The adaptive significance of human language

Function, form and social evolution

Nathaniel Tillman Oesch

St Cross College

University of Oxford

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy in Experimental Psychology to the

Department of Experimental Psychology

Tinbergen Building

University of Oxford

Trinity Term 2014

Supervised by Professor Robin I. M. Dunbar

The adaptive significance of human language: Function, form and social evolution

Nathaniel Tillman Oesch
St. Cross College

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology, Department of Experimental Psychology, University of Oxford

Trinity Term 2014

Abstract

Language is arguably one of the most salient features that distinguish humans from other animal species. However, despite the existence of a large body of relevant theoretical and empirical research, there is currently no consensus as to why language emerged exclusively in the human species or how it evolved its unique communicative structure.

In this thesis, I therefore take a multi-pronged approach to analysing and testing several different hypotheses for the biological function and evolution of language. In Chapter I, I review the evidence and theoretical arguments for each of these proposals and provide, in place, a synthetic perspective which integrates or eliminates each of these ostensibly competing hypotheses for the biological function of language. In Chapter II, I employ the first experimental test of the interdependence hypothesis: the unique proposal offered to explain the emergence and potential coevolution of language and cooperation in the human species. In pursuit of this experiment, I employed a cooperative social foraging task using small and large groups to determine what factors enable individuals to make sense of information from others and converge upon a group consensus. In Chapter III, I take an experimental approach to determine whether aspects of human language can be characterised in terms of honest signalling theory. In this respect, I test several different proposals predicted by the sexual selection and deception hypotheses for human language function.

In Chapter IV, I divert attention away from biological function to focus more closely on language structure. More specifically, I take an experimental approach to the problem of how and indeed whether recursive syntax evolved to be a consistent feature of human language. In pursuit of this experiment, I utilized the Imposing Memory Task (IMT) and a recursive syntax measure, to determine relative performance on each of these cognitive tasks, thereby testing whether recursive syntax may have evolved in tandem with higher-order intentionality (also known as embedded mindreading).

Finally, in Chapter V, I discuss the results and implications of these experiments, and possible suggestions for future studies.

Declaration

I hereby declare that this thesis is of my own composition and it has not been submitted in any previous application for a degree. The work reported in this thesis was executed by myself, except where due acknowledgement is made in the text.

Nathaniel Tillman Oesch
October 2014

Acknowledgements

Before I started writing up this dissertation, I was under the impression it would be a fairly painless and straightforward process, taking myself no longer than three months to complete. I am happy to report that although this initial estimate was entirely incorrect, roughly one year later, I feel relieved to have allowed myself a healthy margin of error of at least 3 standard deviations from the mean. However, I do feel certain that much of the continued inspiration and industriousness was due in no small part to the continued patience, generosity and assistance of those within my 150 person social network, including family, friends and colleagues.

Firstly, special thanks go to my supervisor, Robin Dunbar, for his constant encouragement, thought-provoking and meticulous feedback, and fostering of an exciting and inspiring working environment within the Social and Evolutionary Neuroscience Research Group. It was truly an honour and a privilege to work within a research group at the cutting-edge of evolutionary psychology and social neuroscience as we debated the origin of warfare, altruism, music, religion, language, social networks and complex societies. Secondly, to Andrew King for introducing me to many of the experimental methods contained in Chapter I. Both Robin and Andy first accepted my ideas for the various projects contained herein with just the right mix of caution and enthusiasm. At no point did their enthusiasm wane, and much of what is good about this thesis is due to their advice, suggestions and encouragement.

I would also humbly like to thank my external examiner Daniel Nettle and internal examiner Kate Nation for graciously agreeing to read over my thesis in its entirety and attend my viva voce examination. I sincerely hope they found at least some aspects of this manuscript interesting and engaging.

Thanks are also due to the Social and Evolutionary Neuroscience Research Group, and collaborators to our group, who have taken the time to comment on my work, send me papers, or otherwise assist and provide encouragement. In no particular order, these include: Max van Duijn, Jacques Launay, Anna Machin, Tamas David-Barrett, James Carney, Oliver Curry, Eiluned Pearce, Emma Cohen and Wouter Wolf. Although not directly affiliated with our group, for general intellectual admiration and inspiration, I thank Thom-Scott Phillips, Gareth Roberts and Emma Cohen. For their general statistical expertise, experience and assistance, I thank Sofia Massa and Lori Minini. For granting me the opportunity, even if exhausting experience, of teaching and marking two academic terms of undergraduate Core Practicals in cognitive and behavioural neuroscience, I thank Lori Minini. For subjecting themselves to my often odd and occasionally confusing psychology research studies, I thank all of the research assistants and participants who gave their time and attention to my experiments. Regrettably, for lack of space, I am unable to provide all of their names here. Lastly, the research for and preparation of this thesis was made possible because of the Department of Experimental Psychology and St Cross College at the University of Oxford as well as the very generous funding from a European Research Council Advanced Grant Studentship. I gratefully acknowledge and appreciate their generous support. Not all of these people or institutions will agree with what I've said, done or discovered, of course, and almost none of them are responsible for any mistakes contained herein.

If all I had done over the past four years was work on this thesis, I would have clearly lost the plot long ago, so a heartfelt thank you goes to the music I've listened to and shows I've attended over this period: Jane's Addiction, Red Hot Chili Peppers, Throwing Muses, Ministry, PJ Harvey, My Bloody Valentine and Smashing Pumpkins. In the same vein, I also thank the inspiring city of Oxford, as well as surrounding beautiful cities of Europe and the United Kingdom I've occasionally had the opportunity to travel to during interim breaks.

Second, there are two individuals who weren't even in my life when I started my doctoral studies: Maria Magdalena Arqueros Fernandez and Michal Perlstein whom I thank for their contributed support and enriched life experiences. And for their unquestioning love, pride and support: thanks and much love to my parents Kathy and Eric Burke and Fred Oesch and my brothers Aerian Oesch, Michael Burke and Graeme Burke. I hope they know I'm equally proud of them.

Contents

Abstract ii

Declaration iv

Acknowledgements v

1 The adaptive significance of human language 1

1.1 Introduction 1

1.1.1 Selective scenarios for the emergence of human language 3

1.2 Group bonding and social cohesion 4

1.2.1 Comparative animal behaviour 5

1.2.2 Linguistic anthropology/developmental psychology 6

1.2.3 Human social behaviour 9

1.3 Cooperation 10

1.3.1 Comparative animal behaviour 10

1.3.2 Developmental psychology 12

1.3.3 Human social behaviour 13

1.4 Sexual selection 14

1.4.1 Comparative animal behaviour 15

1.4.2 Developmental psychology 16

1.4.3 Human social behaviour 18

1.5 Cultural learning 20

1.5.1 Comparative animal behaviour 20

1.5.2 Developmental psychology 22

1.5.3 Human social behaviour 23

1.6 Mental tool 26

1.6.1 Comparative animal behaviour 26

1.6.2 Developmental psychology 27

1.6.3 Human social behaviour 28

1.7 Deception 29

1.7.1 Comparative animal behaviour 30

1.7.2 Developmental psychology 32

1.7.3 Human social behaviour 34

1.8 Discussion 36

2 Human group performance in social foraging: An empirical test of the interdependence hypothesis 40

2.1 Introduction 40

2.2 Methods 43

2.2.1 Participants 43

2.2.2 Ethics permission 44

2.2.3 Materials 44

2.2.3.1 Questionnaire measures 45

2.2.4 Procedure 46

2.2.5 Experimental design 47

2.2.5.1 Dependent variables 48

2.2.6 Statistical analysis 49

2.3 Results 50

2.3.1 Group performance effects 50

2.3.2 Self-report questionnaire items 54

2.3.3 Communication and group size effects 57

2.4 Discussion 65

2.4.1 Summary of results 65

2.4.2 Implications 66

2.4.3 Limitations 70

3 Honest signalling theory and human language: An experimental test of four hypotheses 73

3.1 Introduction 73

3.1.1 Vocabulary size as an honest signal 75

3.1.2 Self-disclosure as an honest signal 77

3.2 Experiment I: Effect of low-frequency vocabulary vignettes on heterosexual attractiveness ratings 83

3.2.2 Methods 83

3.2.2.1 Participants 83

3.2.2.2 Materials 84

3.2.2.3 Procedure 85

3.2.2.4 Experimental design 86

3.2.3 Results 86

3.2.3.1 Manipulation check 86

3.2.3.2 Attraction measures 88

3.2.4 Discussion 89

3.3 Experiment II: Effect of poetic vocabulary vignettes on heterosexual attractiveness ratings 91

3.3.1 Methods 91

3.3.1.1 Participants 91

3.3.1.2 Materials 92

3.3.1.3 Procedure 93

3.3.1.4 Experimental design 93

3.3.2 Results 94

3.3.2.1 Manipulation check 94

3.3.2.2 Attraction measures 96

3.3.3 Discussion 97

3.4 Experiment III: Effect of public speaking vignettes on heterosexual attractiveness ratings 99

3.4.1 Methods 100

3.4.1.1 Participants 100

3.4.1.2 Materials 100

3.4.1.3 Procedure 101

3.4.1.4 Experimental design 102

3.4.2 Results 102

3.4.2.1 Manipulation check 103

3.4.2.2 Attraction measures 106

3.4.3 Discussion 108

3.5 Experiment IV: Effect of intentionally deceptive vignettes on heterosexual attractiveness ratings 109

3.5.1 Methods 110

3.5.1.1 Participants 110

3.5.1.2 Materials 111

3.5.1.3 Procedure 112

3.5.1.4 Experimental design 113

3.5.2 Results 113

3.5.2.1 Manipulation check 113

3.5.2.2 Attraction measures 117

3.5.3 Discussion 125

3.5.4 General discussion 127

3.5.4.1 Summary of results 127

3.5.4.2 Implications 129

3.5.4.3 Limitations 130

4 The evolution of grammar in human language: Mentalising performance predicts recursive clause task performance 134

4.1 Introduction 134

4.2 Methods 140

4.2.1 Participants 140

4.2.2 Materials 141

4.2.3 Procedure 142

4.2.4 Experimental design 142

4.2.4.1 Dependent and independent variables 145

4.3 Results 145

4.4 Discussion 152

4.4.1 Summary of results 152

4.4.2 Implications 155

4.4.3 Limitations 159

5 Discussion 161

5.1 Overview 164

5.2 Broader implications 168

5.3 Methodological challenges and limitations 175

5.3.1 Chapter II: The role of language for cooperation during human social foraging 175

5.3.2 Chapter III: The role of honest signalling theory and language function 178

5.3.3 Chapter IV: Mentalising, recursive syntax and language form 179

5.4 Future directions for research 181

5.4.1 Further tests of the social complexity hypothesis 181

5.4.2 Developing accurate measures of social bonding 183

5.4.3 Experimental tests of the social bonding hypothesis 184

5.5 Final conclusions 187

References **189**

Appendix **213**

A.2 Chapter 2 213

A.2.1 Social foraging questionnaire 214

A.3 Chapter 3 215

A.3.1 Male protagonist small and large group public speaking vignettes 216

A.3.1.1 Version 1: Small group public speaker 217

A.3.1.2 Version 2: Large group public speaker 217

A.3.2 Male protagonist vignettes with before-and-after descriptions 218

A.3.2.1 Version 1: Self-enhancing attorney before/after description 218

A.3.2.2 Version 1: Self-enhancing janitor before/after description 218

A.3.2.3 Version 2: Self-deprecating attorney before/after description 219

A.3.2.4 Version 2: Self-deprecating janitor before/after description 219

A.4 Chapter 4 220

A.4.1 Socio-cognitive mentalising task 221

A.4.1.1 Martin, Simon, Charlotte and Jane's story 221

A.4.1.1.1 Questions 221

A.4.1.2 Frank, Betty and Brian's story 223

A.4.1.2.1 Questions 223

A.4.1.3 Gavin, Peter, Fiona and Sophie's story 225

A.4.1.3.1 Questions 225

A.4.2 Proportion of true-false answers to questions for each socio-cognitive story 227

