels in different contexts were slower, and resulted in fewer target matches, than broken labels in different contexts and all other maintained label conditions. While the LAH predicts a different context advantage for broken labels, the cost to processing that different contexts conferred on maintained labels is inconsistent with the LAH in its strongest form, and more consistent with the SAH.

In the case of abstract contexts, the SAH predicts that maintained precedents may determine the target selection, but that reaction time should be slower than in repeated, congruous contexts (and slower than broken labels in different contexts). This was not supported by the data, which shows that reaction times and target selection rates for maintained label, abstract context trials were comparable to maintained labels in repeated contexts.

The overall picture that emerges from this data is somewhat equivocal. On one hand, and in support of the LAH, there was a main effect of Label on both reaction times and target selections. When labels were maintained, participants tended to respond quicker, and match the target shape more often overall. Though incongruent, different contexts led to fewer successful matches and longer response latencies for maintained labels, these trials were still more successful than when labels were broken in other contexts. Maintained precedents also supported comprehension with no contribution from the visual context in abstract configurations. This perhaps lends some tentative support to the core claims of the LAH: that a hearer's representations associated with a linguistic precedent can, in some circumstances, sufficiently support comprehension; and this is because it can provide more reliable means of interpretation than the context, at least some of the time.

One the other hand however, and in support of the SAH, maintained precedents in

different contexts were slower and elicited fewer matches than maintained labels in abstract contexts. In other words, indecipherable contexts that contributed no information to the interpretation of labels did not hinder the use of precedents. But contexts that were *incongruous* with precedent labels exerted a considerable interference in comprehension. A new label for an old referent was processed as quickly as the reference condition of repeated label and contexts, even though all labels were delivered by the same speaker. This lends evidence to a more fundamental role for context in the comprehension of linguistic precedents; a robust finding that recurs in the existing literature on conversational precedents finds that a speaker breaking her own precedent produces the longest response latencies (Kronmüller & Barr, 2015). The present finding suggests that, when accompanied by the correct context, this cost to comprehension need not occur at all.

Limitations

One criticism is that participants do not consider the target shape to be the same referent across the contexts in which it occurs. It could be the case that each tangram arrangement is perceived by participants to be a new set of referents, such that the yellow diamond$_i$ in the establishment phase is considered differently to the yellow diamond$_j$ in the test phase. The use of abstract contexts in experiment 4 offers some evidence that this is not the case; maintained precedent labels led to comparable target matches, with comparably fast reaction times, across abstract contexts. It is reasonable to assume, in these cases, that participants did not hear the label "wheel" and then look for a referent that matched such a label in both pictures. Rather, the repeated label was taken to refer to the same referent in both contexts.

The exit questionnaire invited participants to explain their approach to the task. While the majority of respondents said they simply tried to choose the correct shape through various arrangements, four replies indicated that participants may have still conceived of each presentation as distinct. For example, one participant wrote *"For ones that made no sense sometimes I tried to pick a shape I remembered from a previ-*

ous trial", and another remarked that *"Sometimes the object they wanted me to select was not part of the picture, but had been in the previous one"*. One participant explicitly treated each presentation independently: *"For weird ones, [I went] with [for example] the most "skirt-like" aspect of the gun"*, an approach that would seem to generate a new interpretation for each arrangement. Though responses to each trial took place within a few seconds and participants' insight into their own performance may be variable, responses like these seem to support the criticism that serial tangram presentations may not adequately impress upon participants that recurring shapes constitute the same referent.

Another related limitation of this study concerns the challenge of defining and controlling for other contextual effects. The focus of experiments 3 and 4 was the effect of visual context, and comparisons were made between contexts that matched the linguistic label, conflicted with the linguistic label, or had no correspondence with the linguistic label in the trial or any others. Within the experimental task, the context of tangram configurations clearly exerted effects on participants' comprehension. However, the interpretation of these effects must take into account the wider context that participants found themselves in.

To illustrate, consider the abstract context trials of the task. Participants, after hearing a label that had been established for one of the shapes in front of them, selected the shape again, even though it was no longer accompanied by the establishing context and was instead surrounded by an abstract tangram configuration. This can be interpreted, in LAH terms, as an instance of the hearer's associations with the linguistic stimulus being sufficient to enable an interpretation without any supporting visual context. But there is a wider context beyond the visual display that may perform the same function: The participant is faced with the task of selecting the right shape in an experimental booth, and so will make use of all information offered by the experiment in order to do so. Such demand characteristics can be explained in relevance-theoretic terms that are supportive of the SAH in general: that participants have certain expectations about all of the information presented by the experiment, and so will use this to complete the

task even if it appears nonsensical to do so or does not represent their usual approach to interpretation. The salience of the visual context and the experimental context, respectively, is not easily accounted for, and thus the effect of either can be rather left open to ad hoc interpretation.

Another problem of interpretation concerns how much can be inferred from performances that require interaction with the visual context in general. In different context trials where labels were broken, participants responded quickly and with the most correct matches. In abstract context trials where labels were broken, reaction times were slowest and the fewest target matches were made. One interpretation of this is that the broken labels were not supported by the abstract context of the tangram task, and so comprehension was unsuccessful. This may be considered evidence that comprehension is fundamentally enabled by the context such that an utterance in a completely random context could not be understood by the participant. But there is another conception of what is happening here: that participants have, in fact, understood the instruction, but cannot fulfill its requirements in the situation they are presented with. Long reaction times and unsuccessful matches may be the outcome one would expect from a participant who has understood the label but, upon trying to identify an appropriate referent, cannot find one.

Is this behaviour the result of the visual array affecting comprehension of the utterance, or is it the result of having understood the utterance and then tried to apply the interpretation to the array? This is a problem for the current investigation in particular, because it lies at the very heart of the difference between the LAH and the SAH. Where the LAH holds that linguistic knowledge alone can support comprehension which may then affect interactions with the visual context, the context-driven SAH holds that the linguistic interpretation itself is determined by an appraisal of the context. This problem also applies more generally to the experimental literature about the effects of visual context on processing, including on-line measures such as eye-tracking (as mentioned above in §5.1). Indeed it is hard to imagine a different way of observing participant responses that could overcome this conceptual problem. Observations made at a finer

granularity, such as eye-tracking studies, are faced with the same question; do these fixations represent a 'constraint' on comprehension, or are they the result of comprehending in a particular situation? In the present study, a reference may be unsupported by the context, but this perhaps sheds less light on the success of linguistic comprehension than it would first appear.

5.5 Production of reference across visual contexts

5.5.1 Motivation

Although the focus of this thesis has centred on comprehension, the production of utterances may provide insights relevant to the pragmatic features of linguistic communication. Communication is a fundamentally interactive phenomenon that comprises both a signal and a response (see also §1.1). The interdependent nature of signals and responses is core to the adaptationist definition of communication in evolutionary biology: *"any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved"*(J. M. Smith & Harper, 2003). This adaptationist framing highlights that the capacities of signallers have evolved in order to influence receivers in a particular way, and that the capacities of receivers have likewise evolved to be influenced by signallers. In other words, signals and responses are mutually adaptive and thus depend on one another in a fundamental way (see also Scott-Phillips, 2008).

The seminal characterisation of signal and response coevolution as an 'arms race' between manipulating signallers and mindreading receivers (Krebs & Dawkins, 1984) may give the impression that the evolution of communication involves an inherent conflict of interest. This is perhaps better understood in terms of a game of chess; individual players may try to beat each other, but they must both play roles according to a developed set of rules in order to do so. Just as "Tweedledum and Tweedledee *agreed* to have a battle" (Carroll, 1871), there are mutually binding terms upon which

the ability to pursue any other interest is predicated. In other words, whether a hearer considers a given utterance to be favourable or unpleasant, the capacity to understand the utterance in the first place requires a coevolved, interdependent relationship with the capacities of the speaker.

Given this, we may expect the production of signals to be well-suited to the processing capacities of receivers. One example is the vocalisations of the territorial American bullfrog (*Rana catesbeiana*), which are only evoked by the calls of other American bullfrogs, and are adjusted according to how far away those bullfrogs are (Capranica, 1965; Boatright-Horowitz, Horowitz, & Simmons, 2000). The auditory sensitivity of American bullfrogs and other anurans is partially related to the spectral properties of their species-specific calls (Megela-Simmons, Moss, & Daniel, 1985). The example of territorial urination signals from §1.1 follows similar lines. In the case of canines, the adaptive benefit for both the signaller (of deterring intruders), and the receiver (of avoiding territory disputes), has led to the evolution of signals and responses that depend on each other for communication. Signallers develop autonomous control over fear-response urination, and behaviours such as scratching and leg-raising to increase the scent signal's salience for receivers (Bekoff, 1979); receivers develop highly specialised capacities for scent-detection. In this way, the characteristics of signalling may be informative about the characteristics of receiving, and vice versa.

We can apply similar principles to examine the cognitive features of linguistic production. A range of psycholinguistic literature indicates that speakers are generally sensitive to the context in designing their utterances (see also §2.2). For example, child-directed speech differs from how adults usually speak to each other (Snow, 1972; Lieven, 1994) in a way that young children prefer to listen to (Fernald, 1985) and that may help them with acquisition (Jusczyk & Aslin, 1995).

There is evidence that speakers construct referential descriptions differently when they are intended for others. Fussell and Krauss (1989b) had participants provide referring descriptions for a set of 'nonsense figures' that were either intended to help another person identify the figures, or only for themselves. Participants then had to

match the figures to a set of descriptions; a third of those descriptions were written by the participant herself, a third were written by another participant and intended for others, and the last third were written by another participant but intended only for himself. Participants performed best at matching the figures to descriptions they had written themselves (89%). Of the other descriptions, participants had more success with those written with the intention of being used by somebody else (60%) than with the intention of being only for the person writing the message (49%). This gives some indication that there may be a nontrivial relationship between how utterances are typically constructed, and how hearers are able to process them. However, people in general tend to be suboptimal at assessing the needs of hearers, and have a bias toward their own knowledge which can have variable effects on their message formulation (Fussell & Krauss, 1992). A similar finding by Schober and Clark (1989) indicated that addressees benefit from this audience design more than overhearers who are exposed to all of the exact same words. In a collaborative task where a previously unacquainted 'director' and 'matcher' had to assemble a set of tangram figures, the interactions were recorded. These were then presented to overhearers, who were either privy to all of the trials, or had the first two trials omitted from their playback. The 'early overhearers' assembled the figures quicker, and with more accuracy, than the late overhearers - but still slower and less accurately than the direct addressees of the original interactions.

The importance of interactive collaboration was also emphasised in a key study by Brennan and Clark (1996) (see also §3.1). Recall that participants referred to a number of picture cards presented in different arrays. Partners established disambiguating labels such as "pennyloafer" to identify one of an array of different shoes. Speakers continued to use these labels in subsequent trials with partners they had established the precedent with, even when the label had been rendered overinformative by subsequent arrays (that is to say, arrays for which "shoe" would have sufficed). However, when the same participants spoke to new partners about the same referents, they were more likely to use referring expressions that gave only the amount of information necessitated by the array.

This highlights that the linguistic production of referring expressions appears to be responsive to the needs of hearers in a way far more dynamic and flexible than, say, the calls of bullfrogs or the territory markings of canines. In particular, speakers can use the history of their interaction with a hearer to determine their message formulation, and thus potentially change their productions in a rich variety of ways from interaction to interaction. Beyond the specifics of individual interactions, might there be some overall tendencies that could shed light on which sources of referential information speakers are more likely to leverage?

The LAH and SAH may lead us to expect broad patterns in the production of referring expressions. The SAH holds that hearers interpret utterances on the basis of a rich variety of relevant contextual factors, and that utterances are designed to account for this. The Communicative Principle of relevance proposes that signals *carry a presumption of their own optimal relevance* to the listener. In other words, when speakers produce an utterance, they do so in a way that, given their intended meaning, maximises the relevance of the utterance for the hearer. They accomplish this by taking account of the hearer's knowledge, and the context of the interaction. The listener's expectations of relevance are determined by the context of a given interaction, and inferential comprehension will follow the 'path of least effort' according to these expectations. According to relevance theory, listeners are justified in taking this context-determined path of processing, because they can assume that signals are designed for them. The LAH, by contrast, suggests that speakers may prioritise the consistent use of linguistic precedents over referring expressions that are the best fit to the context. An individual's associative representations may constitute a more stable and easily accessible source of referential information than the multiple factors of a context, and so the comprehension system may be further specialised to derive meaning from linguistic input. If the comprehension system has become specialised to preferentially attend to linguistic input, we may expect that speakers will also preferentially leverage the linguistic precedents available in the discourse - perhaps even doing so regardless of whether they are congruous with the changing context of an interaction.

The following sections present two simple repeated reference tasks in order to investigate how speakers prioritise label consistency and contextual considerations, respectively. In Experiment 5, subjects were asked to produce instructions for future participants in a referential task, who would see the exact same visual displays in the same order. Subjects were exposed to an establishment phase, during which they were presented with a tangram configuration with one shape highlighted. They were then asked to identify the highlighted shape by delivering an instruction to click on it. For example, when presented with a tangram configuration of a candle containing a highlighted yellow diamond, a subject would refer to the diamond using whichever referring expression they thought best in the instruction: "Can you click on the [$\cdots$]?". Subjects would then be exposed to a test phase, where the tangrams would rearrange in an animated display into either the same configuration, or a different arrangement, such as a pram, with the same shape (i.e. the yellow diamond) highlighted. This animation, of the first tangram arrangement into the second tangram arrangement, was intended to impress upon subjects that the shape they were referring to was the same referent in each phase. Subjects would again be asked to issue an instruction for future participants to select the highlighted shape. Subjects' instructions were recorded, and their referring expressions were coded according to whether the labels from the establishment phase of each pair was maintained or broken. Experiment 6 used the same paradigm for a follow up task. The task was identical, except that subjects were presented with either a different tangram arrangement or an abstract tangram arrangement. Abstract tangrams were accompanied by audio blips instead of an establishment phrase (as in §5.4.4).

Subjects are faced with a pragmatic choice in the construction of their utterances: should they prioritise the salient visual context that will be shared with interlocutors, or should they prioritise the consistent use of precedent labels? The SAH predicts that speakers will break their own precedents when the context changes in order to provide hearers with an optimally designed instruction. The LAH predicts that speakers will maintain the same linguistic labels in the face of changing contexts, as hearers are

likely to anchor their comprehension in the referential precedents that have been set.

5.6 Experiment 5

5.6.1 Participants

16 native English speakers (7 male) were recruited from the Edinburgh University campus via mailing lists and word-of-mouth. Participants were paid £5 each.

5.6.2 Design

The experiment was a within-subjects and items design, with one factor of two levels: Context [Same, Different], whereby the tangram configurations used in the establishment phase and test phase of each trial was either the same or different. All participants were therefore exposed to a total of two experimental conditions.

5.6.3 Materials

Visual materials:

As in Experiment 3 (§5.3.3), there were 24 experimental items, each comprising a pair of tangrams. Each pair shared one target shape, whose colour and orientation was the same in both tangrams. The target shape had a different colour and shape combination for each pair, which meant that the target for each experimental item was unique (see also Appendix A).

The pictures for the production experiments were created using a tangram template in Inkscape SVG editing software (Inkscape Team, 2014), and converted to PNGs using ImageMagick software (ImageMagick Studio LLC, 1999-2016). These files were used for animated transitions between pictures during presentation. This allowed participants to directly observe the same referents being used in both the establishment phase and test phase of each trial.

Audio:

The present study used the same establishment phrases for the same visual pairs as in Experiments 3 and 4 (see §5.3.3). This experiment did not, however, use any prompt phrases; the participants' task was to supply these phrases themselves.

5.6.4 Procedure

All stimuli were presented using Python and Javascript.[1]

Instructions:

Participants were given a verbal briefing about the task, where they were told they would be providing stimuli for a simple shape identification task. The verbal briefing included instructions on how the experiment would work that were repeated on-screen for them to read at their own pace. Participants were informed that they were going to see pairs of tangram pictures, that the same shapes were used in each pair, and that they would sometimes see the same picture more than once. Participants were told that for each picture, they would see one shape highlighted. At this point, they were to name it in the following instruction: "Can you click on the $\cdots$?". After delivering an instruction, participants were told click on the shape in order to proceed to the next picture. Participants were also given a shortened version of this format, including the prompt, that they could keep in front of them throughout the experiment.

Critical trials:

Each trial consisted of an establishment phase, immediately followed by a test phase. These phases corresponded to the serial presentation of tangram pictures that formed a 'pair' using the same shape in the same colour and orientation.

Establishment phase: Participants were presented with a black silhouette of the tangram picture, accompanied by an audio establishment phrase (e.g. "This is a lobster").

[1]Credit for the coding of this experiment is due to Justin Quillinan, a colleague at the Centre for Language Evolution.

The black silhouette was then replaced by the multicolour tangram image. All shapes except for the target shape then faded into very faint colour, while the target shape was indicated with a black outline. At this point, participants were to deliver an instruction to select the target shape, for example: "Can you click on the *claw*?". The participant then clicked on this shape themselves, at which point they would see the shapes from the image rearrange themselves into the next picture.

Test phase: Participants were again presented with a black silhouette accompanied by an establishment phrase. The silhouette was either of the same tangram picture, or its corresponding different picture, according to the experimental condition. The black silhouette would again be replaced by the multicolour tangram image, which made the target shape prominent with a black outline and faded background. At this point, participants again delivered an instruction to select the target shape, before clicking on the shape themselves. Trials were separated by a black tangram square and an audio tone, to make the pair format more salient.

5.6.5 Dependent variables

Each participant's experiment was recorded to an mp3 file using a Zoom H1 Digital Field Recorder, and the labels produced for each trial were manually transcribed. Labels produced in the test phase were then coded according to whether they matched the label produced in the establishment phase (1) or not (0). A label qualified as a match if the head noun produced was the same in both referring expressions.

5.6.6 Experiment 5 Results

Produced Labels

Tangram shapes were presented in pairs; the configuration would animate into either a new configuration (Different Context) or the same again (Same Context). Participants' chosen labels on the second instance were recorded and manually coded according to whether they maintained their chosen label (Match = 1), or chose a new label (Match

= 0).

A logistic regression explored the effect of Context on precedent matches, with Same Context trials as the reference level and a maximal random effects structure (by-participant and by-item random intercepts and by-subject random slopes for context). Participants were significantly less likely to maintain their chosen label when the Context around the referent was different (β = -19.53, SE = 3.81, p $<$ 0.001). See also Fig. 5.10.

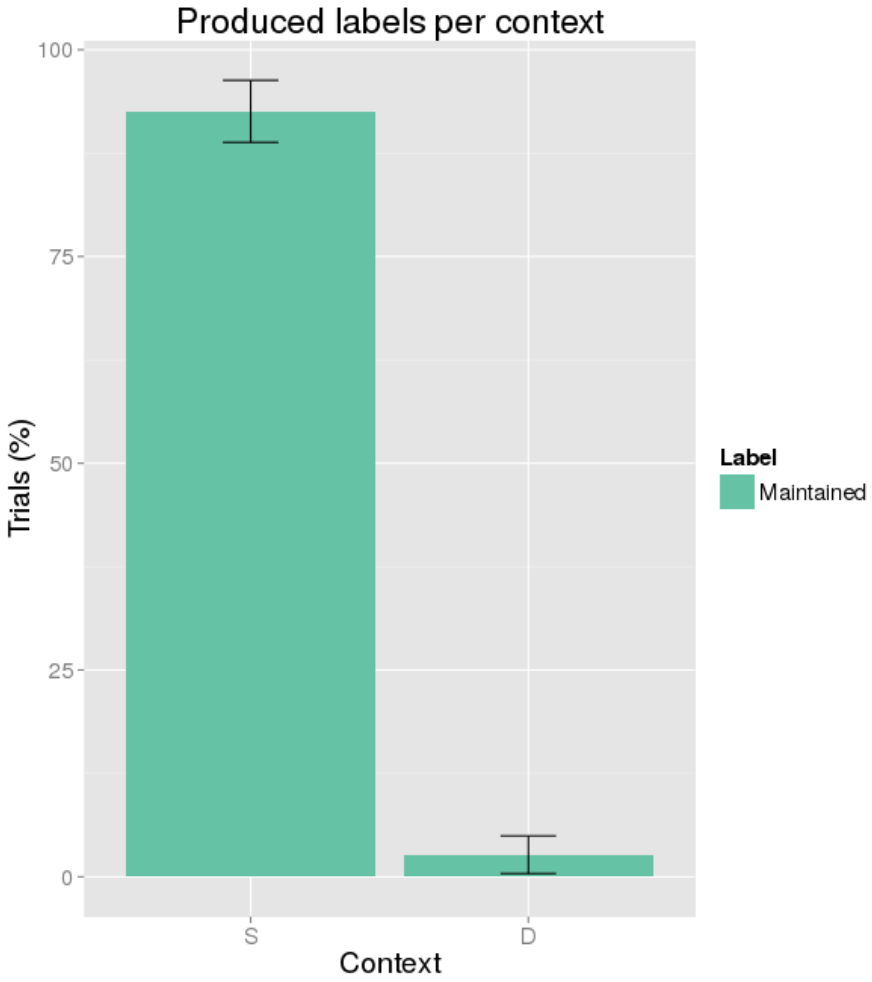


Figure 5.10: Proportion of produced label matches by condition (Context): S = Same Context; D = Different Context. Maintained indicates the production of matched responses in the test phase. Error bars show 95% confidence interval.

5.6.7 Results Summary

The LAH predicts that speakers are likely to maintain their precedents across contexts, as this allows hearers to anchor their comprehension in linguistic precedents. The SAH predicts an effect of context on label production, such that speakers are willing to break their precedents if the context changes. This is in order that hearers can optimally understand their utterances as part of the context of the interaction. The production data showed a main effect of Context: responses to same context trials were matched when labels were maintained, and unmatched when labels were broken. This pattern was reversed for different context trials, where responses were matched for broken labels and unmatched for maintained labels. This is broadly consistent with the SAH, which holds that all utterances are facilitated by the present context; speakers change their referring expressions with the context in order to ensure that they are optimally relevant for the hearer.

The tangrams presented in the establishment phase were animated in the visual display to impress upon subjects that test phases were composed of the *same referents*, however it remains an issue for the interpretation of this data whether participants really did conceive of the target shape in both establishment and test phases as being the same. It may be the case that each tangram presentation was approached as a distinct set of referents, and so the behaviour being measured by experiment 5 does not indicate how speakers make repeated reference to the same object in the task. In experiment 6, speakers are invited to provide referring expressions for referents in both different, and abstract, tangram arrangements. If participants are able to identify the target shape with a common label, this would provide some evidence that participants do indeed conceive of the target shape in each pair as the same referent. Experiment 6 aims to establish whether the assumption that participants recognise the same target referent across presentations is sound.

5.7 Experiment 6

5.7.1 Participants

16 native English speakers (7 male) were recruited from the Edinburgh University campus via mailing lists and word-of-mouth. Participants were paid £5 each.

5.7.2 Design

The experiment was a within-subjects and items design, with one factor of two levels: Context [Abstract, Different], whereby the configuration of the tangram used in the test phase of each trial was either an abstract arrangement with no recognisable picture, or simply a different picture to the establishment phase. All participants were therefore exposed to a total of two experimental conditions.

5.7.3 Materials

Visual materials:

Forty-eight tangram pictures were arranged into twenty-four pairs according to a common shape of the same colour and orientiation, as in Experiment 5 §5.6. An abstract arrangement of the shapes for each pair was also created, resulting in twenty-four additional abstract images; the pairs were counterbalanced into two different lists, wherein 12 pairs contained an abstract picture and 12 contained pairs of conventional arrangements. The target shape had a different colour and shape combination for each pair, which meant that the target for each experimental item was unique (see also Appendix A).

The pictures for the production experiments were created using a tangram template in Inkscape SVG editing software (Inkscape Team, 2014), and converted to PNGs using ImageMagick software (ImageMagick Studio LLC, 1999-2016). These files were used for animated transitions between pictures during presentation. This allowed par-

ticipants to directly observe the same referents being used in both the establishment phase and test phase of each trial.

Audio:

The audio stimuli was as in §5.4.3 with the same establishment phrases used for the same visual pairs, and audio 'blips' used for the abstract tangrams. This experiment did not, however, use any prompt phrases; the participants' task was to supply these phrases themselves.

5.7.4 Procedure

All stimuli were presented using Python and Javascript.[2]

Instructions:

Participants were given a verbal briefing about the task, where they were told they would be providing stimuli for a simple shape identification task. The verbal briefing included instructions on how the experiment would work that were repeated on-screen for them to read at their own pace. Participants were informed that they were going to see pairs of tangram pictures, that the same shapes were used in each pair, and that they would sometimes see more abstract tangram pictures. Participants were told that for each picture, they would see one shape highlighted. At this point, they were to name it in the following instruction: "Can you click on the $\cdots$?". After delivering an instruction, participants were told click on the shape in order to proceed to the next picture. Participants were also given a shortened version of this format, including the prompt, that they could keep in front of them throughout the experiment.

[2]Credit for the coding of this experiment is due to Justin Quillinan, a colleague at the Centre for Language Evolution.

Critical trials:

Each trial consisted of an establishment phase, immediately followed by a 'test phase'. These phases corresponded to the serial presentation of tangram pictures that formed a 'pair', using the same shapes in the same colours.

Establishment phase: Participants were presented with a black silhouette of the tangram picture, accompanied by an audio establishment phrase (e.g. "This is a lobster"). The black silhouette was then replaced by the multicolour tangram image. All shapes except for the target shape then faded into very faint colour, while the target shape was indicated with a black outline. At this point, participants were to deliver an instruction to select the target shape, for example: "Can you click on the *claw*?". The participant then clicked on this shape themselves, at which point they would see the shapes from the image rearrange themselves into the next picture.

Test phase: Participants were again presented with a black silhouette, either of a different tangram picture (accompanied by an establishment phrase), or of an abstract configuration (accompanied by audio 'blips'). The black silhouette would again be replaced by the multicolour tangram image, which made the target shape prominent with a black outline and faded background. At this point, participants delivered an instruction to select the target shape, before clicking on the shape themselves. Trials were separated by a black tangram square and an audio tone, to make the pair format more salient.

5.7.5 Dependent variables

Each participant's experiment was recorded to an mp3 file using a Zoom H1 Digital Field Recorder, from which the labels produced for each trial were manually transcribed. Labels produced in the test phase were coded according to whether they matched the label produced in the establishment phase (1) or not (0). A label qualified as a match if the head noun produced was the same in both referring expressions.

5.8 Experiment 6 Results

5.8.1 Produced Labels

Participants' chosen labels on the second instance were recorded and manually coded according to whether they maintained their chosen label (Match = 1), or chose a new label (Non-Match = 0) in both Abstract and Different Contexts.

A logistic regression explored the effect of Context on precedent matches, with Abstract Context as the reference condition. The model used by-participant and by-item random intercepts. Models failed to converge when random slopes for context were included; this is likely to be due to the small sample size. Participants were less likely to match their referring expression to the precedent when the context was Different (β = -8.073, SE = 2.52, p = 0.001). See also §5.11.

5.9 Comparison of experiments 5 and 6

In the production task for Experiment 5, participants were faced with the pragmatic choice between anchoring their reference in the visual context shared with the intended hearer, and producing consistent labels for referents regardless of a change in context. The SAH predicted that reference is supported by the context in all cases, such that speakers ought to break their own precedents if they are no longer congruous with the context. The LAH predicted that association between the referent and the linguistic label is a principal means by which hearers derive interpretation, and so speakers ought to use their label consistently, regardless of a change in context. The results showed that when the visual context was Different, speakers tended to break their own precedents, which is consistent with the SAH.