List of Figures

2.1 Overhead view of the experimental foraging arena 45

2.2 Time-lapse overhead view of the experimental foraging arena 48

2.3 Mean group consensus time for small and large group sizes 51

2.4 Group performance measured as the ratio of 'good' to 'bad' tokens for small and large group sizes 52

2.5 The relationship between percentage of 'good' tokens and group size 53

2.6 The relationship between consensus time and group size 54

2.7 The relationship between mean number of gesturing subjects and group size 58

2.8 The relationship between mean number of speaking subjects and group size 59

2.9 The relationship between mean number of gesturing subjects and group consensus time 60

2.10 The relationship between mean number of speaking subjects and group consensus time 61

2.11 The relationship between mean number of gesturing subjects and percentage of 'good' tokens collected 62

2.12 The relationship between mean number of speaking subjects and percentage of 'good' tokens collected 63

3.1 Mean values of major dependent variables for male and female target persons for low-frequency vocabulary ratings 88

3.2 Mean values of major dependent variables for male and female target persons for poetic vocabulary ratings 96

3.3 Mean values of major dependent variables for male and female target persons as a function of group size 103

3.4 Mean values of major dependent variables for male and female target persons as a function of speaker performance 104

3.5 Mean values of major dependent variables for attorney and janitor vignettes as a function of truth value 115

3.6 Mean values of major dependent variables for male and female target persons for janitor vignettes 116

3.7 Mean values of major dependent variables for male and female target persons for attorney vignettes 120

4.1 The level of intentionality, recursive syntax and memory at which participants passed perspective-taking questions 146

4.2a The relationship between performance on recursive syntax and intentionality tasks 147

4.2b The relationship between performance on intentionality and recursive syntax tasks 148

4.3 The relationship between performance on memory and mindreading tasks 149

4.4 The relationship between performance on memory and embedded recursive syntax tasks 150

4.5 Bayesian network analysis of intentionality, recursive syntax and memory as dependent variables depicted as a graphical model 152

List of Tables

Table 2.1 Mean group size and pairwise correlation values between percentage of good tokens collected and consensus time for each self-report questionnaire item 56

Table 2.2 Multiple linear regression analysis for variables predicting (a) ln-transformed consensus time and (b) ln-transformed percentage of 'good' tokens collected 64

Table 3.1 Experiment I: ANOVA results summary (dependent variable: intelligence) 87

Table 3.2 Experiment I: ANOVA results summary (dependent variable: sexually attractive) 88

Table 3.3 Experiment I: ANOVA results summary (dependent variable: desirable as a date) 89

Table 3.4 Experiment II: ANOVA results summary (dependent variable: intelligence) 95

Table 3.5 Experiment II: ANOVA results summary (dependent variable: sexually attractive) 96

Table 3.6 Experiment II: ANOVA results summary (dependent variable: desirable as a date) 97

Table 3.7 Experiment III: Small group (version 1) and large group (version 2) public speaking vignettes 101

Table 3.8 Experiment III: ANOVA results summary (dependent variable: public prominence) 104

Table 3.9 Experiment III: ANOVA results summary (dependent variable: awkwardness) 105

Table 3.10 Experiment III: ANOVA results summary (dependent variable: intelligence) 105

Table 3.11 Experiment III: ANOVA results summary (dependent variable: sexually attractive) 106

Table 3.12 Experiment III: ANOVA results summary (dependent variable: desirable as a date) 107

Table 3.13 Experiment III: Complete tabularised summary of the ANOVA results for significant relationships with respect to each personality dimension, measured for both small and large group public speakers, within socially relevant and socially awkward contexts 108

Table 3.14 Experiment IV: Boastful (version 1) and modest (version 2) protagonist vignettes 111

Table 3.15 Experiment IV: ANOVA results summary (dependent variable: honesty overstated janitor) 114

Table 3.16 Experiment IV: ANOVA results summary (dependent variable: honesty understated janitor) 114

Table 3.17 Experiment IV: ANOVA results summary (dependent variable: honesty overstated attorney) 115

Table 3.18 Experiment IV: ANOVA results summary (dependent variable: honesty understated attorney) 116

Table 3.19 Experiment IV: ANOVA results summary (dependent variable: sexually attractive understated janitor) 117

Table 3.20 Experiment IV: ANOVA results summary (dependent variable: desirable as a date understated janitor) 118

Table 3.21 Experiment IV: ANOVA results summary (dependent variable: sexually attractive overstated janitor) 119

Table 3.22 Experiment IV: ANOVA results summary (dependent variable: desirable as a date overstated janitor) 119

Table 3.23 Experiment IV: ANOVA results summary (dependent variable: sexually attractive overstated attorney) 121

Table 3.24 Experiment IV: ANOVA results summary (dependent variable: desirable as a date overstated attorney) 122

Table 3.25 Experiment IV: ANOVA results summary (dependent variable: sexually attractive understated attorney) 123

Table 3.26 Experiment IV: ANOVA results summary (dependent variable: desirable as a date understated attorney) 123

Table 3.27 Experiment IV: Complete tabularised summary of the ANOVA results for significant relationships with respect to each personality dimension, measured for all four protagonists 125

Table 4.1 Mean values of major independent and dependent variables 147

Table 4.2 Multiple linear regression analysis for variables predicting level of recursive syntax performance 151

Table 4.3 Multiple linear regression analysis for variables predicting level of intentionality performance 151

Chapter I

The adaptive significance of human language

1.1 Introduction

Language is arguably one of the most salient features that distinguish humans from other animal species (Deacon, 1997; Fitch, 2010; Radick, 2007). Although many species have evolved the capacity to communicate information about aspects of their social and ecological environments (Bradbury and Vehrencamp, 2011; Cheney and Seyfarth, 1990; Janik et al. 2006; Searcy and Nowicki, 2005), no other species apart from humans are able to transmit abstract ideas, complex information, and unique perspectives through language, radically influencing others with respect to a complex cultural environment (Barrett, Dunbar and Lycett, 2002). Moreover, humans form complex social networks, both in modern industrial societies (Jackson, 2010) as well as traditional hunter-gatherer cultures (Apicella et al. 2012), in which enormous amounts of information are exchanged among highly inter-connected individuals in the social network (Easley and Kleinberg, 2010).

Several biological functions for human language have been posited, especially in social network contexts, including cultural learning (Baumeister et al. 2004; De Backer and Gurven, 2006), social cohesion (Dunbar, 1996, 2004), and the enforcement of social norms (Kniffin and Wilson, 2005; Wilson et al. 2000). Perhaps most fundamentally, however, language can provide communicators with information about the surrounding environment, especially information about other individuals in the social network whom communicators may interact with at some point in the future (Craik, 2009; Laidre et al. 2012). Experimental studies of human gossip have shown individuals integrate information from multiple sources,

comparing and contrasting various pieces of information, to distinguish true from false messages and ultimately converge upon the truth (Collins et al. 2011; Hess and Hagen, 2006; Sommerfeld et al. 2008). It has furthermore been argued that such cognitive heuristics may be an innate and universal aspect of human cognition and behaviour that evolved in response to the demands of large social groups (Dunbar, 1996, 2003, 2004; Hess and Hagen, 2006; Wilson et al. 2000).

Similar results have been reported in various mathematical modelling approaches (Dávid-Barrett and Dunbar, 2013; Laidre et al. 2012). For instance, agent-based simulation models have revealed that the evolution of communicative complexity is dependent on an ecological demand for large social groups (Dávid-Barrett and Dunbar, 2013), as fitness increases with group size as the network gains more independent sources of information (Laidre et al. 2012). Addressing these questions with a purely mathematical or social network approach, however, is prohibitively limited in allowing more definitive conclusions with respect to how human language may have gotten off the ground in the first instance. Although exceedingly valuable, such disparate and incomplete approaches cry out for a multifactorial synthetic approach, which integrates theory and experimental data from multiple perspectives.

Despite the existence of a large body of relevant theoretical and empirical research, there is currently no consensus as to how language originated in the human species (Christiansen and Kirby, 2003; Larson, Déprez and Yamakido, 2010; Számadó and Szathmáry, 2006). Paradoxically, this lack of consensus is often attributed in part to the recent burgeoning of evolutionary linguistics, which has seemingly resulted in an unceasing variety of hypotheses, each of which attempts to provide a unique selective scenario for the evolution of human language (Számadó and Szathmáry, 2006). Unfortunately, many of these hypotheses are often given essentially equivalent credibility, despite the fact that many of

them suffer from significant theoretical and empirical shortcomings. Most critically, it seems, there has been no systematic evaluation of each of these hypotheses in light of recent research in comparative and developmental psychology, sociolinguistics, anthropology and evolutionary psychology.

The following review analyses the evidence and theoretical arguments for each of these proposals, and attempts to provide, in place, a synthetic perspective which integrates or eliminates each of these ostensibly competing hypotheses for the evolution of human language. The burden of theoretical plausibility, and the weight of current debate, rests largely on the issue of functional uses for gossip in social network contexts. Accordingly, disproportionate space is dedicated to these sections, and to the discussion of empirical research which bears weight on this question, including evidence from human social behaviour, comparative animal behaviour, developmental psychology and linguistic anthropology. Further, as language is a behaviour and does not fossilise, I deliberately avoid speculative accounts that invoke archaeological claims as evidence, as is often done by many theorists (Számadó, 2010; Bickerton and Szathmáry, 2011). I conclude by offering several different possibilities for future research programs which may provide further additional insight into the evolutionary mechanisms and biological function(s) of human language.

1.1.1 Selective scenarios for the emergence of human language

In a recent review article, Számadó and Szathmáry (2006) list eleven unique hypotheses for the emergence of human language, including: gossip, grooming, group bonding, cooperative hunting, mental toolkit, mating contract and/or pair bonding, motherese, sexual selection, song, status, and tool making. For present purposes, I characterise the main competing theories for the evolution of language according to six different selective scenarios (including various other subcategories): 1) *group bonding and social cohesion,* maintenance

of frequent reciprocal relationships characterised by stability, predictability, relaxation and reduced hostility (Silk, Cheney and Seyfarth, 2013) (including gossip, grooming, group bonding, pair bonding, motherese and song) 2) *cooperation,* working or acting together for group common/mutual benefit (Tomasello, 2008) (including cooperative foraging and hunting) 3) *sexual selection,* honest signals of genetic fitness (Burling, 1986; Miller, 2000; Rosenberg and Tunney, 2008) (including social status and intelligence) 4) *cultural learning,* gain of information based on the experiences of others (Baumeister et al. 2004; De Backer and Gurven, 2006) (including pedagogy, tool making and enforcement of social norms) and 5) *mental tool*, gain of self-awareness and voluntary control of our actions (Burling, 1993). For present purposes, I further distinguish an additional category 6) *deception,* whereas a signaller benefits from the breakdown of the normative correlation between signaller characteristic and external attribute (Searcy and Nowicki, 2005), as various theoretical proposals have argued it to be a plausible selective scenario for the emergence of language (Dawkins and Krebs, 1978; Krebs and Dawkins, 1984; Scott-Phillips, 2006).

1.2 Group bonding and social cohesion

Dunbar (1993, 1996, 2003, 2004) has suggested that as human groups increased in social complexity over the course of human evolution, language may have been the critical factor that replaced primate social grooming and facilitated social bonding with increasing group size. According to this social complexity hypothesis for animal communication, language is a fundamentally social phenomenon, and may have been the selective solution for evolution to the problem of maintaining stable cohesion in large social groups (Dunbar, 2003, 2004; Freeberg, Dunbar and Ord, 2012). Several studies provide comparative and correlational evidence in support of the social complexity hypothesis, in a wide range of

species, including ground squirrels, bats, birds, and primates (Blumstein and Armitage, 1997; Freeberg, 2006; Freeberg, Dunbar and Ord, 2012; Kroodsma, 1977; Maestripieri, 1999; McComb and Semple, 2005; Wilkinson, 2003).

1.2.1 Comparative animal behavior

The social complexity hypothesis for animal communication posits that groups with complex social systems require more complex communicative systems to regulate interactions and relations among group members (Freeberg, Dunbar and Ord, 2012). Complex social systems, compared with simple social systems, are those in which individuals frequently interact in many different social contexts with many different individuals, often repeatedly with many of the same individuals in networks over time. Complex communicative systems, compared with simple communicative systems, are those which contain a large number of structurally and functionally distinct elements or possess a high number of distinct bits of information (Freeberg, Dunbar and Ord, 2012).