Different Context trials consisted of tangram arrangements that were directly *incongruous* with the preceding arrangements. In other words, when a 'candle' arrangement is reconfigured into a 'pram' arrangement, the precedent label *"flame"* for the

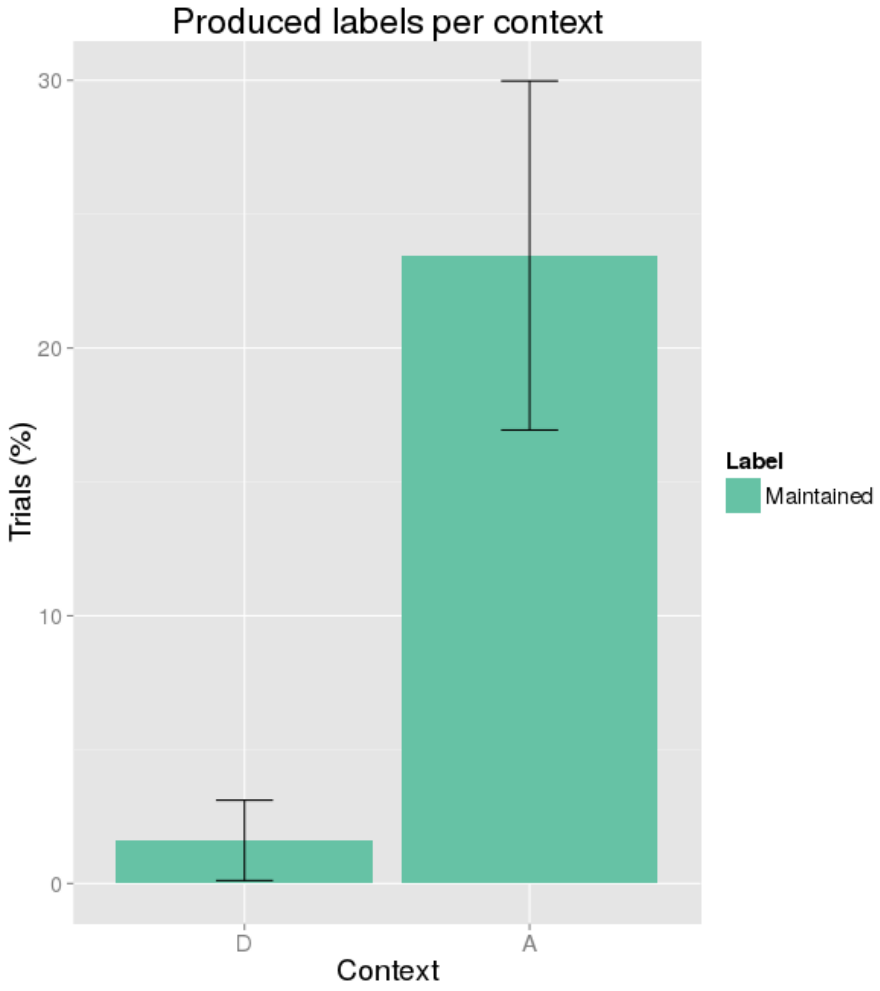


Figure 5.11: Proportion of produced label matches by condition (Context): D = Different Context; A = Abstract Context. Maintained indicates the production of matched responses in the test phase. Error bars show 95% Confidence Interval

common referent is potentially in conflict with a new set of expected labels (such as *"wheel"*). The conventionally recognisable, and explicitly incongruous, arrangements thus elicited new labels from participants. Experiment 6 introduced Abstract context scenarios, whereby precedent labels would be *unsupported* by the visual context without being in direct conflict with them. This tests the extent to which the context-anchored reference strategies in Experiment 5 apply; a reduction in the salience of the changed context may reveal strategies more in line with the LAH. The SAH predicted similar results to Experiment 5, such that speakers will break their own precedents in response to a change in visual context. The LAH predicted that speakers will maintain their reference labels in Abstract contexts. The results show that speakers were less likely to maintain their own labels in Different contexts than in Abstract contexts, which is broadly consistent with the LAH. As illustrated in Fig. 5.11, however, participants still broke their precedents in the majority of Abstract Context trials. A direct statistical comparison of Experiments 5 and 6 is necessary in order to derive effects for the whole dataset, and situate this result alongside Same Context trials.

5.9.1 Results

Responses given in the Different Context condition for experiments 5 and 6 were compared using logistic regression, with Matches = 0 as the reference value, and Different context trials from experiment 5 as the reference level. The model included by-item and by-subject random intercepts. Performance was the same for different context trials across both experiments, with no significant difference in maintained labels (β -1.02, SE = 2.97, p = 0.73).

A logistic regression comparing the Same context trials from experiment 5 and the Abstract context trials from experiment 6 was conducted, with Same context as the reference level. The model included by-item and by-subject random intercepts. Participants were less likely to maintain their label for the target referent in the Abstract condition (β -10.77, SE = 0.002, < 0.001). See also Fig. 5.12.

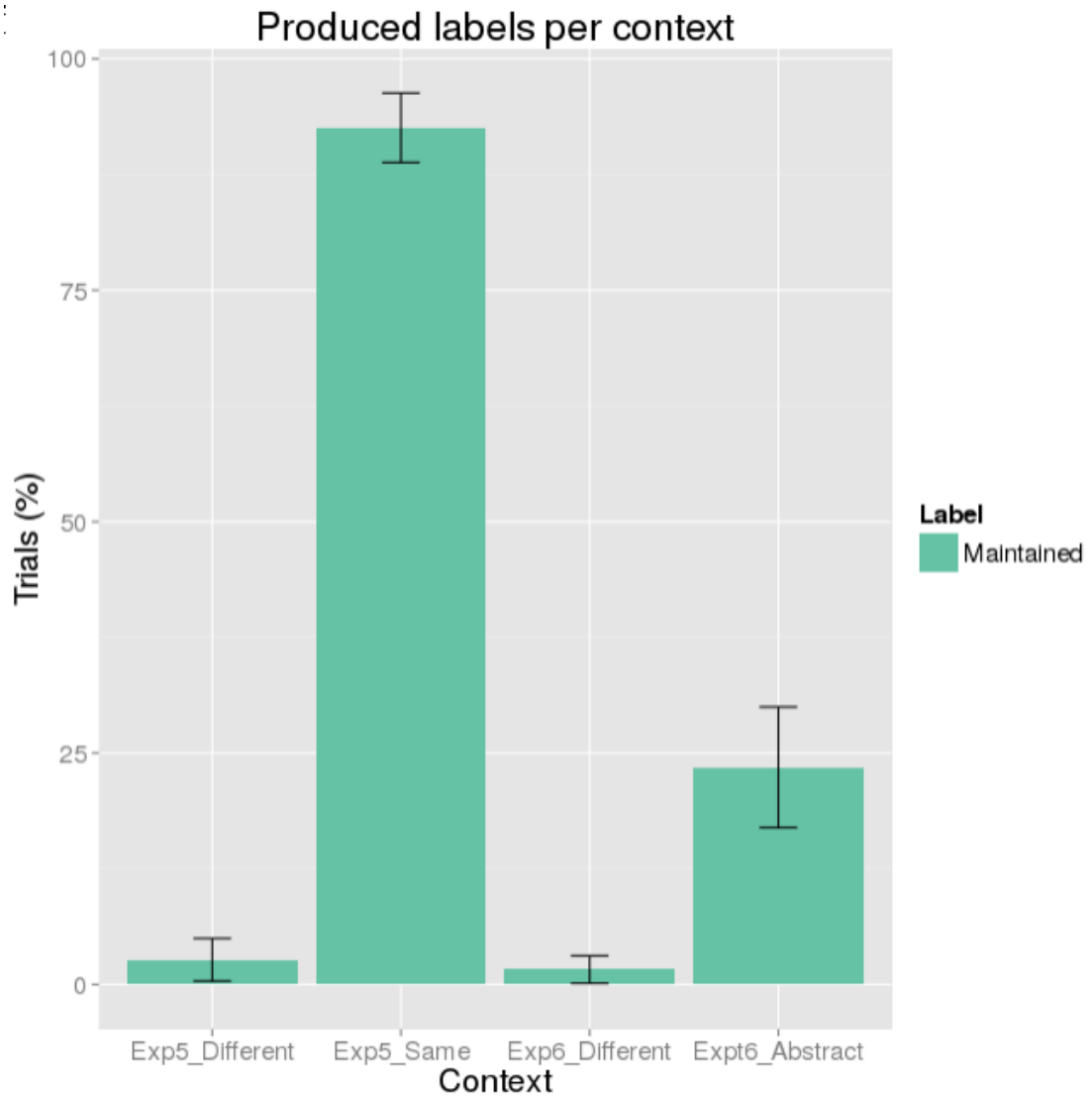


Figure 5.12: Proportion of produced label matches in trial test phases, by Experiment and Context condition. Maintained indicates the production of matched responses in the test phase. Error bars show 95% Confidence Interval.

5.10 Discussion

The LAH predicted that speakers would be more likely to maintain their labels for referents across all contexts. The data is inconsistent with this; there was a significant effect of context in both experiments. In experiment 5, participants broke their precedent labels when the referent was presented to them in a different context 97.3% of the time. In experiment 6, participants broke their precedent labels in a different context 98.4% of the time. Though participants in experiment 6 maintained their precedent labels more often in abstract contexts, they still broke their labels in 75.6% of trials. This was significantly more than same context trials, during which participants maintained their labels 92.6% of the time. The SAH predicted that speakers would maintain or break their labels in accordance with the context. The results are therefore consistent with the SAH; a main effect of context across both experiments reflected that almost all trials with Different contexts elicited different labels, and all trials with a change of context elicited a change of label.

Under the SAH, these results reflect that speakers will necessarily account for the context as far as possible in their production in order to produce well-designed utterances. The broken labels in abstract contexts are an indication that, when forced to prioritise one over the other, speakers weight the present context of their utterance more heavily than the linguistic precedents they have set in the discourse. There is evidence to suggest that participants attempted to give the most context-dependent labels they could in the face of abstract tangram configurations. Examples like *"the shape from the left of the centre towards the bottom"*, *"the shape at the centre of the picture, at the bottom, second from the right"*, and *"the shape directly in the centre of the picture, not shaped like any of the other shapes"* represent a strategy employed by a small number of participants to engage directly with the visual context and produce more elaborate, effortful and abstract descriptions rather than simply repeat labels from the establishment phase which usually consisted of one or two words. This, in SAH terms, represents the tendency of speakers to produce utterances on the basis of the current

context, rather than prioritising linguistic precedents - even if the repetition of a label would ostensibly be an easier strategy.

5.10.1 Limitations and future directions

There are reasons to suspect that the results may reflect the limitations of the experimental paradigm. In particular, the high number of broken labels in abstract context trials warrants caution in the interpretation of these results. The maintained object selection for maintained labels in the abstract contexts of experiment 4 (§5.4.6) provided some evidence that participants conceived of the common shape in both contexts as the same referent, and this accounted for the vast majority of trials at 94.3%. Experiment 6 did not give the same indication, and so it is not clear whether participants conceptualised the referents as being the same object across contexts, or whether each presentation was perceived as a distinct referent.

This large number of broken precedents for abstract contexts was unexpected; participants were informed that the tangrams were presented in pairs that may be rearranged, and the presentation of stimuli for experiments 5 and 6 was animated in order to make it clear that the same shapes appeared in both phases of each trial. Participants saw the first tangram arrangement move on-screen into the second configuration, and these pairs of presentations were separated by an intervening tone and slide. The criteria for a maintained precedent label required only that the referring expressions produced in both the establishment and test phases contained the same head noun. Despite this broad criterion for maintained precedents, labels were maintained in abstract contexts in only 24.4% of trials. It is therefore somewhat equivocal whether the behaviour of participants represents reference to the same referent in different contexts, or simply reference to different referents in all cases.

The broken labels in abstract contexts may be an indication that participants were operating from assumptions about the entire experimental context, rather than the within-task context. In other words, participants may disregard the intended future

audience of the given task, and simply interpret all shape presentations, regardless of context and paired presentation, as being delivered 'by the experiment'. Despite attempts to impress the task upon participants, it may be the case that they operated according to the task of simply labelling shapes presented to them by the experiment. Another possibility is that the geometric shapes themselves may not have been distinct enough *between* pairs for participants to recognise that they were the same *within* pairs.

A minority of responses attempted to interpret the arrangements in a subjective way, much like a Rorschach blot, despite this being a more challenging task in the allotted time than simply repeating their previous label. Examples like *"the upper torso of the rhino in lotus position", "the part where the bottom of the palm would be in a glove",* and *"the eye of the face looking toward the stars"* were all responses to abstract configurations. Although these were tied to the speaker's perception of the context, the results from experiment 4 suggest that these effortful deviations from the established labels may be less likely to elicit a successful response from hearers than simply maintaining the precedent. This strategy is technically consistent with the general SAH account of utterances which are fundamentally motivated by the current context. However, a perhaps more straightforward explanation for these responses is the format of the task; unlike its comprehension counterpart, the production task was not measuring reaction time with a button box, and thus commanded less effortful attention from participants. It may be the case that participants were finding distraction in the stimuli because of a general lack of interest in the task.

It may be the case that, in an interactive version of these tasks, speakers would be more likely to prioritise linguistic precedents in abstract configurations over labels that are closely tied to the visual display. The previous literature that emphasises the role of collaborative grounding of linguistic labels would appear to suggest this (e.g. Brennan & Clark, 1996; Schober & Clark, 1989). A future direction for this line of inquiry may lie in comparisons with precedent maintenance in abstract contexts with a live confederate, or the investigation of the relative weights of linguistic copresence and visual context on the choices speakers make for referring expressions.

5.11 Conclusion

This chapter has presented brief overviews of approaches to visual context and linguistic processing. The LAH predicts that hearers can anchor their comprehension in the linguistic precedents that are available to them in a given discourse. The SAH, by contrast, holds that hearers anchor their comprehension in the attendant context of the discourse, and that this fundamentally underpins linguistic production and comprehension. Though the conversational precedent paradigm is useful for examining the relative effects of linguistic and nonlinguistic contributions to comprehension, visual context has not been studied using this model.