For instance, in a comparative meta-analysis of 42 different non-human primate species, McComb and Semple (2005) found both social group size and social grooming time were strongly positively associated with vocal repertoire size across these species. Similar results have been found using more direct comparisons between closely related species (Gustison, Roux and Bergman, 2012; Wilkinson, 2003). In particular, one comparative study found larger vocal repertoires associated with the long-term courtship bonds found in structurally large groups of gelada baboons, when directly compared to the more temporary courtships found in comparatively smaller chacma baboon societies (Gustison, Roux and Bergman, 2012). A similar effect has also been noted in both birds and bat species (Freeberg, 2006; Wilkinson, 2003). For example, at least one recent study has shown that in both naturalistic field settings and experimentally manipulated aviary environments, Carolina

chickadees in larger social groups used calls with greater vocal complexity than the calls of individuals in smaller social groups (Freeberg, 2006). In a similar case with greater generalizability, Wilkinson (2003) compared the complexity of infant isolation calls of eight different bat species and found a strong positive relationship between colony size and the amount of information in their isolation call systems.

1.2.2 Linguistic anthropology/developmental psychology

With respect to human vocal communication, similar results have been found across a wide variety of the world's different languages (Atkinson, 2011; Lupyan and Dale, 2010). For example, a recent meta-analysis of over 500 different languages found the number of phonemes – perceptually distinct units of sound that differentiate words – in a language is positively associated with the size of its speaker population (Atkinson, 2011). Moreover, when both social group size and social cohesion were taken into account, Lupyan and Dale (2010) found – in a statistical analysis of more than 2,000 languages – the same relationship applies for more complex aspects of language including lexical constructions and morphology. For instance, language speakers with less diverse inflectional morphology were generally identified as part of an *exoteric* niche; typically, non-native speakers or those who use the language to speak with millions of strangers from different ethnic or linguistic backgrounds. In contrast, those languages in the *esoteric* niche, while still most often present in sizable groups of several thousand individuals, were much more likely to be spoken by more socially cohesive groups than languages with many millions of speakers (Lupyan and Dale, 2010). Taken together, these results are consistent with the hypothesis that increases in social complexity (i.e. both group size and social cohesion) may drive increases in vocal complexity (Freeberg, Dunbar and Ord, 2012).

Further evidence of the utility of language for regulating interactions and relations among group members, including the promotion of social identity and group cohesion, has come from the peculiar aspects of linguistic diversity (Cohen, 2012; Deutscher, 2005; Nettle, 1999; Nettle and Dunbar, 1997), including accent, in particular (Cohen, 2012). In a comprehensive review of accent as a social marker, Cohen (2012) notes that language contains valuable social signalling information, irrespective of the actual semantic content in the message conveyed. Accent is, according to Cohen (2012) and several authors (Fuertes, Potere, and Ramirez, 2002; Gluszek and Dovidio, 2010), one of the most important cues of social identity that language provides. For instance, while any two people may use exactly the same words, they may nevertheless be perceived and trusted very differently, depending on their accent or other aspects of linguistic variation (Cohen, 2012; Giles, 1973; Lev-Ari and Keysar, 2010; Nettle and Dunbar, 1997). Nettle and Dunbar (1997) cite one study, in particular, in which participants were asked to rate strangers on the basis of their linguistic variety, and found they ranked their own speech higher on scales of solidarity, friendliness and altruism, but notably not on similar scales which measured competence or social status. Nettle and Dunbar (1997) further cite several additional studies, showing certain highly valued language varieties greatly increase altruistic gestures from strangers.

A recent series of controlled laboratory studies have given further weight to this claim, showing accent is typically a more significant cue than race in determining children's affinity toward other speakers, ranging from 6-month olds to 5-year olds (Kinzler, Dupoux, and Spelke, 2007; Kinzler et al. 2009; Kinzler, Corriveau, and Harris, 2011). Kinzler, Corriveau and Harris (2011), in particular, showed that 5-year olds preferentially learned from native-accented over foreign-accented speakers, despite the fact that both produced semantically meaningless and unintelligible utterances. Similar findings have also been noted in adults, indicating that accent categorization may be a dedicated dimension of social

categorization (Pietraszewski and Schwartz, 2014a; Pietraszewski and Schwartz, 2014b; Rakić, Steffens and Mummendey, 2011). However, it seems clear that social categorization, by itself, is unlikely to be the ultimate biological function, as other research has found that social identity discrimination plays an important role in promoting group integrity, by increasing member loyalty to the group (Van Vugt and Hart, 2004).

In her summary of the accent literature, Cohen (2012: 600) notes that accent may well be an honest signal of group membership and affinity, in that "language acquisition research provides evidence that accent is cheap to acquire in childhood but hard to fake in adulthood" and that "the developmental literature…provides…evidence that accent is a robust and salient guide to early-developing social preferences". Indeed, Cohen (2012) further posits a plausible scenario for the emergence of language itself, as she notes, the ability to produce and discriminate accent (as a means of distinguishing social identity and group membership) has apparent homologies in the vocalizations of many animal species, including killer whales (Yurk et al. 2002), bottlenose dolphins (Janik, Sayigh, and Wells, 2006), greater spear-nosed bats (Boughman and Wilkinson, 1998), Mexican jays (Hopp, Jablonski, and Brown, 2001), yellow-naped Amazon parrots (Wright and Wilkinson, 2001), European starlings (Hausberger, Bigot, and Clergeau, 2008), and even chimpanzees (Herbinger et al. 2009).

Finally, it is noteworthy the aforementioned analysis of accent as a social marker has also been applied more broadly to the phenomenon itself of language diversity (Deutscher, 2005; Nettle, 1999). Namely, the role of particular languages and dialects as a social marker of group membership has been well documented by many sociolinguists, and may therefore not just apply to accent or other aspects of phonetics or morphology (Labov, 1980; 2001; Richerson and Boyd, 2010). In addition, a considerable body of developmental literature strongly suggests the richness and complexity of infant-to-parent vocal interactions are likely related to social bonding and the solicitation of parental investment (Oller and Griebel, 2006).

Consequently, it is conceivable the human phenomenon of nearly 7,000 mutually unintelligible world languages are not an accidental by-product of an underspecified genome (Pinker and Bloom, 1990), but instead constitute a unique adaptation for indicating group membership, facilitating group cohesion and demarcating the social in-group from various social out-groups (Deutscher, 2005; Dunbar, 2003; Nettle and Dunbar, 1997; Nettle, 1999; Roberts, 2010; Roberts, 2013).

1.2.3 Human social behaviour

Additional evidence for the social cohesion function of language comes from detailed studies of human social behaviour. Analyses of actual human conversational content, both in industrial Western cultures as well as more traditional societies, reveal that social gossip topics overwhelmingly dominate most everyday conversations; namely two-thirds of total conversation time (Bischoping, 1993; Dunbar et al. 1997; Haviland, 1991). Moreover, the frequency and ubiquity with which humans engage in social gossip likely has important functional implications within the wider network of social relationships, beyond mere idle chatter. For instance, the spread of reputational information through social gossip promotes prosocial behavior by facilitating partner selection when circumstances require cooperation (Feinberg, Willer and Shultz, 2014). Moreover, Weaver and Bosson (2011) found feelings of interpersonal closeness were facilitated through social gossip; in particular, the effects seemed to be strongest when participants shared a weakly held, negative view of an absent third party. In addition, research has further shown the social use of language may have important implications for social bonding within the context of romantic relationships. Ireland et al. (2011), in particular, found among speed-dating dyads, both romantic initiation and follow-up relationship stability were positively predicted by unconscious conversational language style matching (LSM) scores (determined as a composite index of similar pronouns,

articles, conjunctions, prepositions, verbs, adverbs, negations and quantifiers used in conversation).

In sum, although these studies do not conclusively demonstrate language evolved for the purpose of group social cohesion, such findings are clearly much harder to reconcile with competing hypotheses, such as deception, cooperation, or cultural learning, as the primary biological function for human language. Moreover, as in any biological system that has significant time and energetic costs as well as adaptive benefits (Barrett, Dunbar and Lycett, 2002), it certainly begs the question as to why we spend most of our time in the social use of language, if language primarily evolved for these alternative function(s).

1.3 Cooperation

In contrast to the social cohesion hypothesis for human language function, other theorists have argued that language evolved for the purpose of mutualistic cooperation (Tomasello et al. 2012; Tomasello, 2008). Cooperation, in this sense, is defined as interdependent collaboration with a shared goal, such as group foraging or hunting, as opposed to general altruism, reciprocity or social cohesion (Tomasello et al. 2012). According to this hypothesis, language and cooperation coevolved together in a two stage process: 1) small group cooperation and communication began as the starting point, mainly in the context of cooperative foraging and 2) as group size increased, over the course of human evolution, large group cooperation and communication followed as the entire community needed to work together, often in the context of between-group competition, leading to culture, cultural norms, and complex institutions (Tomasello et al. 2012).

1.3.1 Comparative animal behaviour

As a general rule, chimpanzees and bonobos, as our nearest common ancestor, are almost exclusively individual foragers (Tomasello et al. 2012). Although they may occasionally search for food in small groups, when they eventually discover any resources, each individual typically obtains and consumes the food on its own, while sharing only in special cases (Tomasello et al. 2012). One possible exception to this includes work by Boesch and Boesch (1989) and Watts and Mitani (2002), where male chimpanzees have been observed hunting monkey in small groups; however, typically in these scenarios one individual chimpanzee leads the chase and captures the monkey, while the other chimpanzees block potential escape routes. Therefore, in the majority of cases, chimpanzees appear to hunt opportunistically, although in other cases they do seem to follow directed hunting patrols (Mitani and Watts, 1999). Prior to such patrols, chimpanzees often emit a hunting call, usually emitted to neighbouring conspecifics on sight of their prey; nevertheless, they do not appear to require more sophisticated gestural or vocal signalling in order to capture their target (Mitani and Watts, 1999). Moreover, controlled laboratory studies have found similar results, in that chimpanzees do not generally participate in collaborative activities that require joint attention and involve shared goals, nor do they implement vocal or gestural communication to aid cooperative activities (Warneken, Chen and Tomasello, 2006).

In general, chimpanzee vocalizations normally occur outside the context of group hunting (Goodall, 1986); additionally, they depend on vocal learning and are highly group-specific, suggesting community specific pant hoots may instead serve an affiliative social bonding function (Crockford et al. 2004). Indeed, further evidence has suggested that the main function of chimpanzee hunting itself is likely related to male social bonding and coalition building (Mitani and Watts, 2001; Stanford, 1998). Lastly, although group hunting has been observed in roughly 22 vertebrate species (Boesch, 1994), including hyenas (Mills,

2003), lions (Packer et al. 1990), wolves (Mech and Boitani, 2003), and chimpanzees (Boesch, 1994; Watts and Mitani, 2002), current evidence suggests that they do so in the absence of directed gestural or vocal signalling, and certainly without language (Pinker, 1994).

1.3.2 Developmental psychology

Further evidence for the primacy of human cooperation has come from developmental research (Warneken, Chen and Tomasello, 2006). For instance, Warneken, Chen and Tomasello (2006) have shown pre-linguistic children are fully capable of directed acts of cooperation, while chimpanzees commonly fail to respond, much less understand similar tasks. However, although children often attempt to indicate a cooperative intention through gestural signalling, neither gestural nor vocal signalling seems to be required for effective cooperation in all cases (Warneken, Chen and Tomasello, 2006).

Nevertheless, some have argued that children often best learn word meanings through joint attention (e.g. between a mother and her infant on a word-object association), thereby positing a potentially collaborative foundation for the acquisition of language (Baldwin, 1995; Tomasello, 2003; Tomasello and Farrar, 1986; Tomasello and Todd, 1983). For example, Tomasello and Todd (1983) observed six mother-infant dyads at monthly intervals from 12 to 18 months of age, finding a high correlation between the amount of time infants spent in joint attention with their mothers and the size of the infants vocabulary at the end of this period. However, the proposal that joint attention or collaboration is in some way required for language acquisition is considered a minority viewpoint, as the majority of language acquisition research indicates otherwise (Akhtar and Gernsbacher, 2007; Pinker, 1999; Scofield and Behrend, 2011; Tomasello, 2009). Therefore, the hypothesis that human cooperation and communication are in some way ontogenetically interdependent, thereby

providing a plausible phylogenetic interpretation (Tomasello, 2009), remains an open question.

1.3.3 Human social behaviour

A third line of evidence invokes the need among early humans (and presumably extant hunter-gatherer societies) for cooperative foraging and hunting that would have required sophisticated communication (Bickerton and Szathmáry, 2011; Tomasello et al. 2012). Several authors, for instance, note that eusocial insects communicate displacement (i.e. reference to events remote in time or space), using signals which indicate the distance, direction and richness of food resources to the rest of the group (Bickerton and Szathmáry, 2011; Számadó, 2010). According to this argument, among early humans that had located a potential food resource, some form of communication would have been necessary to inform the rest of the group of the find (Bickerton and Szathmáry, 2011; Számadó, 2010).

However, as has been noted for chimpanzees, although they often emit a hunting call prior to the start of a chase, other more sophisticated forms of signalling do not appear to be necessary either to alert the rest of the group nor aid the chase itself (Mitani and Watts, 1999). Granted, it is conceivable that neither chimpanzees nor early humans would find much success in recruiting individuals greatly removed in time-and-space, beyond the reach of a simple call, thereby language providing some advantage (Bickerton and Szathmáry, 2011; Számadó, 2010). Nevertheless, given both chimpanzees (Mitani and Watts, 1999) and traditional hunter-gatherers often hunt in groups anyway (Smith, 1991), it is questionable why language would have been critical for additional recruits, especially given the proficiency of chimpanzees (and traditional societies in large part) to both recruit and hunt without language. Moreover, recent mathematical modelling attempts have demonstrated this proposal rests on a number of unwarranted assumptions, including speculations on early

hominin scavenging (Clark, 2011). A final argument often raised is that although distance and direction challenges may have been overcome through general primate social and vocal signalling behaviours (or gestural signalling perhaps analogous to eusocial insect signalling), the technical skills and cultural information associated with effective hunting would have definitely required language (Számadó, 2010).

To be sure, evidence from traditional societies indicates that hunter-gatherers do engage in extensive sharing of information, on the location of key resources, in advance of the actual search (Smith, 1991). However, hunting itself is typically done in silence (mainly to prevent alerting the prey), or utilizing only gestures, and in many cases, individuals hunt under solitary conditions (Smith, 1991). Nevertheless, it is important to note, although pre-hunt information exchange may constitute a valid argument, information transfer (even if done mainly in advance of hunting), is arguably a form of cultural learning. As such, it can and likely should be evaluated independently of the actual act of cooperative hunting; to this issue, I will return in the later section on cultural learning.

1.4 Sexual selection

In contrast to the cooperation hypothesis for human language function, other theorists have argued that language evolved for the purpose of sexual advertisement to prospective mates (Burling, 1986; Locke and Bogin, 2006; Miller, 2000). Sexual advertisement, in this sense, is defined as an honest signal of genetic quality or social status which increases the probability of reproductive success, typically grouped into two basic categories: receiver-dependent costs (i.e. those that reflect agonistic, mating and social costs) and receiver-independent costs (i.e. those that reflect underlying biology including production, development, and maintenance costs) (Searcy and Nowicki, 2005). For example, some have

proposed that vocabulary size may have evolved as an honest signal of human intelligence (as a receiver-independent cost), and should therefore be attractive to prospective mates (Miller, 1999, 2000; Rosenberg and Tunney, 2008). Alternatively, others have suggested that self-disclosure of revealing or culpable personal information, a kind of social risk-taking (as a receiver-dependent cost), may be a form of social advertisement, while also signalling imperviousness to danger and social repercussions (Donath, 2008; Bliege Bird and Smith, 2005).

1.4.1 Comparative animal behaviour

Over the past thirty years, a great deal of evidence has accumulated for sexual selection having influenced many animal communication systems, including insects, birds and frogs (Searcy and Nowicki, 2009). In most cases, selection appears to be biased for female preference of male signals (Searcy and Nowicki, 2009). The results from many studies, in both insects and anurans, indicate that females generally prefer male displays with greater acoustic energy, as measured by call amplitude, call rate, call duration, frequency, and asynchrony interval (Searcy and Nowicki, 2009). In birds, female preference is often based on several additional factors, including call rate, locality (i.e. preference for local songs over foreign ones), and song complexity (Searcy and Nowicki, 2009). Various functions have been posited for the adaptive significance of female preference, including honest signals of genetic quality and conspecific recognition in insects, birds and anurans (Ryan, 2013; Ryan and Rand, 1993; Searcy and Nowicki, 2009), as well as pair-bonding in monogamous birds (Bolund et al. 2012). In addition, some limited circumstantial evidence indicates sexual selection may have shaped some primate communication systems (Hauser, 1999). For example, in free-ranging vervet monkeys, it has been shown that high-ranking individuals are generally more vocal than middle to low-ranking individuals (Hauser, 1999).

However, as has been noted by several theorists, such selective factors may have limited applicability, when used to explain the emergence of human language (Briscoe, 2008; Fitch, 2005; Hurford, 2007; Zahavi and Zahavi, 1997). Firstly, if comparisons are to be made to other species (such as birds), with respect to genetic quality, it is typically the case that song complexity tends to be correlated with high levels of sexual selection (Hamao, 2012). Although it is not entirely clear what sort of data would be required to demonstrate high degrees of sexual selection for vocal complexity in human populations (but see Miller, 2000, for one potential argument), at present, it would seem fair to say there is limited evidence for this hypothesis (Briscoe, 2008; Fitch, 2005; Hurford, 2007; Zahavi and Zahavi, 1997). Second, sexually selected traits typically result in exaggerated characteristics in males, which are notably absent in females (Searcy and Nowicki, 2009). That is, as a viable hypothesis for the emergence of language, the genetic quality argument fails to explain the presence of equal degrees of vocal complexity, namely language, in both human males and females (Fitch, 2005; Hyde and Linn, 1988). Third, traits under sexual selection typically mature at puberty (Fitch, 2005); yet, by the time normal human infants are 17-months-old, they are able to link meanings to segmented words (Graf Estes et al. 2007). In any event, it is certainly worth noting vocal complexity can be sexually selected for other functions, apart from genetic quality, including species recognition and pair-bonding (Bolund et al. 2012; Ryan and Rand, 1993), both of which would seem more congruent with the social bonding hypothesis (Freeberg, Dunbar and Ord, 2012).

1.4.2 Developmental psychology

A critical assumption, of the sexual selection hypothesis for human language evolution (including those which emphasize either receiver-dependent costs or receiver-independent costs), is that several developmental milestones must necessarily first be reached

in order to acquire full adult language competency (Locke and Bogin, 2006; Miller, 2000). For instance, Locke and Bogin (2006) have argued that two developmental stages, namely childhood and adolescence, are relatively new evolutionary innovations unique to humans, and thereby provide the foundation for the emergence of language. However, several authors have pointed out the vast majority of social animals (which rely on communicative systems for reproductive success), have similar developmental stages that appear to be critical for adult communicative competency (Ragir and Brooks, 2006; Schleidt, 2006; Weisfeld, 2006). Moreover, altriciality is a relatively widespread phenomenon, and has been documented in many animal species, including birds and mammals, none of which have acquired a capacity for language (Eisenberg, 1981; Starck, 1998).

Nevertheless, Locke and Bogin (2006) maintain adolescence occupies a unique place within human development, whereby vocal and verbal performances increase fitness through the facilitation of status attainment and associated mating opportunities. In support of this thesis, Locke and Bogin (2006) cite examples of oratory, speech, attention and power in traditional societies, as well as numerous examples of verbal performance in both traditional and modern industrial societies. However, an important limitation of these examples is that most are largely anecdotal, with very little experimental evidence of an association between increased status or reproductive success. Granted, experimental research has identified a range of traits which tend to correlate with group leadership, including extraversion, sociability, and assertiveness, as well as verbal and social intelligence; even so, the causal evidence for a direct link between verbal proficiency and status or leadership, thereby leading to increased reproductive success, remains unclear (Van Vugt, 2006).

Nevertheless, Locke and Bogin (2006) draw further attention to recent work demonstrating a more direct link between vocal attractiveness and increased reproductive success (Hughes et al. 2004). However, most of the evidence for increased mating success

appears to be limited to aspects of language shared with many other species, such as formant dispersion (an aspect of language phonetics), and no experimental evidence currently suggests that aspects of morphology, syntax, semantics or pragmatics lead to increased reproductive opportunities (Locke and Bogin, 2006). Moreover, at least one critic notes the hypothesis requires males to be the primary benefactors of these vocal performances; problematically, however, life history evolution generally acts primarily on females, a difficulty which advocates of the hypothesis fail to address (Kappeler, 2006). In sum, several authors argue Locke and Bogin (2006) fail to demonstrate a necessary causal connection between novel developmental stages and sexual selection as the primary selective pressure for the evolution of language (Bickerton, 2006; Ragir and Brooks, 2006; Schleidt, 2006; Weisfeld, 2006).

1.4.3 Human social behaviour

As noted previously, sexual selection hypotheses are typically grouped into two basic categories: receiver-dependent costs (i.e. those that reflect agonistic, mating and social costs) and receiver-independent costs (i.e. those that reflect underlying biology including production, development, and maintenance costs) (Searcy and Nowicki, 2005). Hypotheses which have tended to emphasize receiver-independent costs have often focused on the peculiar intricacies of human vocabulary, while hypotheses which emphasize receiver-dependent costs have generally focused on status and mating opportunities (Burling, 1986; Miller, 2000; Rosenburg and Tunney, 2008).

For instance, human vocabulary size has been noted to differ quite dramatically among people, is at least 60 percent genetically heritable, and has about an 80 percent correlation with general intelligence (Jensen, 1998; Plomin et al. 1997; Stromswold, 2001). Moreover, intelligence has been established as one of the single most important traits desired

in a long-term partner for both males and females (Buss, 1989; Buss, Abbott and Angleitner et al. 1990). In addition, studies have also found that people in long-term relationships usually have vocabularies of similar sizes (Mascie-Taylor, 1988).

Further evidence for this hypothesis has shown that males do in fact use more low-frequency vocabulary when in the presence of attractive females, whereas females seem unmotivated to speak in this respect when in the presence of attractive males (Rosenberg and Tunney, 2008). All the same, the reverse association does not appear to be supported, in that recipients of linguistic fluency often judge users of low-frequency vocabulary as *less* intelligent than more plain-spoken individuals (Oppenheimer, 2006). However, this particular study did not specifically investigate use of low-frequency vocabulary in the context of mate attraction.

On the other hand, recent work by Donath (2008) and others have instead argued that human honest signalling theory should emphasize status and influence, with concurrent costs including the risk of unfavorable social reputation in an information-based society (Donath, in press; McCracken, 1998; Thornton, 1996). For example, at least one social anthropologist has argued that nowhere would this phenomenon be more salient among traditional societies than in the context of oratory and public speaking (Burling, 1986). Indeed, according to charismatic leadership theory, several characteristics are commonly attributed to extraordinary leadership ability, including having an appealing vision and a willingness to take personal risks to achieve that vision (Robbins and Judge, 2013). A range of traits have further been identified which tend to correlate with group leadership, including extraversion, sociability, and assertiveness, as well as verbal and social intelligence (Van Vugt, 2006) suggesting the salience brought by public speaking may be an honest signal of these traits (with concurrent costs of social embarrassment and unfavourable social reputation). However, such proposals have appeared to most theorists as speculative and unconvincing, as

there is currently inadequate empirical evidence to support either of these claims (Briscoe, 2008; Fitch, 2005; Hurford, 2007; Zahavi and Zahavi, 1997).

1.5 Cultural learning

In contrast to the sexual selection hypothesis for human language function, other proposals have argued that human language may have evolved for the function of gaining information based on the experiences of others (Baumeister et al. 2004; De Backer and Gurven, 2006; Pinker and Bloom, 1990). According to this hypothesis, listeners may benefit from the accrued advantageous experiences of other speakers, while avoiding the potentially costly pitfalls of mistakes or experiences gone poorly. In an often cited quotation, it is stated: "[there is] an obvious advantage to being able to acquire… information second-hand: by tapping in to the vast reservoir of knowledge accumulated by other individuals, one can avoid having to duplicate the possibly time-consuming and dangerous trial-and-error process that won that knowledge" (Pinker and Bloom, 1990: 712).

1.5.1 Comparative animal behaviour

The phenomenon of cultural learning, once thought unique to humans, has now been documented in a wide variety of species, including chimpanzees (Whiten et al. 1999), cetaceans (Rendell and Whitehead, 2001), black rats (Terkel, 1996), and guppies (Dugatkin, 1992) (although see Tomasello et al. 1993, for an alternative perspective). Chimpanzees, in particular, have been observed practicing over 65 different categories of behaviour, including the use of sticks, leaves, branches, and stones for communication, play, food gathering, eating, and comfort (Whiten et al. 1999). Moreover, the presence, absence and/or variety of these behaviours seem to vary dramatically between chimpanzee communities (Whiten et al.