Experiments 3 and 4 used visual arrays of tangrams to study the effect of visual context on the comprehension of linguistic precedents. In experiment 3, participants were asked to select a shape from a tangram configuration, and then to select the shape from a second configuration which was either the same or rearranged into a different picture. The referring expression used in the second presentation was either the same label, or a new label. The shape was the same size, shape, colour and orientiation across both contexts. Under the LAH, we would therefore expect a main effect of label, such that broken labels led to fewer target selections overall, and longer reaction times. Maintained labels in different contexts were a key contrastive condition for both accounts. The LAH predicts that maintained labels in different contexts should elicit comparable target selections and reaction times to maintained precedents in repeated contexts. The SAH predicts that maintained labels in different contexts will elicit slower reaction times and fewer target matches, as the label is unsupported by the context. In experiment 4, participants were given the same task of selecting the same shape across two contexts per trial. However, the second presentation in these trials was an abstract arrangement of shapes with no discernible order or straightforward interpretation. The referring expression used in the second presentation was either the same label, or a new label. In these circumstances, participants should be particularly likely to comprehend maintained labels across abstract contexts by making recourse

to the label established in the previous round; this provides evidence that participants recognise the same referent across presentations. In comparison to repeated contexts, the SAH predicts that maintained labels in abstract contexts will be comprehended slower, and with fewer successful target selections. The LAH predicts that maintained labels in abstract contexts will be comparable to maintained labels in repeated contexts.

The results of experiment 3 were broadly consistent with the SAH; broken precedents that were accompanied by different contexts elicited reaction times as fast as maintained labels in the same contexts, and a comparably high rate of target selections. The contrastive condition of maintained labels in different contexts was not entirely conclusive; reaction times were faster than broken labels in repeated contexts, but slower than maintained labels in repeated contexts. Similarly, maintained labels in different contexts elicited more target matches than broken labels in same and abstract contexts, but fewer matches than broken labels in different contexts. The majority of trials in the maintained label, different context condition did result in matches. This supports an LAH interpretation that linguistic precedents can support interpretation in incongruent contexts at least some of the time. However, this interaction with the context can also be interpreted in terms broadly supportive of the SAH, whereby participants make recourse to linguistic 'copresence' with the experimental rounds in lieu of a supportive context.

In the abstract contexts of experiment 4, participants responded to maintained labels with reaction times and target selections similar to the reference condition of maintained labels in the same contexts. This is evidence that suggests the results of experiments 3 and 4 measure the comprehension behaviour of participants toward the same referent across presentations, rather than conceiving of the shape as a new referent in each configuration.

The comprehension of maintained labels in abstract contexts may be interpreted as some evidence in support of the LAH. The reaction times and target selections for maintained labels in abstract contexts were comparable to those in repeated contexts, which we would expect if comprehension preferentially weighted the linguistic prece-

dent in interpretation. That a linguistic precedent alone can enable comprehension in the absence of any supporting context indicates that a hearer's associations with a label are enough to deliver reaction times and target selection that are as fast and accurate as repeated labels in a repeated context. It should be noted that this experimental condition of an indeterminate context may not reflect the normal processing of hearers out in the world. Though it is a point in favour of the LAH that it is *possible* for a precedent to enable comprehension in this limited and abstract condition, speakers and hearers rarely (if ever) find themselves in this situation. There is also a sense in which the participants themselves were not in that situation, either. An SAH interpretation of comprehending maintained labels in abstract contexts is that participants appealed to the 'common ground' they had with the experiment itself, and made recourse to the previous exposure in order to succeed at the experimental task.

§5.4.8 also raised a conceptual limitation to the study of visual context as a factor in referential comprehension. At first blush this may appear to be the result of using outcome measures like reaction times and selections, which cannot shed light on the details of processing, by definition. But the problem also extends to on-line measures. Eye-tracking studies are faced with the same question concerning fixations on aspects of visual arrays: do these fixations represent the application of an understood proposition, or do they represent the visual context providing input to understanding the proposition in the first place? In other words, on-line measures are faced with a transposition of the original problem at a finer granularity.

Experiments 5 and 6 used visual arrays of tangrams to study the effect of visual context on the production of linguistic precedents. In experiment 5, participants were asked to refer to a shape as part of an instruction intended for players of a simple game. They were first presented with a tangram configuration, and when a shape from the configuration was highlighted, they delivered an instruction. In the second part of each trial, participants were presented with either the same configuration or a different picture, and the same shape was highlighted in order to elicit another instruction. The participants' instructions were recorded, transcribed and coded according to whether

they chose to maintain the same label, or use a different one. The LAH predicts that speakers are likely to maintain their precedent labels across contexts, as this is the information that hearers will preferentially use in comprehension. The SAH predicts that speakers will form their referring expressions by making fundamental recourse to the context of the referent, as the expectations of hearers will be determined by the context in which the reference occurs. Experiment 6 followed the same procedure, except the second phase of each trial used a different configuration, or an abstract configuration. This follow up, as in experiment 4, sought to provide evidence that participants conceived of the target shape in each pair as being the same referent. The LAH predicts that speakers will maintain their precedent labels. The SAH predicts an effect of context on production.

The results of experiments 5 and 6 were consistent with the SAH overall. There was a main effect of context across both experiments, reflecting that participants broke their precedent labels in the majority of trials in which the context changed. In other words, participants maintained their labels when the same context was repeated, but broke their labels across different and abstract contexts. Under the SAH, this effect is underpinned by the context-driven cognition of interlocutors. Speakers produce utterances that are optimally relevant to the hearer. The hearer's expectations of relevance are crucially determined by the context of the utterance itself, and so the speaker's utterances will take this into account. This picture is consistent with the effect of context on the maintenance of referring expressions.

Although participants were more likely to maintain their labels in abstract contexts than in different contexts, they still broke their precedent labels for a large majority of trials. This raises a question concerning the participants' engagement with the task. Specifically, it was not clear that the experiments were eliciting referring expressions for the same referent across contexts. There were additional indications from the qualitative data that participants were conceiving of the target shape in each pair as a distinct referent with each presentation. Though this is explicable in relevance-theoretic terms that are supportive of the SAH, this methodology may simply be unsuitable for the

investigation of repeated reference. Methodological issues are discussed further in the next chapter.

Chapter 6

Summary and conclusions

In this thesis I aimed to situate psycholinguistic investigation within the new and unfolding conversation about pragmatics in evolutionary linguistics. In particular, a current debate about the integration of context in linguistic comprehension has implications that are relevant to theories about how the cognitive architecture of linguistic communication evolved. Using an adaptationist rubric, I outlined that context-driven linguistic comprehension specifies the comprehension system in terms of adaptation to the *social* environment. This is in contrast to language-driven comprehension, which specifies the comprehension system in terms of adaptation to the *linguistic* environment. The overarching approach was to investigate the contrastive hypotheses derived from each of these adaptationist accounts with a series of experiments that collectively aimed to address the following question: how sensitive is the comprehension process to linguistic input qua linguistic input, relative to various other grades of relevant contextual information? These investigations were consistent with the hypothesis that individual associations with linguistic input can be sufficient for comprehending referring expressions, at least under certain conditions. This is inconsistent with the Social Adaptation Hypothesis, and provides evidence for an alternative hypothesis that highlights the evolutionary impact of the emergence of language on human cognition. In this final chapter I will appraise the empirical work presented in Chapters 3-5, and consider the Social Adaptation Hypothesis and Linguistic Adaptation Hypothesis in

the light of this evidence.

I will begin with a commentary on the broad contributions of the thesis. I will then give a summary of the theoretical synthesis, and an appraisal of the empirical investigations that followed. The chapter ends by noting some issues raised by this work that remain open, for which I will briefly suggest some avenues of future study.

6.1 An evolutionary psycholinguistic approach to pragmatics

The work in this thesis has taken a new, interdisciplinary approach to the issue of language and context integration in the processing of reference. The relevant fields of language evolution, psycholinguistics and pragmatics were synthesised using an adaptationist rubric. This allowed a new mechanistic account of processing (the LAH model) to feature in the evolutionary investigation of language, and provided a new line of evolutionary support for language-driven models of processing in psycholinguistics. Here I will outline the merits of this approach for evolutionary pragmatics more broadly.

The field of language evolution has broadly represented an extension of traditional linguistics geared toward addressing questions about the origins and development of language(s) over phylogeny. An example of the dominant research questions in language evolution, identified in one introduction to the field, is laid out in Fig. 6.1 below. It is notable in particular that the questions concerning *mechanisms* refer to the structure of the linguistic form itself; questions about the psycholinguistic processing that constitutes an evolved capacity for language use are more peripheral in the field at large, or else are considered in terms of cognitive pre-adaptations in historical accounts of language emergence (see Scott-Phillips, 2010a, for a detailed discussion).

Conversely, psycholinguistic accounts of processing generally have not incorporated evolutionary considerations (Tooby & Cosmides, 1990), despite an explicit focus

Source	Questions
KIRBY 2007	1. *Structure*: Why is language structured the way it is and not some other way? How can an evolutionary approach explain the particular language universals we observe? **Mechanism** 2. *Uniqueness*: Why are we unique in possessing language? What is so special about humans? **Phylogeny** 3. *Function*: How could language evolve? What were the selective pressures involved? **Function** 4. *History*: What is the evolutionary story for language? When did it evolve? Were there intermediate stages? **Phylogeny**

Figure 6.1: Research questions in language evolution identified in Kirby (2007), with each labelled according to its classification in Tinbergen's (1963) paradigm for evolutionary study.

on the performance of cognitive mechanisms in their relevant task domains. All functional outcomes are ultimately attributable to the operation of adaptations (Tooby & Cosmides, 1990), and so mechanistic accounts of function implicitly entail adaptationism. Making these particulars explicit provides another line of evidence that can illuminate such accounts; compatibility with the evolutionary scenario is pertinent information for theories about cognitive design. In this thesis, an evolutionary approach to existing psycholinguistic models of language-driven comprehension derived a new line of supporting evidence. In particular, this thesis has identified one way in which language, conceived as a cultural environment with its own predictable structural properties, may have played a formative role in the adaptation of the comprehension system. This analysis makes a language-driven account of comprehension available to the existing work in evolutionary pragmatics, which had previously discounted such models on evolutionary grounds (Scott-Phillips, 2014).

Pragmatics has, so far, been situated at the periphery of evolutionary linguistics, and linguistics more broadly. One reason for this is that the interface between language and its use out in the world is less amenable to the kind of formalised abstractions that are fruitful for other linguistic subdisciplines (Scott-Phillips, in press). However, it is precisely because of this focus that inquiry about the evolved nature of human communication, and its particular appearance of complex design, stands much to gain from pragmatics. The evolutionary study of "where the linguistic rubber hits the communicative road" (Scott-Phillips, in press, p.1) necessarily requires an account of the

environmental particulars that have shaped, and continue to enable, our cognitive tools for communication.

This thesis has demonstrated that adaptationism is a useful theoretical tool for this purpose. Adaptationism is aimed at identifying the *design features* of evolved systems, and provides a successful rubric for doing so. Adaptations are, by definition, the interaction of a mechanism with particular features of the environment in order to produce a functional outcome. An account of adaptations thus not only requires the identification of a particular mechanism of processing, but an account of the environmental features that could have interacted with this architecture over phylogeny. These components, mechanism and environment (for this thesis, the cultural/linguistic environment), represent the dominant foci of psycholinguistics and language evolution, respectively. The work in this thesis thus demonstrates an adaptationist synthesis of these components for investigating the design features of context integration in language comprehension.

As well as making mechanistic accounts available to evolutionary approaches to language, and conversely providing evolutionary evidence that can bolster (or challenge) psycholinguistic accounts, the application of an adaptationist rubric lends itself especially well to addressing questions in pragmatics. The study of language use in the world, and explicit acknowledgement of the *functions* it performs, is a natural bedfellow of adaptationist functionalism; both are concerned with how complex, adaptive features (in this case, language) function within the task environment.

This thesis has demonstrated an adaptationist approach to context integration, which is a universal feature of language use (Levinson, 2011). As adaptationism is geared toward the identification of design features in complex systems, universals in language use represent reasonable candidates for adaptationist study. Pragmatics has identified a number of universals (Levinson, 2011), including: deixis (the use of language to denote referents in time and space relative to the conversation); presupposition (the relegation of information to the 'background' of a given utterance's foregounded point); conversation structure (avoidance of overlap with conversational partners through turn-taking); and politeness. For a given universal, an adaptationist rubric should identify

1) the reliably developing property in language users, 2) stable features outside of that property (either in the environment or elsewhere in the language user), and 3) the functional pragmatic outcome that results from the coordination of 1 & 2. As an example, politeness could be investigated along the following lines (based on Brown & Levinson (1987)): 1) particular modes of reasoning from ends, to means that will achieve these ends, 2) attributes of others that constitute 'face', and 3) interactions that avoid threatening those attributes (and thus aid social cohesion). Evidence for this coordination over phylogeny can then be appraised, and empirical testing can proceed with specific manipulations of the identified features that predict functional and nonfunctional outcomes.

6.2 Theoretical synthesis

In chapter 2 I reviewed current debates in pragmatic theory and the psycholinguistic empirical literature that focus on the relative roles of linguistic convention and contextual information in comprehension. I drew parallels between these debates along their common theoretical approaches, despite their distinct lines of scholarly inheritance. The broad polarisation between context-driven comprehension and language-driven comprehension may also imply polar adaptationist particulars, and I sought to make the requirements that each account requires of the evolutionary scenario explicit. This derived the Social Adaptation Hypothesis, and the alternative Linguistic Adaptation Hypothesis. The Social Adaptation Hypothesis (SAH) holds that linguistic comprehension is performed by relevance-oriented inferential mechanisms that have been selected for by a social environment (i.e. inference-using conspecifics). In particular, the SAH holds that linguistic conventions are attended to in the same way as other ostensive stimuli and contextual information, and because of their relevance to communicative interactions. The Linguistic Adaptation Hypothesis (LAH) holds that linguistic comprehension is performed by specialised cognition that has been selected for by a linguistic environment (i.e. language-using conspecifics) that was established

subsequent to, and as a consequence of, the emergence of inferential communication. In particular, the LAH holds that linguistic conventions are a privileged domain of input for the comprehension system.