1999). However, there is currently no evidence chimpanzees learn any of these culturally acquired skills through the use of language, nor directed gestural or vocal signalling (Whiten et al. 1999). Granted, at least in the context of group hunting, although chimpanzees often emit a hunting call prior to the start of a chase, other more sophisticated forms of signalling do not appear to be necessary either to alert the rest of the group or aid the chase itself (Mitani and Watts, 1999). Moreover, there is absolutely no evidence that chimpanzees engage in gestural or vocal planning or strategizing prior to the start of a hunt (Boesch and Boesch, 1989; Mitani and Watts, 1999; Tomasello et al. 2012; Watts and Mitani, 2002), nor for that matter, any other culturally transmitted behaviour (Whiten et al. 1999) (although see Bard 1993, for an alternative account of "very subtle tutorial activity" in a sign-language trained chimpanzee).

Nevertheless, chimpanzees do appear to have preferred roles with different duties during the hunting pursuit, including those of the driver, chaser, blocker, and ambusher (Boesch, 2002). Moreover, a few social carnivores, including lions, seem to engage in similar behaviours (Stander, 1992). Puzzlingly, advocates of the cultural learning hypothesis often acknowledge these facts, but dismiss them; at least in the case of chimpanzees, it is argued hunting roles must be learned, and due to the absence of language, require 20 years to master though cultural learning (Számadó, 2010). However, studies of hunter-gather societies have revealed the presence of similar training periods, in that young males typically begin to hunt at around age 15, and do not achieve full mastery until about 35 years of age (Boesch, 2002; Kaplan et al. 2000). In sum, although extensive periods of cultural learning are necessary for effective group hunting, in both chimpanzees and humans, it would appear debateable whether sophisticated gestural or vocal signalling are critical, given effective competence in chimpanzees. Such findings draw further attention to the fact that group hunting has been observed in over 22 vertebrate species (Boesch, 1994), including hyenas (Mills, 2003), lions

(Packer et al. 1990), and wolves (Mech and Boitani, 2003), all of which do so in absence of directed gestural, vocal or linguistic communication (Pinker, 1994).

1.5.2 Developmental psychology

Advocates of the cultural learning hypothesis often claim human ontogeny as an important period of social learning which allows for the fidelity of information transmission among conspecifics (Tomasello et al. 1993). Moreover, it has been claimed that human ontogenetic cultural learning manifests in three unique developmental stages; *imitative learning* (age 9-14 months), *instructed learning* (age 4-5 years), and *collaborative learning* (age 6-7 years) (Tomasello et al. 1993). As defined, *imitative learning* is often typified as a learner's internalization of a demonstrator's behavioural strategies; *instructed learning*, as learning in which a learner internalises the instructions from a teacher for subsequent attentional and cognitive self-regulation; and *collaborative learning*, as learning which requires peers working together to construct something which did not exist prior to the interaction (Tomasello et al. 1993).

Granted, it is beyond doubt that children acquire complex socio-cognitive abilities throughout development; however, it is questionable whether such stages exist in the first place and perhaps more importantly, if so, whether they provide sound evidence for the importance of human language as the primary conduit for cultural information transmission (Hauser, 1993). For instance, as Tomasello et al. (1993) themselves concede, many children in both Western and non-Western cultures are often instructed by means of nonverbal demonstration; as such, cultural acquisition by way of verbalised *instructed learning* would seem largely inapplicable in these cases (Greenfield and Lave, 1982; Rogoff, 1990). Nevertheless, Tomasello et al. (1993) argue that during the *collaborative learning* stage, considerable planning, responding, clarification, criticism and integration occur, using

reflective and recursive language intended to come to a joint consensus (e.g. involving a problem-solving task).

However, collaborative learning has been argued to exist in many nonhuman primates (without the use of directed gestural or vocal signalling), including chimpanzees, typically in the context of group hunting and other social activities (Bard, 1993; Hauser, 1993; Stammbach, 1988; Watts and Mitani, 2002). Moreover, as has been noted by several commentators, this particular ontogenetic account, as it stands, lacks an explicit functional proposal of how or why collaboration, or cultural information transmission, should evolve in the first place (Collier, 1993; Hauser, 1993).

1.5.3 Human social behaviour

As several studies have now shown social gossip topics consume nearly two-thirds of total conversation time (Bischoping, 1993; Dunbar et al. 1997; Haviland, 1991), more recent studies of cultural learning in humans often assume a fundamental role for social gossip (Baumeister et al. 2004; Collins et al. 2011; De Backer and Gurven, 2006; Hess and Hagen, 2006; Sommerfeld et al. 2008). For instance, experimental studies of human gossip have suggested that individuals integrate information from multiple sources, comparing and contrasting various pieces of information, to distinguish true from false messages and ultimately converge upon the truth (Collins et al. 2011; Hess and Hagen, 2006; Sommerfeld et al. 2008). Baumeister et al. (2004) has also shown, in one study of social gossip content, that lessons about how to effectively live in one's society were more frequent than discussions of specific individuals. Similar results have been found in mathematical modelling approaches, where it has been shown that effective linguistic communication can solve the adaptive problem of individual learning, as social information transmission enables the accumulation of fitness-relevant knowledge (De Backer and Gurven, 2006).

However, an important limitation, of the aforementioned mathematical modelling approach, is that it relies on particular assumptions, including fully cooperative agents and conditions where social status trades for information (De Backer and Gurven, 2006). Unfortunately, there is currently little evidence to support either of these suppositions, in existing nonhuman primates, putative early humans, or modern humans (where cooperative conditions may be a given, but conditions where status trades for information are less supported) (Fitch, 2005). In addition, experimental research has shown humans are often insensitive to serious conversational incoherence, calling into question whether spontaneous human conversations should best be modelled in terms of faithful information transmission (Galantucci and Roberts, 2014). Moreover, each of the aforementioned studies stress the primary importance of gossip and social bonding, and propose cultural learning as a likely additional derived use of language function (Baumeister et al. 2004; De Backer and Gurven, 2006; Galantucci and Roberts, 2014).

Other recent studies have attempted more sophisticated approaches, to tease out derived applications from the more primary use of language function. Similar to a critical tests format (Calhim, Shi and Dunbar, 2006; van Schaik and Dunbar, 1990), the predictions made from different competing hypotheses were pitted against each other to determine a cognitive bias for one sort of information over another (Mesoudi, Whiten and Dunbar, 2006; Redhead and Dunbar, 2013). For example, Mesoudi, Whiten and Dunbar (2006) demonstrated that social information transmits with significantly greater quantity and accuracy than equivalent non-social information, such as technology or the physical environment, suggesting a bias in cultural information transmission toward social interactions. In a similar study, it was shown at least in terms of general interest and attention, memory recall was often more accurate for social narrative content than equivalent factual, ecologically related information (Redhead and Dunbar, 2013).

In addition, the apparent human cognitive bias for social information may have important functional implications in real-world situations. For instance, in a study of Maine lobster fisherman's radio exchanges, it was found that fisherman were significantly less likely to broadcast the location of lobster clusters to strangers and tourists in the community, as when compared to close family and friends (Palmer, 1991). In other words, it would appear that information exchange is typically something we engage in once we have already used language to establish a relationship in the first place. Accordingly, the results of these studies taken together would seem to suggest a primary role for language in social gossip and social cohesion contexts, with cultural transmission of fact-based ecologically relevant information as a possible derived communicative function (Millikan, 1984; Origgi and Sperber, 2000).

Finally, evidence from traditional societies indicates that among the !Kung Bushmen of the Kalahari desert, there is near absence of pedagogy, at least in the modern sense of the term (Bruner, 1972). For instance, anthropologists state, "Most of what we would call instruction is through showing…. Among the !Kung children there is very little 'telling'" (Bruner, 1972: 11), and more recent studies tend to confirm this finding (Rogoff, 1990). On the other hand, within the context of group foraging, it appears that hunter-gatherers do engage in extensive sharing of information, on the location of key resources, in advance of the actual search (Smith, 1991). However, the conditions under which extensive information-sharing occur, prior to foraging, are often highly variable and condition dependent (Smith, 1991). Both mathematical modelling and empirical evidence indicate that information-sharing occurs most often when prey are both spatiotemporally unpredictable and concentrated to a particular locale (Smith, 1991). In most cases, these correlated patterns of unpredictable resource distribution and information-sharing are often highly seasonal; for instance, Inujjuamiut foraging parties most often exchange information when caribou aggregate in sizable, but unpredictable, herds during the winter months (Smith, 1991: 336). In

sum, although information sharing does indeed seem to occur within traditional societies, prior to hunting, it would appear to be inconsistent and highly context-sensitive (Smith, 1991).

1.6 Mental tool

In contrast to the cultural learning hypothesis, other proposals have argued that human language may have evolved primarily for use as a mental tool, and only secondarily as a method of communication (Burling, 1993; Herwegen and Karmiloff-Smith, 2006). According to this hypothesis, language is not only a communicative tool, but more importantly, a domain of knowledge that provides a means for thinking about the world (Burling, 1993; Herwegen and Karmiloff-Smith, 2006).

1.6.1 Comparative animal behaviour

In support of this hypothesis, it is argued that language is not the most important trait which sets humans apart from other animal species; indeed, many animals have complex communication, intricate sensorimotor abilities, and rich mental representations (Herwegen and Karmiloff-Smith, 2006). However, the validity of this hypothesis has often been criticised on several grounds (Burling, 1993). First, research has shown the mentality of nonhuman primates, especially great apes, to be strikingly similar to that of human beings, including abilities such as spatial and numerical cognition, symbolic capacity, and complex social relationships (Adachi, 2014; Allen and Gardner, 1980; de Waal, 2007; Sakura, 1993). Therefore, the proposal that language was primarily selected in humans as a method of thinking about the world seems questionable, given the complex cognitive abilities present in closely related species without language. Second, mental activity, regardless of content, is

typically expressed in action or inaction, and is therefore most traditionally selected for its behavioural outcome (Parker, 1993). In other words, in contrast to the majority of language function proposals, the mental tool hypothesis offers no coherent selective scenario for the evolution of language (probably due, at least in part, to the absence of particular behaviours which could in principle be selected for clear survival or reproductive advantage). Third, if language evolved primarily for communication with the self, it is challenging to explain the existence of the complex human articulatory speech apparatus necessary for speaking, which often results in death by choking, and is clearly absent in closely related primate species (Parker, 1993; Pinker, 2000).

1.6.2 Developmental psychology

According to Karmiloff-Smith (1992), language is not merely a method of communication, but a "problem space" or "domain of knowledge" in its own right, allowing pre-linguistic infants to interpret and perceive the world in terms of language. For example, by as early as nine months, infants have begun to isolate the phonological sequences of their native language, allowing them to parse fluent speech into words (Herwegen and Karmiloff-Smith, 2006). Moreover, by 4 years of age, once children have become fluent speakers, they have already begun to learn the mappings between words and referents, as well as more complex aspects of morphology (Herwegen and Karmiloff-Smith, 2006). Clearly, while these examples demonstrate a clear developmental interplay between cognition and language, they nevertheless fail to provide convincing evidence that 1) "domains of knowledge" must always occur in connection with language, 2) language itself is a "domain of knowledge" which was primarily selected for this adaptive function, or 3) provide a satisfactory explanation for why language has any behavioural or verbal aspects at all (Parker, 1993).

As further evidence that language is not just a communicative tool, but a "problem space" in its own right, Herwegen and Karmiloff-Smith (2006) argue human populations exist in which linguistic communication is not impaired, but language as a "domain of knowledge" is impaired. For example, despite low IQs in the 60-point range, individuals with the neurodevelopmental disorder Williams syndrome nevertheless have proficient language skills (Donnai and Karmiloff-Smith, 2000). Additionally, among individuals with Down syndrome, impairments are often more severe, where both the ability to represent and organise knowledge is typically just as impaired as language proficiency skills (Karmiloff-Smith, 2006). Moreover, peer acceptance and normal socialization skills are often problematic among both populations, despite a general lack of reluctance to join in social conversations (Herwegen and Karmiloff-Smith, 2006).