A recurring theoretical contrast across theoretical pragmatics, psycholinguistics, and cognitive science more generally, concerns the characterisation of mechanistic processes in terms of either task-level requirements, or implementation-level resources. As it pertains to the integration of nonlinguistic contextual information in linguistic comprehension, this distinction characterises two broad accounts of interpreting a speaker's meaning. The task-level, or *computational* characterisation of comprehending speaker meaning gives primacy to the relevant parameters of contextual input that can facilitate inferences about what is intended by an utterance. In particular, this approach identifies metapsychological information about the speaker - that is, inferred information that pertains directly to the speaker's mind state - as the key requirement for understanding what a speaker means. This grounding of the comprehension system in a fundamentally inferential process geared toward metapsychological knowledge holds that linguistic knowledge can only constitute an input to the comprehension process. That is to say, the linguistic input contributes to comprehension only in so far as it aids inference about the speaker's intentions, rather than linguistic association sufficing for interpretation by itself. Linguistic knowledge here is defined as the conventional or associative representations a hearer relates to a given linguistic utterance. The implementation-level, or *algorithmic* characterisation of comprehending speaker meaning gives primacy to the most stable and easily accessible resources that can be brought to bear on linguistic communication. Specifically, this account holds that a hearer's own associative linguistic knowledge can constitute the output of comprehension on at least some occasions, without the need for contextual integration. This latter model - which has thus far been absent from the evolutionary literature - conceives of linguistic associations as a reliable, 'good enough' solution to the problem of deriving a speaker's meaning from linguistic utterances. The former model is thus a context-driven account of comprehension, while the latter is language-driven. The

label ‘language-driven’ may give an overstated impression that any and all linguistic utterances will always be processed with a focus on linguistic meaning. This is not quite an accurate model; the language-driven model contrasts with the context-driven account only by its core contention that comprehension has specialised toward linguistic representations to the extent that it is *possible* that they can support the output of comprehension.

A striking feature of this polarisation across the literatures was the recurring disagreement over whether linguistic knowledge is considered a stable enough resource for comprehension. The relevance-theoretic account of context-driven inference, which forms the core of the social adaptation model, rests upon the radical underdeterminacy of all utterances. This underdeterminacy - which can be otherwise characterised as the unreliability of linguistic knowledge alone - warrants the powerful solution of metapsychological inference. In examples from the semantics/pragmatics border wars in the theoretical literature, semantic content is argued to be fundamentally lacking in a way that requires the context to ‘pick up the slack’ - to the extent that semantic content is argued not to exist in its intuitively-understood form at all. Put differently, the associative representations an individual has are insufficient for a computational task that requires nothing less than mindreading. Or are they? This central question pertains to the evolutionary scenario in a nontrivial way. The question, in adaptationist terms, concerns whether the linguistic environment is stable and reliable enough in coordinating the representations of speakers to consistute a functional proxy for the representations of others, and thus make egocentric knowledge a useful candidate for interpreting utterances. The same border war that occupies philosophers of language also concerns the way in which language has occupied our cognitive territory.

In chapters 1 and 2 I made the case that we can move toward answering this type of question by investigating the predictions of the LAH and the SAH with psycholinguistic methods. In particular, the LAH predicts that associative representations related to linguistic input constitute a specialised domain for the comprehension system that can support the output of an interpretation. By contrast, the SAH holds that linguistic

associations can never constitute an output to comprehension, which must necessarily be inferred from the relevant context in all cases. The processing capabilities predicted by these contrastive accounts should be manifest in two behavioural measures: the interpretation that a hearer arrives at, and the ease (and hence, speed) with which she is able to do this in the presence or absence of this special input. In the experiments that followed, I used designs in which the two accounts made contrasting predictions in order to seek empirical evidence that would favour one account over the other. All of the experiments rested on some form of the conversational precedent paradigm, which looks at participants' treatment of referring expressions for objects that have either been established earlier in the interaction (i.e. precedented references), or referring expressions that have been newly introduced (i.e. unprecedented references). In experiment 1, participants selected objects from an array under the instruction of two different speakers. All of the referring expressions took the format of speaker-specific facts about the referents, for example: "This thing reminds me of my sister". During the task, participants would identify the referent after hearing an instruction that used a referring expression with a precedent, or a new referring expression that broke the precedent. These instructions were delivered either by the same speaker who established the precedent, or a different speaker. In experiment 2, participants were given the same task, but with the added manipulation of cultural copresence; a precedent such as "This thing reminds me of France" was established for either a corresponding cultural object, such as a miniature Eiffel Tower, or a neutral everyday object. In experiment 3, participants were tasked with identifying a referent from an instruction, with either a maintained or broken precedent, when it was presented in either the same or an incongruent visual context. This was achieved with the use of tangrams; seven geometric shapes were reconfigured into different arrangements to make an incongruent visual context, and participants identified constituent shapes. In experiment 4, participants identified the shapes from either maintained or broken precedent labels, but across completely abstract visual context arrangements. In experiment 5, participants described the constituent shapes of tangram pictures across either repeated or incon-

gruent visual contexts, in order to measure the effect of visual context on speakers' likelihood of breaking their own linguistic precedents. Participants in experiment 6 described the constituent shapes of tangram pictures across either incongruous or abstract visual context arrangements. Though the results were variable, there was evidence to suggest that linguistic associations can indeed suffice for interpretation, as evidenced by hearers' responses, at least some of the time. These empirical contributions will be reviewed below.

6.3 Empirical evidence

Although there were some methodological issues that make interpretation of the experimental data problematic (as discussed below), I gathered some results that are consistent with the possibility that associative meanings with linguistic labels are sufficient for producing a comprehension response, without the integration of relevant contextual knowledge or metapsychological inference.

Experiment 1 used an established conversational precedent paradigm to look at the relative effects of precedent label maintenance and speaker identity on the processing of speaker-specific facts about referents. An example of a speaker-specific fact as a referring expression is *"The thing that reminds me of my nephew"*. Referring expressions of this type ensure that the knowledge state of the speaker is a highly relevant parameter for successful comprehension. Under the LAH, we would expect that hearers use the linguistic precedent to determine their comprehension, and so would have slower reaction times and fewer precedent object matches when the precedents were broken. Under the SAH, we would expect that hearers' comprehension would integrate the knowledge of the speaker in all cases, rather than forming an interpretation on the basis of their egocentric associations with the label. Thus, in trials where a new speaker maintained the precedent of an old speaker, the SAH predicted that participants would have longer reaction times and fewer object matches. The results indicated a main effect of precedent label, and no effect of speaker; when a new speaker maintained a

label established by someone else, participants interpreted this according to their own knowledge. This experiment also found that the maintenance of precedent labels exerted effects on referent selection and reaction time measures, while the identity of the speaker did not. This was surprising, as the paradigm used labels that depended on the knowledge-state of the speaker in order to yield an interpretation.

Experiment 2 looked at the relative effects of cultural salience, precedent label maintenance and speaker identity on the speed and accuracy of referential comprehension. The SAH predicted that hearers would understand referring expressions in terms of the speaker and the expectations conferred by the culturally copresent items, such that there would be a main effect of both on comprehension. The LAH predicted that linguistic precedents will facilitate the referential comprehension of hearers, and that contextual knowledge would be employed when linguistic precedents are broken. There was no effect of speaker, and a main effect of label which interacted with both speaker and cultural objects; the data were generally consistent with the LAH. During reference condition trials, a speaker referred to a culturally salient object, for example a Christmas tree, with a congruous label, such as "This reminds me of Christmas". In a subsequent trial of the game, the same speaker would deliver the instruction "Click on the thing that reminds me of Christmas". In the key contrastive condition for this experiment, a culturally neutral object such as a glove was assigned a precedent label that was culturally salient, for example "This reminds me of France". When a different speaker asked for "The thing that reminds me of France" in a subsequent round, the participants clicked on the glove in the majority of trials. This was surprising, because the same array contained a miniature Eiffel Tower; in relevance-theoretic terms, this culturally salient object in the visual array ought to have fulfilled the expectations of relevance generated by a new speaker's reference about France. However, not only did participants select the glove despite the presence of the Tower in the majority of trials, but they did so with comparable reaction times to the reference condition. This provides some evidence that a hearer's associations with a linguistic precedent in the discourse can suffice as an output for the comprehension process, such that the indi-

vidual's associated meaning with the linguistic label determines the comprehension response. Additional evidence for hearers' ability to anchor their comprehension in linguistic precedents came from experiment 4, where hearers were able to identify referents in indeterminate, abstract visual contexts on the basis of the precedent label that had been previously assigned to it.

Taken together, the evidence above is suggestive of hearers' ability to comprehend referring expressions on the basis of linguistic precedent alone, and is thus consistent with the LAH. The referential meaning associated with a linguistic precedent can, at least some of the time, determine the interpretation that hearers ultimately derive - even if this may yield errors such as reaching for a glove instead of the Eiffel Tower when asked for an object associated with France. This evidence is directly inconsistent with the SAH, which predicts that, if a new speaker uses a culturally salient label that was used for a neutral object *by somebody else*, the hearer ought to instead assume the interpretation that is culturally copresent with the new speaker.

Some results that were more consistent with the SAH came from experiments using tangrams to manipulate the visual context around referents. Experiments 3 and 4 looked at the effect of visual context on the comprehension of referring expressions. In experiment 3, hearers had to identify a referent that either appeared in the same visual context twice, or appeared in a different visual context on its second presentation. During this second presentation, the shape would be referred to by either a maintained or broken precedent. For example: in the reference condition, a yellow square would be presented as part of a visual arrangement that looked like a candle, and accompanied by the instruction *"can you click on the flame?"*. This candle arrangement would also appear on a second presentation, with the same instruction repeated (i.e. the precedent was maintained). By contrast, maintained precedents in a different context showed the yellow square as part of a configuration that looked like a pram on the second presentation, and then asked the participant to *"click on the flame"*. The SAH predicts that the visual context confers expectations that determine hearers' comprehension of referring expressions, such that a label that is unsupported by the context will incur longer reac-

tion times and fewer target matches. However, a broken label that is supported by the accompanying context will be as fast and accurate as the reference condition. In trials where a label is maintained in a different context, the SAH predicts a cost to comprehension. By contrast, the LAH predicts that hearers can use linguistic precedents alone to determine their comprehension of referring expressions, and so predicts a main effect of label. In cases where the context is different but the label is maintained, the LAH predicts that processing should still proceed successfully.

The overall pattern of effects in experiment 3 was consistent with the SAH in some key respects. When participants heard new labels for old referents in an accompanying context that was also new, reaction times and target selection were on par with the reference condition. In the case of maintained precedents in a different context, the results were somewhat more equivocal. Though faster and with more target selections than broken precedents in repeated contexts, the responses were also slower and less accurate than those elicited by broken precedents in different contexts. This is incompatible with the strong version of the LAH, whereby linguistic associations can produce fast and efficient comprehension output in all cases. On the other hand, maintained precedents did contribute to faster processing and more successful referent identification despite the incongruent visual context.

6.3.1 Methodological limitations

There are reasons to believe the results in experiment 1 reflect confounds introduced by the methodology. In particular, participants were unable to correctly identify the target referent in the reference condition where the same speaker asked for the same object twice. This indicates that speaker-specific facts as referring expressions may not succeed in testing precedent comprehension in the kind of set-up used in Chapter 3. At first blush, the non-interactive paradigm may appear to be responsible for preventing participants' engagement with the task; participants are forced to engage with the prerecorded speakers in a more abstract way, which is extra cognitive effort when the task

involves tracking the mindstates of speakers. In addition to this, a lack of feedback may seem to have given participants no extra incentive to really comprehend the instructions beyond guesswork.

However, there are reasons to believe the cause lies elsewhere - specifically, in the lack of referents in the array that were explicitly indicated by the labels. Noninteractive paradigms are successful for testing conversational precedent comprehension in general, generating robust effects across multiple studies (see Kronmüller & Barr, 2015), and speaker-specific facts about referents are an established method of testing referential comprehension even in children (Diesendruck & Markson, 2001). The combination of these methods elicited successful task completion and a reasonable pattern of data when they were employed in Experiment 2, which also contained no feedback. The principal difference was that, in Experiment 2, the prompt label always had a 'culturally salient' referent that could be picked out in response to the instruction. This contrasted with experiment 1, where all the referents were equally compatible with the referring expressions. Interrogating the relative contributions of speaker-specfic knowledge and conversational precedents using speaker-specific facts as referring expressions, then, requires some correspondence between the label in the instruction and the objects in the array.

The key contrastive condition for experiment 3, where precedent labels were maintained across incongruent context arrangements, elicited responses that were halfway between the predictions of both the LAH and SAH. In attempting to interpret these results, we come up against a conceptual problem. The overarching agenda for these experiments was to compare linguistic input with various grades of contextual information, of which the visual context constituted one instance. At first it appears that visual context, like speaker-knowledge and cultural associations, can be measured *against* the linguistic precedent in terms of its influence on comprehension. This approach was used in experiments 1 and 2, where the selection measures indicated whether the hearer *made use of* particular sources of information - such as cultural association, speaker knowledge or linguistic precedent - in their interpretation. However, the question we

are presented with in the case of visual context is more subtle than whether hearers weight comprehension more on visual context *or* labels.

This issue extends to results of experiment 3 that were consistent with the SAH. In trials where the second presentation of the referent was in a different context (e.g. a pram arrangement), and accompanied by a different label (e.g. *"Click on the wheel"*), reaction times and target matches were comparable to the reference condition of repeated contexts and labels. In this case we cannot tell whether the responses result from understanding the linguistic label and applying it to the visual array, or having expectations generated by the visual array which then facilitates easy comprehension. In other words, the visual context is qualitatively different to the distinct sets of knowledge that constitute part of the common ground or mutual cognitive environment, and as a result the precedent paradigm cannot measure the manner in which it is brought to bear on comprehension in the same way. It is not clear whether this problem is surmountable, as it conceptual as well as methodological; on-line measures like eye-tracking are faced with the same problem.