In any event, while both examples point to an apparent deficiency in the "reorganization of knowledge" (Karmiloff-Smith, 2006), they fail to provide convincing evidence that language evolved specifically for this purpose; rather, they apparently draw attention to the rather commonplace observation that the two abilities can be considered relatively independent faculties. Moreover, it is equally important to note both these cases are examples of *clinical disorders*, which are not generally considered representative of the human universals of which evolutionary psychologists seek to explain. In sum, while it may be true that the ability for "reorganization of knowledge" may be unique to humans, it is still far from clear how or why language evolved specifically to enable this capacity.

1.6.3 Human social behaviour

Lastly, evaluating this hypothesis in terms of human social behaviour is perhaps the most problematic, as the proposal argues that language itself should be thought of as a mental tool, rather than as a method for gestural or vocal signalling (Burling, 1993; Herwegen and

Karmiloff-Smith, 2006). As noted above, the mental tool hypothesis, as currently formulated, offers no obvious selective scenario for the evolution of language, likely due to the absence of particular behaviours which could in principle be selected for clear survival or reproductive advantage. In other words, mental activity, regardless of content, is generally expressed in terms of action or inaction, and is therefore most traditionally selected for its behavioural outcome (Parker, 1993). In sum, no known evolutionary mechanism exists which could ignore behavioural outcomes, while selecting primarily for specific mental activities.

1.7 Deception

Impressed with the apparently ubiquity of deception found throughout the animal kingdom, some have argued that animal communication systems have evolved solely and primarily (as opposed to secondarily as a derived function) for the function of deception (Dawkins and Krebs, 1978; Krebs and Dawkins, 1984; Scott-Phillips, 2006). According to this hypothesis, the speaker (or signaller) may benefit from the breakdown of the normative relationship between speaker attribute or referent and the actual nature of the external world, while the listener (or receiver) may incur a significant cost from this misrepresentation (Searcy and Nowicki, 2005).

For instance, in a classic series of articles by Dawkins and Krebs (1978) and Krebs and Dawkins (1984), it was argued that animal signalling, and by extension, human language ought to be viewed as an evolutionary arms race in which signallers evolve to become better at manipulating receivers, while receivers evolve to become more resistant to manipulation. Moreover, at least one prominent evolutionary linguist has voiced a very similar assertion, in that human language evolved primarily and specifically for the function of mind-reading and manipulation of group conspecifics (Scott-Phillips, 2006). According to this hypothesis,

human language is viewed as a fundamentally competitive method of communication.

1.7.1 Comparative animal behaviour

The evolutionary strategy of deception, once thought unique to humans, has now been documented in a relatively small number of nonhuman primate species, including chimpanzees (Green, 2005), baboons (Byrne and Whiten, 1985), and capuchins (Wheeler, 2009) as well as birds (Bugnyarf and Kotrschal, 2002; Moller, 1990) and ungulates (Brø-Jorgensen and Pangle, 2010). Chimpanzees, in particular, have been observed using both nonverbal signals in the wild (Byrne and Whiten, 1991; deWaal, 1986; Kirkpatrick, 2007), as well as sign-language mimicry in trained experimental conditions (Green, 2005). However, there are even fewer cases of animal communication which might be described as the sort of manipulative *tactical deception*, often referred to as functional deception, which occurs in humans (Byrne and Whiten, 1990; Byrne and Whiten, 1991; Hauser, 1997; Talwar and Lee, 2008). Moreover, in highlighting the relatively paucity found throughout the animal kingdom, Byrne and Whiten (1990) define tactical deception, accordingly, as "acts from the normal repertoire of [an] agent, deployed such that another individual is likely to misinterpret what acts signify, to the advantage of the agent". As such, this sort of deception can prove potentially costly to the user, as tactical deception mostly occurs in social animals, which may lose trust of fellow group-members when their deceit is discovered (Cheney and Seyfarth, 1990; Hauser, 1992; Wheeler, 2009).

In the vast majority of documented cases, tactical deception most frequently occurs in cases where alarm calls or warning cries are used deceptively to obtain access to increased food supplies or reproductive opportunities (Brø-Jorgensen and Pangle, 2010; Bugnyarf and Kotrschal, 2002; Byrne and Whiten, 1985; Wheeler, 2009). For example, tufted capuchin subordinate monkeys have been found to use alarm calls (normally reserved for predator

sightings) to elicit responses from fellow group members, when competing with dominant monkeys over valuable food resources (in effect, taking advantage of their distraction to pilfer food) (Wheeler, 2009). However, because the costs of deception are believed to be relatively high for the deceiver if discovered, tactical deception has, so far, been documented in very few cases (Hauser, 1992; Wheeler, 2009). As such, it is believed to be more common in cases where the costs of ignoring the possibly deceptive act are even higher than the costs of believing (Hauser, 1992; Wheeler, 2009).

For example, observations of rhesus monkeys have found that on forty-five percent of occasions when monkeys serendipitously discover food, they announce their discoveries to the rest of the group (Hauser, 1992). Conversely, monkeys who discover food but fail to call, and are later detected with food by other group members, typically receive significantly more physical aggression than vocal discoverers (Hauser, 1992). In the case of tufted capuchins, the costs of ignoring the aforementioned alarm calls could potentially result in death, leading to more conservative courses of action (e.g. fleeing the scene even in the presence of food), even when the caller is known as a deceiver (Wheeler, 2009).

Deception and active signal falsification have further been found within animal communication to manipulate mates for the purpose of reproductive benefits (Brø-Jorgensen and Pangle, 2010). For example, male topi antelopes have been found to alarm snort deceptively in order to retain receptive females within the males' territory (Brø-Jorgensen and Pangle, 2010). In sum, deceptive alarm snorting would appear to be a secondary or derived function, rather than the primary function for animal communication, and beyond any further considerations, could reasonably be argued a derived function for human language (Millikan, 1984; Origgi and Sperber, 2000).

Therefore, while deception may accurately characterise certain types of animal communication in certain contexts, there are strong reasons to doubt that it can be identified

as the primary function for the majority of animal communication systems, including human language. Firstly, behavioural ecologists have argued on theoretical grounds that Dawkins and Krebs' (1978, 1984) signal evolution models (which argue that signallers evolve to manipulate receivers, while receivers evolve to resist manipulation) are in fact critically flawed. Searcy and Nowicki, for example, have argued that if there is, on average, no information of benefit to the receiver of a signal, then receivers should evolve to ignore the signal (then signalling no longer has any benefit to the signaller and the communication system collapses) (Searcy and Nowicki, 2005: 8).

Granted, a partial solution to this paradox was proposed by Zahavi (1975), but this was limited specifically to signals with respect to mate choice or aggressive contexts. However, while males might be inclined to deceive or manipulate females (or other male rivals) with respect to their genetic superiority, such males were actually 'handicapped', thereby ensuring the honesty of the signal (Searcy and Nowicki, 2005: 9). For instance, the female preference for the exaggerated tails of adult male birds, as found in some species, is often cited as a prototypical example; but not all bird species can be characterised accordingly (Searcy and Nowicki, 2005). Clearly, then such models apply only in a relatively narrow range of contexts, and it is furthermore unclear whether animal models of this kind can be generalised to human language (as there is currently no existing data to support or indicate they do apply to human language). In conclusion, functionally deceptive signalling is believed to be rare, as well as have a generally low cost for the deceived; otherwise, the signal would simply be ignored and become ineffectual (Fitch and Hauser, 2002; Searcy and Nowicki, 2005).

1.7.2 Developmental psychology

Further arguments for the primacy of human deception have come from developmental research (Keating and Heltman, 1994; Lee, 2013; Talwar and Lee, 2008). Experimental studies examining children's antisocial lies have found that self-serving lies (e.g. to conceal a transgression), typically occur between 4 and 7 years of age, with lies below this age range occurring much more infrequently, where the default seems to be honesty (Lee, 2013; Talwar and Lee, 2008). Moreover, while children seem, in most cases, more willing to lie to familiar others than strangers, self-interest seems to be a generally stronger motivator for deception (Talwar and Lee, 2008). In addition, several studies indicate that deception often occurs through nonverbal expressive control in a variety of deceptive situations (Keating and Heltman, 1994; Lee, 2013; Talwar and Lee, 2008).

On the other hand, while non-human primates have been observed using tactical deception for self-interest, prosocial deception appears to be unique to humans (Talwar and Lee, 2008). Interestingly, studies have also shown that older children beyond age 7 are generally more inclined to tell prosocial "white lies", despite potential costs to themselves (Lee, 2013; Talwar and Lee, 2008). Accordingly, as children grow older, they become more likely to be concerned with the feelings and needs of others, and therefore more likely to tell prosocial white lies, as well as prosocial "blue lies" lies (i.e. lies intended to benefit an in-group collective) (Lee, 2013; Talwar and Lee, 2008). Children are, therefore, generally more inclined to tell lies for self-interest, but will often conform to social norms and etiquette in telling prosocial lies, in order to maintain social relationships, when costs to self-interest are relatively low (Lee, 2013; Talwar and Lee, 2008). In sum, it would appear plausible that antisocial deception, as well as prosocial white and blue lies, could serve adaptive functions in children (Talwar and Lee, 2008).

However, some authors have questioned this interpretation for several reasons (Talwar and Lee, 2008). First, it may be unreasonable to assume, at least for antisocial deception, whether lying is especially well-suited to serve an adaptive function in children, given their general lack of physical strength and social power (Talwar and Lee, 2008), although some evidence may suggest the existence of viable nonverbal cues (Keating and Heltman, 1994). Second, deception is often highly context-specific; in some situations, lying may be appropriate, but in others honesty may be a more ideal strategy (Lee, 2013; Talwar and Lee, 2008). Research suggests this is not a trivial decision, as it requires adequate knowledge of various social norms, as well as the child's intuitive understanding of how to represent the mental states of themselves and other people (Lee, 2013). No current evidence exists, however, as to whether children are capable of making these sorts of cost-benefit analyses prior to the decision to deceive (Talwar and Lee, 2008). Third, with respect to potential adaptive functions for language, considerable evidence shows deception often occurs through nonverbal expressive control in a variety of deceptive situations, demonstrating language is not always a necessary requirement (Keating and Heltman, 1994; Lee, 2013; Talwar and Lee, 2008).

Nevertheless, some have suggested a harsh disciplinarian approach to parenting may be directly related to the development of deception in children (e.g. in order to conceal transgressions), as an adaptive response to avoiding severe punishment (Talwar and Lee, 2008). However, given lying typically occurs in many varieties, including antisocial and prosocial lying (e.g. white and blue lies) with prosocial lying occurring later in development, it seems a social bonding function could be just as legitimately posited as an antisocial function (Talwar and Lee, 2008). Indeed, several studies have shown an increase in antisocial lying in children and adolescents with significant behavioural problems, suggesting a lack of normative development in these cases (Lee, 2013; Li, Kelley, Evans and Lee, 2011).

1.7.3 Human social behaviour

As previously noted, analyses of human conversational content reveal social relationship topics overwhelming dominate over two-thirds of total everyday human conversation time (Bischoping, 1993; Dunbar et al. 1997; Haviland, 1991). Indeed, as previously argued, increasing evidence suggests language evolved primarily for creating and maintaining social relationships (Cohen, 2012; Dunbar, 2003, 2004; Freeberg, Dunbar and Ord, 2012; Galantucci and Roberts, 2014; Ireland et al. 2011; Nettle and Dunbar, 1997; Pietraszewski and Schwartz, 2014a; Roberts, 2010; Weaver and Bosson, 2011). Moreover, as in any biological system that has significant time and energetic costs as well as adaptive benefits (Barrett, Dunbar and Lycett, 2002), it certainly begs the question as to why we spend most of our time in the prosocial use of language, if language reputedly evolved primarily for manipulation and deception of group conspecifics.

Second, studies have further shown that, in humans, lying and deception are relatively rare occurrences in the general population (DePaulo et al. 1996; Hancock, Thom-Santelli and Ritchie, 2004; Serota, Levine and Boster, 2010). For example, Serota, Levine and Boster (2010) found, in a sample of 1,000 Americans, the average number of lies told per day was 1.65. Moreover, the distribution was highly skewed; 22.7% of all lies were told by one percent of the sample, and 50% of all lies were told by five percent of the sample. Studies conducted in controlled laboratory environments have obtained similar results (Abeler, Becker and Falk, 2012). Given so many clear incentives to deceive, often in the interest of manipulating rivals (Hauser, 1992; Dawkins and Krebs, 1978; Whiten and Byrne, 1988), many investigating the evolution of language have questioned how linguistic communication can remain primarily honest in the everyday use of language (Fitch, 2010; Knight, 1998; Nettle, 1999). One possible explanation, as previously mentioned, is the risk of unfavourable

social reputation and implementation of strict social sanctions against lying and deception (Bliege Bird and Smith, 2005; Donath, 2008; Lachmann et al. 2001).