Significant methodological limitations became apparent during tasks that looked at label production across visual contexts. Experiment 6 looked at whether speakers maintained their label for a target shape across different contexts, and abstract contexts. Speakers did not consistently identify target referents for the majority of trials that used an abstract context arrangement. This, along with the nature of the labels that were produced, indicates that participants are likely to consider each presentation of shapes as a distinct array of referents, rather than the same set of referents that have simply been reconfigured. This may be particular to tasks where participants have to produce labels for constituent shapes, as experiment 4 indicated that *hearers* have no problem with comprehending the target shape a second time using the same referring expression without any supporting context. On reflection, the performance of participants when it comes to production may be due to the nature of tangrams themselves. The geometric shapes are designed and arranged in order to encourage viewers to conceive of the shapes as *something else*. While tangrams may appear to offer a means of controlling

the visual context in ways that are not possible for naturalistic contexts, the nature of the tangram puzzles themselves invite conceptualisations from the participants which prove more inscrutable. Furthermore, the shape and colour combinations may have been insufficient to make individual shapes memorable across presentations; it may be the case that the presentations were too visually similar across all trials, and so speakers simply treated each trial as separate. The presentation of tangrams in experiments 5 and 6 were animated in order to preclude any tendency to conceive of the target shape in each pair as separate on both presentations, but speakers appeared to do so regardless. Successful use of tangrams for production may be better suited to tasks where the labelling of the constituent shapes themselves is not the object of study (as in Schober & Clark, 1989).

The experiments designed to address the role of visual context in reference processing produced results broadly consistent with the SAH, which holds that the comprehension and production of reference is fundamentally driven by the context. In experiments 3 and 4, the SAH successfully predicted that Broken labels in Different contexts would be processed as quickly as Maintained labels in the Same context, and that Maintained labels in Different contexts would incur a cost to processing that would be reflected in reaction times. The SAH also successfully predicted that Maintained labels in Abstract contexts would elicit slower reaction times than in Same contexts, as any comprehension anchored in the visual context would be compromised when this became unavailable. For experiments 5 and 6, the SAH successfully predicted that speakers would produce novel labels for old referents in all novel contexts.

The methodological limitations of these experiments raise significant issues for the interpretation of their results. The ideal empirical investigation would be able to compare the relative contributions of contextual and linguistic input to the comprehension process. However, there is no straightforward measure with which to compare the contributions of these different domains; in other words, it is not clear how much informational value is carried by a given contextual item as compared to a given linguistic one (see also 2.3.2). In order to measure these effects, a *relative* measure was

used with an operationally-defined baseline condition in which both the context and linguistic label were made available to the participants. This allowed a comparison of comprehension performance when either of these were made unavailable. However, this alone is not sufficient for establishing the contributions made by each factor; the outcome measure must also be bounded in such a way that makes clear which set of information the comprehension process has relied on.

Experiments 1 and 2 demonstrate such an approach, whereby the selection of a particular referent indicates the use *either* of the speaker's knowledge *or* the participant's association with the precedent label. To illustrate, consider the results from the key contrastive condition in Experiment 2. Here, the referent selected by participants indicated which particular source of information they had used in their interpretation. Recall that in trials where the precedent was a Neutral Object (glove), and the label ("reminds me of Paris") was Maintained by a Different speaker, participants selected the glove even though the Eiffel Tower was available in the array. There is no clear reason why inferential processes would identify a glove as the appropriate referent for "reminds me of Paris" when uttered by a new speaker. The selection of the glove must have made use of the participants' individual association with the linguistic label by the previous speaker. In this way, the outcome of comprehension reasonably indicates which source of information has been employed. Motivated interpretation of the more continuous reaction time measure is thus made available with the object selection measure.

In the case of visual context, however, this becomes problematic. Unlike experiments 1 and 2, the referent selection cannot clearly disambiguate which source of information has been drawn upon in comprehension. Visual information is not easily decoupled from linguistic understanding; is a participant's performance the result of visual context *determining* their interpretation in the first instance, or is it the result of trying to *apply* their interpretation to the context? This poses a problem for the SAH, which predicts slower reaction times (compared to the baseline condition) when labels are Maintained in a Different context. The SAH account for this result is that compre-

hension is fundamentally enabled and determined by the context, and so reaction times are slower because participants will find it more difficult to comprehend a label in a context that does not support it. However, if participants *did* comprehend the linguistic label and simply took longer to locate the referent in a new and salient visual configuration, this also would produce slower reaction times. The object selection measure cannot provide disambiguation on this issue; the majority of participants selected the precedent shape in Different context, Maintained label trials.

The clearest line of disambiguating evidence comes from the Abstract trials of Experiment 4. In these trials, the context was unable to support the linguistic label and also lacked the salience of a recognisable arrangement. In other words, the visual context could not enable the understanding of, or conflict with, the linguistic label. These trials elicited more precedent object selections than the baseline condition, and comparable reaction times, when the label was Maintained. This is consistent with an LAH account where the hearer's association with the label is enough to support comprehension.

However, this interpretation also warrants caution. Unlike Experiments 1 and 2, the outcome measure for Experiments 3 and 4 does not adequately indicate which sets of information must have been brought to bear on the task. The history of the discourse constitutes another set of contextual information that can support comprehension, and so the comprehension of Maintained labels in different visual contexts is also consistent with the use of shared knowledge with the speaker in order to comprehend linguistic utterances. By contrast, Experiments 1 and 2 manipulated features of the discourse history itself, such that interpretations could *either* use the knowledge they shared with the speaker *or* the linguistic label that had been established.

This problem also applies to Experiments 5 and 6, which looked at the contribution of visual context to the production of reference. These experiments measured whether participants repeated their reference across Different and Abstract contexts, or produced new labels when the referent was presented in a new context. The results appear to provide clear support for the SAH; speakers broke their precedents in 98.4% of Dif-

ferent context trials, and 75.6% of Abstract context trials. Labels were maintained in 92.6% of Same context trials. The LAH account predicts that speakers will tend to maintain their own precedents, which is not supported by this pattern of results.

However, the unbounded nature of the participants' responses presents a difficulty for the interpretation of results. Put another way, any systematic role of the visual context in the participants' responses was not made clear by the outcome measure of Maintain vs Break, which cannot categorise the response in terms of their relation to particular sets of contextual information. Controlling for the contextual information available to each individual in their production of novel labels is particularly difficult. Many of the new labels were distinctly idiosyncratic - and even creative - interpretations of the tangram arrangements. For example, references made in Abstract contexts such as "the upper torso of the rhino in lotus position" are difficult to reconcile with the SAH account, which expects that speakers will design their utterances according to the context they share with the hearer. Rather, examples like this indicate that speakers were able to make recourse to their own conceptualisations about the tangram arrangements, which are less likely to be shared by a hearer. The task did not constrain participants' answers according to particular contextual parameters, and this makes it difficult to establish exactly which sets of information each speaker employed in the creation of their labels. In other words, while these results are incompatible with the LAH, their support for the SAH is less conclusive.

An additional consideration for the appraisal of experiments 5 and 6 is the use of a non-interactive paradigm to elicit audience-designed labels. Previous lexical choice tasks have found that speakers are likely to maintain their labels when these have been established with a live conversational partner (Brennan & Clark, 1996). The LAH approach is motivated, in part, by the features of the usual discourse environment that enable interlocutors to anchor their language use in linguistic associations common to speakers and hearers. These features include the opportunity for feedback when a label is unsuccessful, and previous experience that their chosen description of the referent was successful. A non-interactive paradigm that does not provide these features in the

task environment may result in strategies of production that do not represent the behaviour of speakers in a normal discourse setting. This may be particularly important for production; the corresponding comprehension experiments provided referring expressions with a bounded set of possible referents. That participants' responses were narrowed to these pre-defined options relieves some of the need for confirmation that their response was appropriate. By contrast, speakers in experiments 5 and 6 were faced with a potentially infinite range of possible responses, and so feedback may have been a crucial omission from the task environment. Taken together with the possibility that participants conceived of each presentation as a novel referent, and the difficulty in establishing exactly which sets of contextual information were employed in the participants' responses, this avenue of support for the SAH is considerably weakened.

A final methodological consideration that may affect the performance of participants across experiments is their awareness that they are taking part in an experiment, rather than participating in normal discourse exchanges. The phenomenon of demand characteristics, whereby taking part in an experiment is characterised as "a special form of social interaction" (Orne, 1962), results in participants performing in a way they believe will aid the experimenter, or give responses that reflect something about wanting to give correct answers. Unlike social psychology studies, more cognitively-oriented studies that measure processing tend to offset a lot of demand characteristics, as the object of study is relatively indifferent to social considerations. However, if participants were of the primary belief that the instructions were delivered by *the experiment*, rather than by *different speakers*, this may lead to an assumption (despite instruction to the contrary) that the speakers share all knowledge within the experiment, and that any differences between speakers that would usually apply are irrelevant to the overall aim of the task. This limitation applies to non-interactive experiments in general, but may also apply to interactive experiments where the participants are aware that it is an experiment; knowledge may be assumed to be shared between all confederates or participants in a study, which would present experimenters with the same problem when interpreting data with fewer speaker effects than anticipated. Note that this issue

applies even to studies that use deception about the nature of the experimental task - the knowledge *that* it is an experimental task poses the problem.

Methodological considerations for evolutionary pragmatics

Several of the methodological limitations reviewed above are pertinent for the psycholinguistic study of evolutionary pragmatics, and pragmatics more broadly. Firstly, the outcome measures of the experimental paradigm ought to specify the contextual factor of interest. This is particularly relevant for the investigation of different grades of contextual information against linguistic information, where the relative weighting and contribution of each factor is difficult to determine. In addition to comparing the relative effects that result from manipulation of these factors (such as comparing reaction times), the outcome measures must also provide reliable evidence about the contribution of one particular set of information. The interpretation of experiments 3 and 4 was difficult because reaction times alone cannot indicate which set of information has been brought to bear on the experimental task. Consequently, the slower selection of a precedented object in a new visual array could not adequately establish support for the SAH. This outcome is consistent with a case where the new visual array has *determined* the interpretation, and also the case where linguistic knowledge has been *applied* to a new visual array. The ideal design for experimental outcome measures, therefore, must provide the means to identify exactly which set of knowledge has been employed for interpretation.

The experiments in this thesis explored hypotheses derived from two rival evolutionary accounts of context integration in linguistic communication. On a related note to the issues identified above, the outcome measures must be designed such that they discriminate between these two accounts. This design should ideally preclude the possibility for results that support one hypothesis without also falsifying the other. In other words, the ideal experimental paradigm for contrasting rival accounts ought to specify two distinct outcomes that each pertain to one of the hypotheses. The task must ensure that only one of these outcomes is possible for the participant.

Outcome measures must not only be designed to specify the contextual factor of interest, but also to control for (or eliminate) effects from contextual factors that do not pertain to the investigation. The outcome measure for Experiments 5 and 6 did not control for the deployment of other kinds of contextual knowledge that were available to the participants. As a result, the task elicited a wide range of responses that were not straightforwardly attributable to the contextual factors under study. The ideal design for outcome measures provides a means of eliminating these other contextual effects from the data.

The success of an experimental paradigm for the investigation of one contextual factor may be insufficient for the investigation of another. Contextual factors that differ in qualitative terms may require a deeper redesign or reinvention of a previously successful experimental paradigm. Experiment 1 used a 2 x 2 conversational precedent paradigm to investigate the roles of shared knowledge with the speaker and the consistency of label use on comprehension. This was applied and expanded in Experiment 2, with more success, for the investigation of shared speaker knowledge, label consistency, and cultural associations with referents. However, the application of the conversational precedent paradigm for the investigation of visual context encountered conceptual problems, specifically related to the difficulty of evaluating the effects of the visual environment. The conversational precedent paradigm translated well across studies of shared knowledge with speakers and cultural associations with referents. These contextual factors are targeted on the domain of hearers' knowledge about the referents and speakers. The paradigm did not successfully translate to the investigation of visual context, which concerns a qualitatively different domain of information. The term 'context' encompasses sets of information that pertain to different domains of processing. As such, paradigms that successfully address one grade of contextual information may not easily translate to another. As well as the contextual factor's role in the pragmatic theory, consideration must also be made to the processing differences it may entail in order to ensure an appropriate experimental design.

The final methodological consideration is of particular importance to the evolu-

tionary approach advocated in this thesis: ideally, *all* features of the environment that pertain to the adaptation of interest must be accounted for in the experimental task environment. Broadly, the LAH account of linguistic communication outlines the coordination of 1) the ability to use associative linguistic meaning, with 2) the features of the discourse, including feedback from the interlocutor, such that 3) cognitively efficient interpretations are produced/comprehended. A lack of feedback in the experimental tasks may bias the results against the LAH account. This feature of the task domain may be particularly important for production experiments 5 and 6; there is a significantly larger range of possible responses in comparison to the comprehension tasks of referent selection. As a result, the role of feedback in the production of consistent linguistic labels may be more important. The absence of this feature in the experimental task domain therefore runs the risk of biasing experimental results against the LAH, or even rejecting it, without testing it under appropriate conditions.

6.3.2 Conclusion

In light of the considerations above, experiment 2 appears to have given the most reliable results. The statistical interactions with Speaker present in the data indicate that participants were in fact sensitive to the difference between speakers despite the non-interactive format of the task. Response rates were considerably better than experiment 1, as participants were better able to understand and complete the task without all objects in the task array being equally compatible with the instruction phrase. The overall pattern of effects was a main effect of Label on both reaction time and object selection, with interactions with the cultural salience of the precedent object and the speaker. The key contrastive condition of experiment 2 in particular showed that hearers can anchor their comprehension in a linguistic precedent, and that this can suffice for a full comprehension response such that the end result of comprehension matches the linguistic convention held by the individual, *despite* this interpretation being a mismatch to both the shared discourse history with the speaker and the culturally copresent alternative

referent.

This evidence is incompatible with the SAH, which holds that egocentric, associative linguistic knowledge is never a candidate output for the comprehension system, which necessarily takes into account the knowledge of the speaker and the mutually available context that accompanies the utterance. The LAH, by contrast, suggests that an the associative linguistic knowledge held by the individual can suffice as an interpretation, without further integration of contextual factors, at least on some occasions. The wider implication of the LAH is that the linguistic environment may have selected for the direct application of linguistic knowledge in the derivation of meaning, in parallel to social inference.

6.4 Future directions

While experimental results consistent with comprehension that is anchored in associative linguistic conventions is some evidence for the LAH, there remain some unexplored questions. Although this thesis has pursued evidence that contrasts linguistic input with types of contextual information, it remains to be seen whether linguistic input is treated differently in comprehension to other ostensive stimuli. This thesis also has not pursued empirical evidence to investigate whether a comprehension strategy that is more focussed on linguistic input than metapsychological inference results from the emergence and evolution of linguistic structure. Here I will make some brief suggestions for future avenues of study.

First, while the LAH has suggested that the cultural evolution of language into fully-fledged, reliable and stable associative systems introduces a pressure to prioritise these kinds of representations in communicative processing, there remains little experimental evidence regarding the role of emerging language structure and a resultant change in the comprehension or production strategies of communicators. A growing body of alien language experiments in the language evolution literature has studied the emergence and cultural evolution of linguistic structure (Kirby et al., 2008;

Kirby, Tamariz, Cornish, & Smith, 2015) including stable semantic categories (Silvey et al., 2015). Experimental semiotics also studies the transmission of symbols, often with communication games in which participants collaborate on referential tasks (see Galantucci & Garrod, 2011, for a review). However, such studies have not explored the effect of emerging linguistic structure on the communication strategies of learners exposed to the alien language. For example, learners of an alien language could play a referential game where one person directs another to, say, arrange the alien fruit on a tree according to an alien gardening manual, *without* any instruction on how they should achieve this. Several measures could indicate the strategies they rely on most, for example i) nonlinguistic ostensive hand gestures, ii) their native language, and iii) the alien language from the learning task. Learners at each generation would learn the language, pass it on to the next generation, and we should expect the linguistic structure to stabilise. We might expect, if exposure to a reliable set of structured associations creates the wherewithal for preferences in communication, that early generation alien fruit gardeners would exhibit more instances of i) and ii) in their completion of the task, whereas later generation gardeners who had been exposed to a structured symbol set for alien fruit may spontaneously employ more of iii) in completing the task. Such a set-up, or some variation like it, may be a first step toward exploring the effect of developing linguistic structure on communication strategies.

Secondly, this thesis has explored the relative contributions of linguistic and non-linguistic information to the comprehension process, but has not pursued empirical evidence about any differences in comprehension elicited from linguistic stimuli compared to *other ostensive stimuli*. The LAH would predict that linguistic stimuli are brought to bear on comprehension in a way that is qualitatively different from other ostensive stimuli such as coughs, grunts, hand gestures, or facial expressions. The SAH, on the other hand, would consider both linguistic and nonlinguistic ostension to be processed in the same way, as input to a central inferential system. One way of interrogating any differences that may exist in the comprehension system's treatment of various types of ostensive stimuli may lie in gesture-speech mismatch paradigms;

several studies by Goldin-Meadow et al. suggest that gestures do in fact contribute to understanding and affect the comprehension of speech (Alibal & Goldin-Meadow, 1993; Goldin-Meadow, 1999; Singer & Goldin-Meadow, 2005). Perhaps on-line measures such as eye-tracking could provide measurements at a finer granularity in order to weight any preferences in comprehension between, for example, linguistic signs and nonlinguistic gestures in the comprehension of sign-language users.

Finally, an interactive replication of experiment 2, during which an old or new live confederate refers to either neutral or cultural objects with maintained or broken precedents, would interrogate the noninteractive results that are in favour of the LAH. Experiment 2 did not result in any effects of speaker, which was perhaps surprising. One possibility is that the different speaker identities were not sufficiently salient to the participants during the task. An interactive version of this same task, with speakers in the same space, may elicit speaker effects that challenge the current results.

6.5 Overall conclusions

In this thesis, I have presented an adaptationist formalisation of two rival mechanistic accounts of contextual integration in linguistic comprehension, with a view to contributing to the development of an adaptationist psycholinguistics. Specifically, I have synthesised context-driven and language-driven accounts of meaning derivation across scholarly domains. From these integrated accounts, I derived two rival adaptationist hypotheses. The Social Adaptation Hypothesis represents an established view in the evolutionary linguistic literature that linguistic comprehension, as one of many applications of social inference, is fundamentally enabled by the relevant context of a discourse, and receives linguistic input alongside other ostensive stimuli and processes them in the same way. An alternative view, the Linguistic Adaptation Hypothesis, suggests that the post-inference emergence and development of language into a stable cultural environment has facilitated the specialisation of the comprehension system toward context-independent linguistic input. Both were found to be plausi-

ble with respect to the evolutionary literature. Experimental evidence in the form of comprehension responses facilitated by individual linguistic associations, without integrated information from the context, lends some support to the Linguistic Adaptation Hypothesis. This new turn in the evolutionary linguistic conversation on pragmatics means that several issues are yet to be investigated. An adaptationist psycholinguistic approach to these questions is still relatively underexplored, but this thesis has shown that psycholinguistics and evolutionary linguistics can indeed make relevant contributions to each other.

References

Ackley, D. H., & Littman, M. L. (1994). Altruism in the evolution of communication. In R. A. Brookes & P. Maes (Eds.), *Artificial life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (pp. 40–48).

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In K. Hirsch-Pasek, L. Bloom, & H. B. Smith (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford University Press.

Alibal, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468–523.

Aoki, K. (1986). A stochastic model of gene-culture coevolution suggested by the 'culture historical hypothesis' for the evolution of adult lactose absorption in humans. *Proceedings of the National Academy of Sciences*, *83*(9), 2929–2933.

Ariel, M. (2002). The demise of a unique concept of literal meaning. *Journal of Pragmatics*, *34*(4), 361–402.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26.

Audacity Team. (2014). Audacity: Free audio editor and recorder, version 2.0.0 http://audacity.sourceforge.net/ [Computer software manual].

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris*, *102*(1), 35–39.

Bach, K. (2001). You don't say? *Synthese*, *128*(1-2), 15–44.

Bach, K. (2010). Impliciture vs explicature: What's the difference? In R. Carston (Ed.), *Explicit Communication* (pp. 126–137). Palgrave.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 131–158). Psychology Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *38*(6), 795-798.

Bar-On, D. (2013). Origins of meaning: Must we go Gricean? *Mind & Language*, *28*(3), 342–375.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*(2), 391–418.

Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In H. L. Colston & A. N. Katz (Eds.), *Figurative Language Comprehension: Social and Cultural Influences.* Routlege.

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed.). Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Barrett, H. C., Frankenhuis, W. E., & Wilke, A. (2008). Adaptation to moving targets: Culture/gene coevolution, not either/or. *Behavioral and Brain Sciences*, *31*(05), 511–512.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*(3), 628.

Batali, J. (1998). Computational simulations of the emergence of grammar. *Approaches to the Evolution of Language: Social and Cognitive Bases*, *405*, 426.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using LME4. *Journal of Statistical Software*, *67*(1), 1–48.

Bekoff, M. (1979). Ground scratching by male domestic dogs: A composite signal. *Journal of Mammalogy*, *60*(4), 847–848.

Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, *34*(4), 433–456.

Bezuidenhout, A., & Morris, R. (2004). Implicature, relevance and default pragmatic inference. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics.* Springer.

Blackburn, P. L. (1999). *The Code Model of Communication: A Powerful Metaphor in Linguistic Metatheory* (Unpublished doctoral dissertation). University of Texas at Arlington.

Bloom, L., & Lahey, M. (1978). *Language Development and Language Disorders.* John Wiley & Sons.

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind & Language*, *17*(1-2), 37–54.

Boatright-Horowitz, S. L., Horowitz, S. S., & Simmons, A. M. (2000). Patterns of vocal interactions in a bullfrog (rana catesbeiana) chorus: Preferential responding to far neighbors. *Ethology*, *106*(8), 701–712.

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational interaction in the scanner: Mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, *25*(9), 3219–3234.

Borg, E. (2009). On three theories of implicature: Default theory, relevance theory and minimalism. *International Review of Pragmatics*, *1*(1), 63–83.

Borg, E. (2013). Minimalism versus contextualism in semantics,. In G. Preyer G & Peter (Ed.), *Context-Sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics* (pp. 339–359.). New York: Oxford University Press.

Botha, R., & Knight, C. (2009). *The Cradle of Language* (Vol. 12). Oxford University Press.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of Animal Communication*. Sunderland MA: Sinauer Associates.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functionalist perspective. *Child Development*, 529–548.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*(3), 177–226.

Brothers, L. (1990). The neural basis of primate social communication. *Motivation and Emotion*, *14*(2), 81–91.

Brothers, L. (2002). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Foundations in Social Neuroscience*, 367–385.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage* (Vol. 4). Cambridge University Press.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue and Discourse*, *2*, 11–33.

Buller, D. J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. MIT Press.

Byrne, R., & Whiten, A. (1989). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press.

Cappelen, H., & Lepore, E. (2008). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. John Wiley & Sons.

Capranica, R. R. (1965). *The Evoked Vocal Response of the Bullfrog; A Study of Communication by Sound.* (Tech. Rep.). Defense Technical Information Centre.

Carroll, L. (1871). *Through the Looking Glass: And What Alice Found There*. Rand, McNally.

Carruthers, P. (2006). Simple heuristics meet massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Culture and Cognition* (Vol. 2, p. 181. Oxford University Press). Oxford University Press, USA.

Carston, R. (1988). Implicature, explicature and truth-theoretic semantics. In R. Kempson (Ed.), *Mental Representations: The Interface Between Language*

and Reality (pp. 155–181). Cambridge University Press.
Carston, R. (1998a). Informativeness, relevance and scalar implicature. *Pragmatics And Beyond New Series*, 179–238.
Carston, R. (1998b). *Pragmatics and the Explicit-Implicit Distinction* (Tech. Rep.). University College London. doi: 10.1.1.400.8649
Carston, R. (2005). Relevance theory, Grice and the neo-Griceans: A response to Laurence Horn's 'Current issues in neo-Gricean pragmatics'. *Intercultural Pragmatics*, *2*(3), 303–319.
Carston, R. (2008). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. John Wiley & Sons.
Carston, R. (2009). The explicit/implicit distinction in pragmatics and the limits of explicit communication. *International Review of Pragmatics*, *1*(1), 35–62.
Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.
Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687.
Chandler, M. J., & Greenspan, S. (1972). Ersatz egocentrism: A reply to H. Borke. *Developmental Psychology*, *7*(2), 104-106.
Chater, N., Clark, A., Goldsmith, J. A., & Perfors, A. (2015). *Empiricism and Language Learnability*. Oxford University Press.
Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and Beyond*, *3*, 39–103.
Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In A. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (pp. 157–168). Cascadilla Press.
Chomsky, N., & McGilvray, J. (2012). *The Science of Language: Interviews with James McGilvray*. Cambridge University Press.
Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509.
Churchland, P. M. (1986). *Scientific Realism and the Plasticity of Mind*. Cambridge University Press.
Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (Vol. 1, p. 33). Lawrence Earlbaum.
Clark, H. (1992). *Arenas of Language Use*. University of Chicago Press.
Clark, H. (1996). Communities, commonalities, and communication. In J. J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity* (Vol. 17, pp. 324–355). Cambridge University Press.
Clark, H., & Carlson, T. B. (1981). Context for comprehension. In J. B. Long & A. D. Baddeley (Eds.), *Attention and Performance IX* (pp. 313–330). Erlbaum Associates.
Clark, H., & Marshall, C. (1978). Reference diaries. In B. Nash-Webber (Ed.), *Proceedings of the 1978 Workshop on Theoretical Issues in Natural Language*

Processing (pp. 57–63). Association for Computational Linguistics.

Clark, 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.* Cambridge University Press.

Clark, H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*, 287–299.

Clark, H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, *26*(2), 209–225.

Clark, H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 245–258.

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

Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, *59*(1), 187–205.

Cosmides, L., & Tooby, J. (1987). From evolution to behaviour: Evolutionary psychology as the missing link. In J. Dupré (Ed.), *The Latest on the Best: Essays on Evolution and Optimality* (pp. 277 – 306). MIT Press.

Culbertson, J., & Kirby, S. (2015). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, *6*(1964).

Danks, J. H. (1970). Encoding of novel figures for communication and memory. *Cognitive Psychology*, *1*(2), 179–191.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381–385.

Dawkins, R. (1983). *Universal Darwinism*. Cambridge University Press.

Dawkins, R. (1986). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. WW Norton & Company.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630.

Dunbar, R. (1998a). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.

Dunbar, R. (1998b). The social brain hypothesis. *Brain*, *9*(10), 178–190.

Dunbar, R. (2003). The social brain: Mind, language and society in evolutionary perspective. *Annual Review of Anthropology*, *32*, 163-181.

Dunbar, R. (2012). Social cognition on the internet: Testing constraints on social network size. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1599), 2192–2201.

Dunbar, R., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*(5843), 1344–1347.

Durham, W. H. (1991). *Coevolution: Genes, Culture, and Human Diversity*. Stanford University Press.

Egan, F., & Matthews, R. J. (2006). Doing cognitive neuroscience: A third way. *Synthese*, *153*(3), 377–391.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*,

32(05), 429–448.

Evans, V. (2014). *The Language Myth: Why Language is not an Instinct*. Cambridge University Press.

Feest, U. (2003). Functional analysis and the autonomy of psychology. *Philosophy of Science*, *70*(5), 937–948.

Feldman, M. W., & Cavalli-Sforza, L. L. (1985). *On the Theory of Evolution under Genetic and Cultural Transmission with Application to the Lactose Absorption Problem*. Stanford Institute for Population and Resource Studies.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195.

Fitch, W. T. (2000). The Evolution of Speech: A Comparative Review. *Trends in Cognitive Sciences*, *4*(7), 258–267.

Flavell, J. H. (1968). The development of role-taking and communication skills in children. *Young Children*, *21*(3), 164–177.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.

Fodor, J. A. (2001). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1-2), 3–71.

Frankenhuis, W. E., & Barrett, H. C. (2013). Design for learning: The case of chasing. In M. Rutherford & V. Kuhlmeier (Eds.), *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention* (pp. 171–195). MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance 12: The Psychology of Reading* (pp. 559–586). Lawrence Erlbaum Associates.

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 165–176.

Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*(3), 203–219.

Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*(6), 509–525.

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*(3), 378.

Futuyma, D. J. (2013). Natural selection and adaptation. In J. Losos (Ed.), *The Princeton Guide to Evolution* (p. 189). Princeton University Press.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: A review. *Frontiers in Human Neuroscience*, *5*, 11.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3-4), 455–479.