Furthermore, previous findings have also revealed that individuals seem to be much more willing to deceive when the potential stakes are quite high, as for example when attempting to advertise to prospective mates, suggesting a context-sensitive signalling device (Ellison, Heino and Gibbs, 2006; Hall, Park, Song and Cody, 2010; Toma, Hancock and Ellison, 2008). For instance, one study of online daters in New York City found that 87% of men and 76% of women lied about at least one attribute in their personal profiles; men tended to increase their height, whereas women tended to lower their weight (Toma, Hancock and Ellison, 2008). Deviations from truth were generally quite small in magnitude, however, suggesting that individuals were at least partially aware of the possible social repercussions from being discovered as deceitful (Toma, Hancock and Ellison, 2008).

Finally, computational modelling experiments of human social networks demonstrate deception can be just as often used for beneficial prosocial purposes, as often as for antisocial uses (Iñiguez et al. 2014; Rauwolf, Mitchell and Bryson, 2014). For instance, recent studies have shown antisocial lying causes social networks to become increasingly fragmented, while prosocial lies can be beneficial in smoothing the flow of interactions and facilitating larger, more integrated networks (Iñiguez et al. 2014; Rauwolf, Mitchell and Bryson, 2014).

1.8 Discussion

Despite a steadily increasing amount of evidence that language evolved primary for facilitating social bonding and group cohesion in large social groups (Cohen, 2012; Dunbar, 2003, 2004; Freeberg, Dunbar and Ord, 2012; Galantucci and Roberts, 2014; Ireland et al. 2011; Mesoudi, Whiten and Dunbar, 2006; Nettle and Dunbar, 1997; Pietraszewski and

Schwartz, 2014a; Pietraszewski and Schwartz, 2014b; Redhead and Dunbar, 2013; Roberts, 2010, 2013; Weaver and Bosson, 2011), there is currently no consensus as to how language originated in the human species (Christiansen and Kirby, 2003; Larson, Déprez and Yamakido, 2010; Számadó and Szathmáry, 2006).

Perhaps most problematically, nearly every study to have considered the function(s) of language has simply presented evidence in favour of its preferred hypothesis. Fewer attempts have been made to test between competing hypotheses. Indeed, many proposals often fail to recognise that there may in fact be competing hypotheses. Confirmatory evidence is always, at best, weak evidence for a hypothesis, not least because it is quite common for evidence to be compatible with several mutually incompatible hypotheses. A comparison between competing hypotheses thus provides a heuristically more powerful way of testing a particular hypothesis. Only then can one definitively conclude that the best supported hypothesis is in fact the most likely explanation. Accordingly, a comprehensive review of the contemporary literature suggests that there are probably three serious contenders for the adaptive significance of human language: the interdependence or cooperation hypothesis (Tomasello et al. 2012), sexual selection or honest signalling theory (Bliege Bird and Smith, 2005; Lachmann et al. 2001) and the social bonding hypothesis (Dunbar, 2004; Freeberg, Dunbar and Ord, 2012).

In this thesis, I therefore take a multi-pronged approach to testing these three different hypotheses for the biological function of language, with clear implications for the origin and subsequent evolution of language. In Chapter II, I take an experimental approach to the problem of group social bonding and effective communication with increasing group size: the interdependence hypothesis. In pursuit of this experiment, I employed a cooperative foraging task using small and large group sizes with varying degrees of group familiarity (as a measure of group cohesion) to determine what factors enable individuals to make sense of

information from others and ultimately converge upon the truth. The first prediction was that language would be used more frequently in large group coordination than in small group coordination. The second prediction was that performance would increase in larger groups as a result of more independent sources of information. The third prediction was that performance would be greater in groups with strong social ties than groups with weaker social ties (i.e. groups of friends versus groups of strangers). In summary, the experiment in Chapter II tested several proposals predicted by the interdependence hypothesis (and by extension the social bonding hypothesis), for communication with respect to human language.

In Chapter III, I take an experimental approach to the question of whether human language can be characterised in terms of honest signalling theory. In this respect, I test several proposals predicted by the sexual selection hypothesis and deception hypothesis for human language function. Experiment I tests the prediction that low-frequency vocabulary evolved as an honest signal of human intelligence (as a receiver-independent cost), and should therefore be attractive to prospective mates. Experiment II tests a subtle modification of this proposal, that use of fluent poetic vocabulary evolved as an honest signal of human intelligence (as a receiver-independent cost), and should therefore be attractive to prospective mates. Experiment III tests the prediction that public speaking may constitute an honest signal of intelligence, status, and influence (as a receiver-dependent cost), with concurrent potential costs of unfavorable social reputation. Experiment IV tests the prediction that self-disclosure of revealing or culpable personal information, as a kind of social risk-taking (as a receiver-dependent cost), may be a form of social advertisement while signalling imperviousness to danger and social repercussions. More specifically, Experiment IV tests the prediction that prosocial courtship deception leads to increased attractiveness ratings, while antisocial courtship deception leads to decreased attractiveness ratings (as a direct test

of both increased social status and social sanctions for deception within the context of mate attraction). In summary, Chapter III tested four competing, but not necessarily mutually exclusive, predictions characterising human language in terms of honest signalling theory.

In Chapter IV, I take an experimental approach to the question of whether recursive syntax evolved to be a consistent feature of human language. Interestingly, several primatologists and evolutionary psychologists have theorised the human capacity for higher-order intentionality may have played a critical role in the evolution of language, including especially the ability for recursive syntax comprehension and production (Cheney and Seyfarth, 2007; Dunbar, 1998; Dunbar, 2009). However, no empirical studies to date have yet examined how higher-order intentionality capacities beyond second-order may influence the acquisition of recursive syntax comprehension, in either normal adults or children. In pursuit of this experiment, I utilized the Imposing Memory Task (IMT) and a recursive syntax task to determine relative performance on each of these cognitive measures. The first prediction was that intentionality performance and recursive syntax performance should be highly correlated as evidence of a cognitive bootstrapping effect. The second prediction was that simple fact-based social memory tasks should be more efficient and cognitively less demanding, than either equivalent fact-based mentalising or recursive syntax problem-solving. The third prediction was that recursive syntax task problem-solving should be more efficient and cognitively less demanding, than equivalent intentionality task problem-solving. In summary, the experiment in Chapter IV tests several predictions of the cognitive bootstrapping hypothesis for the evolution of recursive syntax with respect to human language.

Finally, Chapter V provides an overview of the thesis and assesses the broader implications of the findings, the methodological challenges and limitations, future directions for research and final conclusions of this integrated study.

Chapter II

Human group performance in social foraging: An empirical test of the interdependence hypothesis

2.1 Introduction

Language is arguably one of the most salient features that distinguish humans from other animal species (Deacon, 1997; Fitch, 2010; Radick, 2007). Although many species have evolved the capacity to communicate information about aspects of their social and ecological environments (Bradbury and Vehrencamp, 2011; Cheney and Seyfarth, 1990; Janik et al. 2006; Searcy and Nowicki, 2005), no other species apart from humans are able to transmit abstract ideas, complex information, and unique perspectives through language, radically influencing others with respect to a complex cultural environment (Barrett, Dunbar and Lycett, 2002). Moreover, humans form complex social networks, both in modern industrial societies (Jackson, 2010) as well as traditional hunter-gatherer cultures (Apicella et al. 2012), in which enormous amounts of information are exchanged among highly inter-connected individuals in the social network (Easley and Kleinberg, 2010).

Several biological functions for human language have been posited, especially in social network contexts, including cultural learning (Baumeister et al. 2004; De Backer and Gurven, 2006), social cohesion (Dunbar, 1996, 2004), cooperation (Tomasello et al. 2012) and the enforcement of social norms (Kniffin and Wilson, 2005; Wilson et al. 2000). Perhaps most fundamentally, however, language can provide communicators with information about the surrounding environment, especially information about other individuals in the social network whom communicators may interact with at some point in the future (Craik, 2009; Laidre et al. 2012). Experimental studies of human gossip have shown that individuals

integrate information from multiple sources, comparing and contrasting various pieces of information, to distinguish true from false messages and ultimately converge upon the truth (Collins et al. 2011; Hess and Hagen, 2006; Sommerfield et al. 2008).

It has furthermore been argued that such cognitive heuristics may be an innate and universal aspect of human cognition and behaviour that evolved in response to the demands of large social groups (Dunbar, 1996, 2003, 2004; Hess and Hagen, 2006; Wilson et al. 2000). For instance, some theorists have argued that as human groups increased in social complexity over the course of human evolution, language may have evolved for the purpose of mutualistic cooperation with increasing group size (Tomasello et al. 2012; Tomasello, 2008). Cooperation, in this sense, is defined as interdependent collaboration with a shared goal, such as group foraging or hunting, as opposed to general altruism, reciprocity or social cohesion (Tomasello et al. 2012). According to this proposal, known as the interdependence hypothesis, language and cooperation coevolved together in a three stage process: 1) small group cooperation and communication began as the starting point, mainly in the context of cooperative foraging and 2) as group size increased, over the course of human evolution, large group cooperation and communication followed as the entire community needed to work together, often in the context of between-group competition, leading to culture, cultural norms, and complex institutions. As a third and final step: 3) human cognition and social behaviour became ever more cooperative, prosocial and altruistic as humans became ever more interdependent (Tomasello et al. 2012).

Consequently, there may well have been a need among early humans (and presumably extant hunter-gatherer societies) for cooperative foraging and hunting that would have required sophisticated communication (Bickerton and Szathmáry, 2011; Tomasello et al. 2012). Several authors, for instance, note that eusocial insects communicate displacement (i.e. reference to events remote in time or space), using signals which indicate the distance,

direction and richness of food resources to the rest of the group (Bickerton and Szathmáry, 2011; Számadó, 2010). According to this argument, among early humans that had located a potential food resource, some form of communication would have been necessary to inform the rest of the group of the find (Bickerton and Szathmáry, 2011; Számadó, 2010).

However, very few studies to date have experimentally tested whether communication facilitates cooperation in humans, thereby limiting the credibility of these theoretical arguments with respect to human language. Nevertheless, such experimental tests are perfectly feasible in many animals, including humans (Freeberg, Dunbar and Ord, 2012; King and Sueur, 2011). For example, social coordination games have been argued to be an effective method for understanding the social behaviour of groups of individuals in a variety of contexts and species, including flocks of birds, shoals of fish, humans and other non-human primates (King and Sueur, 2011).

In this study, I therefore adapted the paradigm developed by King et al. (2011), which used a cooperative social foraging task with small and large group sizes of varying familiarity (as a measure of group cohesion) to determine what factors enable individuals to use information from others to converge upon a group consensus. Data were collected on both gestural and linguistic communication to determine what effect this had on overall group performance, defined in terms of faster group consensus time and greater percentage of 'good' tokens collected. Previous research has had success with this methodology, demonstrating communicating groups forage more effectively than non-communicating groups (King et al. 2011).

As foraging was humanity's first and most successful adaptation (Lee, 2005), the social foraging model ensured the highest possible degree of ecological validity. In addition, among traditional hunter-gather societies, foraging has generally been found to be more commonly practiced among women than men (Bird, 1999; Marlowe, 2007; Marlowe, 2010;

Zihlman and Tanner, 2001). Moreover, in accordance with expectation states theory, men will often have an advantage over women in mixed sex-groups, other things being equal (Wittenbaum, 1998). Participants were thus restricted to females over age sixteen.

The current study examined three specific predictions following from the interdependence hypothesis. Previous studies have shown, according to the wisdom of the crowd effect, that many opinions are often better informed than fewer, provided they are made up of diverse opinions and ideologies (James, 2004; Vul and Pashler, 2008). Therefore, the first prediction was that performance would increase in larger groups as a result of more independent sources of information. As the wisdom of the crowd effect suggests, complex communication should possess a high number of distinct bits of information (James, 2004; Vul and Pashler, 2008). The second prediction was that performance would be greater in groups with strong social ties than groups with weaker social ties (i.e. groups of friends versus groups of strangers). The third prediction was that language and gesture would be used more frequently in large group coordination than in small group coordination.