Gallistel, C. R. (1990). Representations in animal cognition: An introduction. *Cognition*, *37*(1), 1–22.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, New York.

Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. *Current Biology*, *18*(11), R457–R460.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, *15*(1), 30–33.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, *205*(1161), 581–598.

Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, *41*(01), 34–50.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, 377–388.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3). Academic Press.

Grice, H. P. (1991). *Studies in the Way of Words*. Harvard University Press.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Harnad, S. (1992). Connecting object to symbol in modelling cognition. In A. Clark & R. Lutz (Eds.), *Connectionism in Context* (pp. 75–90). Springer.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88-96.

Hoefler, S. H. (2009). *Modelling the Role of Pragmatic Plasticity in the Evolution of Linguistic Communication* (Unpublished doctoral dissertation). The University of Edinburgh.

Hoefler, S. H., & Smith, A. D. (2009). The pre-linguistic basis of grammaticalisation: A unified approach to metaphor and reanalysis. *Studies in Language*, *33*(4), 886–909.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context:*

Linguistic Applications (pp. 11–42). Georgetown University Press.

Horn, L. (1990). Hamburgers and truth: Why Gricean explanation is Gricean. In K. Hall, J. Koenig, M. Meacham, S. Reinman, & L. A. Sutton (Eds.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society* (p. 454 - 471). Braun-Brumfield.

Horn, L. (2005). Current issues in neo-Gricean pragmatics. *Intercultural Pragmatics*, *2*(2), 191- 204.

Horn, L. (2006a). The border wars: A neo-Gricean perspective. *Where Semantics Meets Pragmatics*, *16*, 21–48.

Horn, L. (2006b). More issues in neo- and post-Gricean pragmatics: A response to Robyn Carston's response. *Intercultural Pragmatics*, *3*(1), 81–93.

Horn, L. (2008). *On F-implicature: Myth-analysis and rehabilitation.* Yale University. (Unpublished manuscript)

Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition*, *40*(1), 113–126.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

ImageMagick Studio LLC, T. (1999-2016). Imagemagick software, apache license: https://imagemagick.org/script/license.php [Computer software manual].

Inkscape Team, T. (2014). Inkscape svg editor [gnu general public license] http://sourceforge.net/projects/inkscape/ [Computer software manual].

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*(1), 55–80.

Keeley, B. L. (2000). Shocking lessons from electric fish: The theory and practice of multiple realization. *Philosophy of Science*, 444–465.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

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*(1), 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), 1–20.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on Theory of Mind use in adults. *Cognition*, *89*(1), 25–41.

Kingdon, J. (1993). *Self-Made Man and His Undoing*. Simon & Schuster London.

Kirby, S. (2001). Spontaneous evolution of linguistic structure- An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681-10686.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language* (pp. 121–147). Springer.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Kotchoubey, B. (2005). Pragmatics, prosody, and evolution: Language is more than a symbolic system. *Behavioral and Brain Sciences*, *28*(02), 136–137.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*(1), 2.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (Vol. 2, pp. 380–402). Blackwell.

Kripke, S. (1977). Speaker's reference and semantic reference. In P. A. French & H. K. Wettstein (Eds.), *Midwest Studies in Philosophy* (Vol. 2, pp. 255–276). Wiley Online Library.

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, *83*, 1–19.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., ... E, H. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161.

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press.

Levinson, S. (2011). Universals in pragmatics. In P. Hogan (Ed.), *The Cambridge Encyclopedia of the Language Sciences* (p. pp.654-657). Cambridge University Press.

Lieven, E. V. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Galloway & B. J. Richards (Eds.), *Input and Interaction in Child Language Acquisition* (Vol. 56, pp. 436–455). Cambridge University Press.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.

Lloyd, E. A. (2015). Adaptationism and the logic of research questions: How to think clearly about evolutionary causes. *Biological Theory*, *10*(4), 343–362.

MacLennan, B. J., & Burghardt, G. M. (1993). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior*, *2*(2), 161–188.

Markman, E. M. (1991). The Whole-Object, Taxonomic, and Mutual Exclusivity assumptions as initial constraints on word meanings. *Perspectives on Language and Thought: Interrelations in Development*, 72–106.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Vol. 2). MIT Press.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial

organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences*, *200*(1140), 269–294.

Mayr, E. (1983). How to carry out the adaptationist program? *American Naturalist*, 324–334.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287.

Megela-Simmons, A., Moss, C. F., & Daniel, K. M. (1985). Behavioral audiograms of the bullfrog (R. Anacatesbeiana) and the green tree frog (H. Ylacinerea). *The Journal of the Acoustical Society of America*, *78*(4), 1236–1244.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The On-line Study of Sentence Comprehension; Eyetracking, ERPs and Beyond* (pp. 15–32). Psychology Press.

Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2-to-6-year-old children. *Developmental Psychology*, *12*(1), 85.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Nettle, D., Cronin, K. A., & Bateson, M. (2013). Responses of chimpanzees to cues of conspecific observation. *Animal Behaviour*, *86*(3), 595–602.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*(2), 135–183.

Noveck, I., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*(11), 425–431.

Noveck, I., & Sperber, D. (2004). *Experimental Pragmatics*. Springer.

Noveck, I., & Spotorno, N. (2013). Narrowing. In L. Golstein (Ed.), *Brevity* (pp. 280–296). Oxford Univrsity Press.

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, *107*(3), 850–903.

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, *36*(4), 313–322.

Oliphant, M. (1996). The dilemma of saussurean communication. *BioSystems*, *37*(1), 31–38.

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

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition* (pp. 140–169). Cambridge University Press.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776.

Panchanathan, K., & Frankenhuis, W. E. (2016). The evolution of sensitive periods in a model of incremental development. *Proceedings of the Royal Society B*, *283*,

20152439.
Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, *78*(4), 1265–1287.
Perreault, C., & Mathew, S. (2012). Dating the origin of language using phonemic diversity. *PloS ONE*, *7*(4), e35289.
Piaget, J. (1959). *The Language and Thought of the Child* (Vol. 5). Psychology Press.
Piccinini, G. (2006). Computational explanation in neuroscience. *Synthese*, *153*(3), 343–353.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190.
Pigliucci, M., & Müller, G. B. (2010). Elements of an extended evolutionary synthesis. *Evolution: The Extended Synthesis*, 3–17.
Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(04), 707–727.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.
Potts, C. (2005). *The Logic of Conventional Implicatures* (No. 7). Oxford University Press.
Psychology Software Tools Inc. (2014). E-Prime 3.0 [Computer software manual]. Retrieved from `www.pstnet.com`
Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.
Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge University Press.
Quine, W. V. (1960). *Word and Object*. MIT Press.
R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `www.R-project.org`
Recanati, F. (2004). *Literal Meaning*. Cambridge University Press.
Recanati, F. (2010). Pragmatic enrichment. In G. Russell & D. Graff-Fara (Eds.), *Routledge Companion to the Philosophy of Language.* Routlege.
Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. *Metaphor and Thought*, 284–324.
Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. McWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (Vol. 87, p. 237). John Wiley & Sons.
Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' Theory of Mind and semantic development. *Child Development*, *72*(4), 1054–1070.
Schiffer, S. R. (1972). *Meaning*. Clarendon Press Oxford.
Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.
Scott-Phillips, T. (2008). Defining biological communication. *Journal of Evolutionary Biology*, *21*(2), 387–395.
Scott-Phillips, T. (2010a). Evolutionary psychology and the origins of language. *Journal of Evolutionary Psychology*, *8*(4), 289–307.

Scott-Phillips, T. (2010b). The evolution of relevance. *Cognitive Science*, *34*(4), 583–601.
Scott-Phillips, T. (2014). *Speaking Our Minds: Why Human Communication is Different, and how Language Evolved to make it Special*. Palgrave MacMillan.
Scott-Phillips, T. (2015a). Meaning in animal and human communication. *Animal Cognition*, *18*(3), 801–805.
Scott-Phillips, T. (2015b). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56–80.
Scott-Phillips, T., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*(9), 411–417.
Scott-Phillips, T., & Kirby, S. (2013). Information, influence and inference in language evolution. In U. Stegmann (Ed.), *Animal Communication Theory: Information and Influence* (p. 421). Cambridge University Press.
Scott-Phillips, T. (in press). Pragmatics and the aims of language evolution. *Psychonomic Bulletin & Review*.
Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*(4), 569–588.
Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science*, *39*(1), 212–226.
Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, *39*(4), 553–569.
Simon, H. A. (1991). *The Architecture of Complexity*. Springer.
Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85–89.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3.
Smith, J. M., & Harper, D. (2003). *Animal Signals*. Oxford University Press.
Smith, K. (2001). The evolution of learning mechanisms supporting symbolic communication. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society.* Lawrence Erlbaum.
Smith, K. (2003). Learning biases for the evolution of linguistic structure: An associative network model. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life: 7th European Conference* (pp. 517–524). Springer.
Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, *228*(1), 127–142.
Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*(4), 371–386.
Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.
Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574–608.
Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 549–565.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. *Metarepresentations: A Multidisciplinary Perspective*, Oxford University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance Theory explains the selection task. *Cognition*, *57*(1), 31–95.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*(1), 40–46.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (Vol. 142). Cambridge MA: Harvard University Press.

Sperber, D., & Wilson, D. (1987). Précis of relevance: Communication and cognition. *Behavioral and Brain Sciences*, *10*(04), 697–710.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1-2), 3–23.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140.

Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, *8*(4), 339–356.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to Theory of Mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656.

Strawson, P. F. (1964). Intention and convention in speech acts. *The Philosophical Review*, 439–460.

Symons, D. (1990). Adaptiveness and adaptation. *Ethology and Sociobiology*, *11*(4-5), 427–444.

Tanenhaus, M. K., Chambers, C., Hanna, J. E., & Hall, M. (2004). Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 279–317). Psychology Press.

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*(5217), 1632.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., ... Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Tooby, J., & Cosmides, L. (1990). Toward an adaptationist psycholinguistics. *Behavioral and Brain Sciences*, *13*(04), 760–762.

Trueswell, J., & Gleitman, L. (2004). Childrens eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. Henderson & F. Ferreira (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 319–346). Psychology Press.

Trueswell, J., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Trueswell, J., & Tanenhaus, M. K. (2005). *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*. MIT Press.

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133.

Vazquez, M. D., Delisle, S. S., & Saylor, M. M. (2013). Four-and six-year-olds use pragmatic competence to guide word learning. *Journal of Child Language*, *40*(02), 291–306.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*(4), 357–380.

Von Frisch, K. (1967). *The Dance Language and Orientation of Bees*. Harvard University Press.

Weaver, W. (1949). The mathematics of communication. *Scientific American*, *181*(1), 11–15.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of Theory-of-Mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wharton, T. (2003). Natural pragmatics and natural codes. *Mind & Language*, *18*(5), 447–477.

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1599), 2119–2129.

Williams, G. C. (2008). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wilson, D. (1994). Relevance and understanding. In G. Brown (Ed.), *Language and Understanding* (chap. 3). Oxford University Press.

Wilson, D., & Carston, R. (2007). *A Unitary Approach to Lexical Pragmatics: Relevance, Inference and Ad Hoc Concepts*. Palgrave.

Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, *90*(1), 1–25.

Wilson, D., & Sperber, D. (2008). Relevance Theory. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics.* John Wiley & Sons.

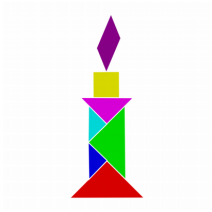
Wilson, D., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

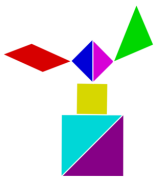
Zaitchik, D. (1991). Is only seeing really believing?: Sources of the true belief in the false belief task. *Cognitive Development*, *6*(1), 91–103.

Appendices

.1 Paired Tangrams (Experiments 3-6)

Arrangement 1	*Arrangement 2*	*Target Shape*
		
This is a candlestick	*This is a house*	*Can you click on the base/roof?*
		
This is a man in the rain	*This is a lobster*	*Can you click on the umbrella/claw?*
		
This is a gun	*This is a lady*	*Can you click on the handle/skirt?*
		
This is a lamp	*This is a puppy*	*Can you click on the lampshade/hindlegs?*
		
This is a bird	*This is santa claus*	*Can you click on the wing/beard?*

.2 Paired Tangrams (Experiments 3-6)

 *This is a chicken*	 *This is a turtle*	 *Can you click on the beak/flipper?*
 *This is a waiter*	 *This is a tie*	 *Can you click on the food/knot?*
 *This is a rabbit*	 *This is cupid*	 *Can you click on the ear/bow?*
 *This is a boat*	 *This is a fan*	 *Can you click on the sail/propellor?*
 *This is a camel*	 *This is an old camera*	 *Can you click on the hump/base?*

.3 Paired Tangrams (Experiments 3-6)

 *This man is eating*	 *This man is dancing*	 *Can you click on the bowl/foot?*
 *This is a shark*	 *This is a duck*	 *Can you click on the fin/bill?*
 *This is a helicopter*	 *This man is sneezing*	 *Can you click on the propellor/hankerchief?*
 *This is a man*	 *This is a moutain*	 *Can you click on the hat/peak?*
 *This is an apple*	 *This is a swan*	 *Can you click on the leaf/head?*

.4 Paired Tangrams (Experiments 3-6)

 *This is a bull*	 *This is a cat*	 *Can you click on the horn/tail?*
 *This man is running*	 *This is a t-shirt*	 *Can you click on the arm/sleeve?*
 *This is a teapot*	 *This is a girl*	 *Can you click on the spout/ponytail?*
 *This is a christmas tree*	 *This is a smiley face*	 *Can you click on the stump/eye?*
 *This is a potted plant*	 *This is an angry face*	 *Can you click on the flower/nose?*

.5 Paired Tangrams (Experiments 3-6)

 *This is a pram*	 *This is a candle*	 *Can you click on the wheel/flame?*
 *This is a footballer*	 *This is a question mark*	 *Can you click on the ball/dot?*
 *This is a cottage*	 *This is a dog*	 *Can you click on the chimney/snout?*
 *This person is reading*	 *This person is playing cricket*	 *Can you click on the book/bat?*