2.2 Methods

2.2.1 Participants

Participants were 242 anonymous and voluntary female prospective and current university students from a range of backgrounds from the United Kingdom (56.20 %), North America (18.60 %), China (8.70 %), India (4.50 %), Asia (2.50 %), Trinidad (2.50 %), Brazil (0.40 %), Pakistan (0.40 %), Serbia (0.40 %), Italy (0.40%) and several other European countries. High school education was completed by 97.9 % of participants, with 24.8 % having attained a diploma or some college experience, and 42.1 % having completed a bachelor's degree or higher. All participants were fluent English speakers. To avoid

developmental effects and confounds due to declining social engagement in old age, participants were restricted to the age range 16–35 years (M = 19.78, SD = 3.76).

2.2.2 Ethics permission

Ethics approval was provided by the University of Oxford ethics board (CUREC). All participants provided written informed consent before participation, which included permission to record their behaviour using a digital camera. Subjects were paid £10 for participation.

2.2.3 Materials

Participant behaviour was observed in a circular arena (diameter 6.4m) containing six perimeter *foraging patches* (flip-door plastic waste containers) arranged radially and equidistant at 3.2m from a central *home base* (two plastic waste containers with slits for depositing tokens) equidistant from one another at 3.4 m apart (King et al. 2011). Foraging patches each contained a mix of 500 ‘good’ (red) and ‘bad’ (yellow) tokens, and foraging patches varied in quality from 5, 35, 50, 65, to 95 percent red tokens (Figure 2.1).

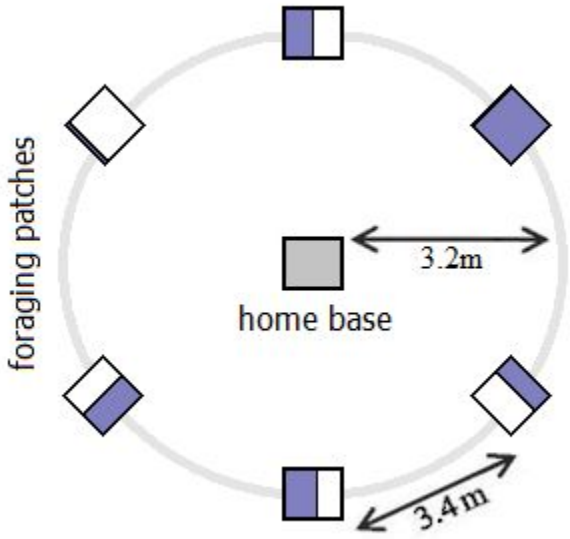

Figure 2.1. Overhead view of the experimental foraging arena. The six foraging patch containers (depicted here as squares and diamonds) were positioned 3.2 m from the central home base and 3.4 m apart from every other container. The shaded portions of the diagram indicate 'good' forage while the unshaded portions indicate 'bad' forage.

The experiments were originally taped using MP3 digital voice recorders. Individually recorded conversations proved difficult to analyse, however, due to a combination of background noise and the limited range of the directional microphones (often only picking up the speaker directly facing the microphone). Individual recording was therefore abandoned in favour of group auditory and video recording using a Sony HDR-CX190E Handycam.

2.2.3.1 Questionnaire measures

To determine what group processes and relationship statuses facilitated performance within the group, participants were asked to complete a post-experiment questionnaire comprising 25 seven-point Likert scale questions to examine group processes and group relationships (Table 2.1). The questions covered seven basic areas of intergroup processes and relationship status including: information sharing, morale building, planning, monitoring, relationship quality, and cooperation which have been found to mediate the effect of communication on group performance (Jenh and Shah, 1997; Mesmer-Magnus and

DeChurch, 2009). For example, participants were asked to rate how well they agreed (i.e. 'very little' – 1 or 'a lot' – 7): 'To what extent were you cooperative with other group members?', 'How much did you gesture to other group members?' and 'How much did you talk to other group members?' Participants were also asked to provide information about their age, nationality, occupation and highest level of attained education. The complete questionnaire as presented to participants can be found in Appendix A.2.

2.2.4 Procedure

Following King et al. (2011), a whistle was blown to begin and participants were tasked with collecting as many red tokens as possible from six radial foraging patches and to return them to a central home base (Figure 2.1). Immediately before the beginning of each trial, each group was verbally given the following standard set of instructions:

> Imagine yourself as a traditional hunter-gather collecting forage for survival. Each foraging patch contains both red and yellow tokens, but you should attempt to collect as many red tokens as possible, as you will be penalised for any yellow tokens. However, whatever token you remove cannot be put back inside the container and must be deposited at the home base. You are allowed to visit any of the foraging patches, as many times as you want, but you are not allowed to look inside any of the containers. You are only allowed to remove one token at a time. Any rule violations will result in a whistle being blown and 10 red tokens deducted from the end total. You are expected and encouraged to cooperate and work together as a team, and will be competing against other groups. So, try to do your absolute best on the task. Try to work as fast and efficiently as possible, since you will be timed; however, I can't tell you precisely how long you will have.

Participants had no restrictions on which patches they visited and were told at the outset, before receiving the standard set of instructions, that they would be "taking part in a short experiment", so as not to divulge the actual duration of allotted time for each trial. This ambiguous stipulation was included so as to discourage frantic and potentially dangerous task completion as well as to more accurately simulate real-world human foraging. In fact, all trials lasted 10 minutes.

Participants were instructed that if any task rules were broken (e.g. looking inside the containers, carrying multiple tokens, moving the containers, coaching from nonparticipant friends, etc.) a whistle would be blown and 10 red tokens deducted from the group's total. In practice, it was very difficult both to detect as well as police serious rule violations during the experiment. Nonetheless, serious rule violations were identified during subsequent detailed analyses of the videos, in six of the 35 trials. Accordingly, these were excluded from the analyses, yielding a final sample of N=29, distributed roughly evenly across the two group size conditions.

2.2.5 Experimental design

Participants were assigned at random to groups of two different sizes. Group sizes were chosen based on previous studies which have shown that information often has a restricted range of direct transmission, typically no more than 4 to 5 individuals at a time (Dunbar et al. 1995; King and Sueur, 2011; Waller et al. 2011). I therefore defined two sizes of group: small groups, consisting of 3 subjects, and large groups, consisting of between 6-20 subjects (M = 14.07, SD = 4.67). Large groups varied in size because it was not always possible to get everyone to turn up at the same time, even when they had been asked to come at a set time. In total, 29 groups of participants were observed (14 large groups and 15 small groups). All individuals were naïve in regard to the purpose of the experiment.

Only one experiment took place at any one time, and each individual participated in only one experiment. In addition, by randomly assigning subjects to a group, I tried to ensure that subjects were not allocated to groups in which they knew everyone. To check for this, self-report measures of familiarity were taken using a post-experiment questionnaire.

2.2.5.1 Dependent variables

Foraging patches each contained a mix of 500 'good' (red) and 'bad' (yellow) tokens. Therefore, group performance was measured according to two dependent variables in the study: percentage of 'good' (red) tokens and mean group consensus time. The percentage of good tokens was calculated according the number of red tokens relative to the number of yellow tokens the group collected and brought back to the home base by the end of the trial. Accordingly, as performance was rated based on the percentage of red tokens, and not on the absolute number of tokens collected, the group was therefore penalised by accumulating any yellow tokens.

I also used consensus time as a second measure of foraging success in that it measured how long it took a group to converge on the most preferred solution. Group consensus time was defined as the moment during the trial when at least fifty percent of the group had converged upon a single foraging patch for a duration of at least one minute (Figure 2.2). In order to ensure precise annotation, group consensus time was identified during detailed video analyses following the experiment.

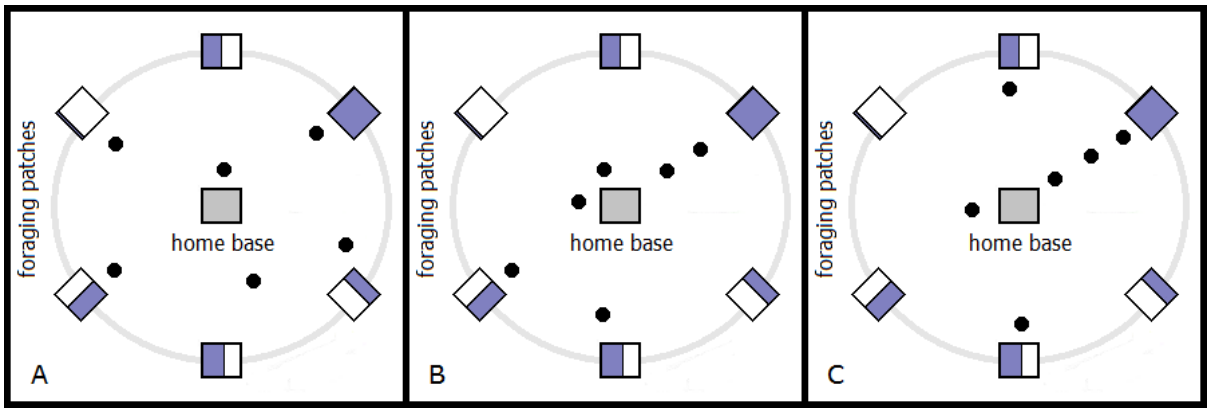

Figure 2.2. Time-lapse overhead view of the experimental foraging arena, immediately after the whistle was blown, at the: A) first minute B) second minute and C) third minute interval of the experimental trial. As shown in section C, this particular group consisted of six participants (each depicted as a dark filled circle), of which three participants, or 50% of the group, had achieved consensus by the third minute interval. Consensus is defined as foraging at the same patch. The shaded portions of the diagram indicate 'good' forage, while the unshaded portions indicate 'bad' forage.

To assess the amount of communication, the videos of each trial were examined and the number of subjects who were either speaking or gesturing scored separately as an instantaneous scan sample taken at 30-second intervals. The communication categories were delineated a priori so as to reflect language and gesture as relevant categories. For instance, verbal utterances were only counted if it was clear that the speaker was directing their utterance at one or more other individuals. Verbal utterances not obviously directed at anyone in particular, such as laughter, were not counted. Similarly, for gestures to count, they had to be clearly related to the task at hand, such as pointing at a particular foraging patch while looking at a particular individual. Gestures such as scratching the head or shrugging the shoulders, without being associated with attention to another individual, were not counted. All communication analyses were conducted via half-speed video playback to ensure all participants were accurately annotated. The study used a 2 (small/large group size) between-subjects x 2 (language/gesture communication condition) within-subjects design.

2.2.6 Statistical analysis

In many cases, plots of quantitative relationships between one of the dependent variables and actual group size are clearly nonlinear, with many having shapes suggestive of asymptotic relationships. SPSS only allows a limited number of standard functions to be fitted to a distribution, which makes it difficult to identify the best fit model it is it not one of these basic forms. Nonetheless, to try to capture this, I fitted a variety of nonlinear functions to the data, using the value of r^2 to identify the simplest model with the best fit.

2.3 Results

2.3.1 Group performance effects

Group performance was analysed to determine if large social foraging groups performed more effectively than small social foraging groups. Group performance was measured according to two dependent variables: percentage of 'good' tokens collected and mean group consensus time, with group size as the independent variable. The data for this section were analysed using non-parametric statistics, but regression analyses were also used to try and gain deeper insights into the possible quantitative patterns in the data.

Overall, the time participants in large groups took to achieve consensus was approximately three minutes (M = 196.29 seconds), while the majority of small groups rarely achieved consensus at any point during the ten minute trial (M = 558.67 seconds) (Figure 2.3). The difference is significant: Mann-Whitney test, $p < 0.001$. In addition, the difference between the percentage of 'good' tokens collected in large groups (M = 71.03 %) and small groups (M = 63.70 %) was also significant: $p = 0.026$ (Figure 2.4).

The relationship between percentage of 'good' tokens and actual group size is strongly suggestive of a nonlinear asymptotic relationship (Figure 2.5). The best fit was yielded by a cubic function, $r^2 = 0.177$, providing some principled evidence for suggesting that the relationship really is asymptotic. The relationship between group mean consensus time and actual group size yielded a power function best fit with a considerably higher goodness of fit, $r^2 = 0.590$ (Figure 2.6), again indicative of an asymptotic relationship.

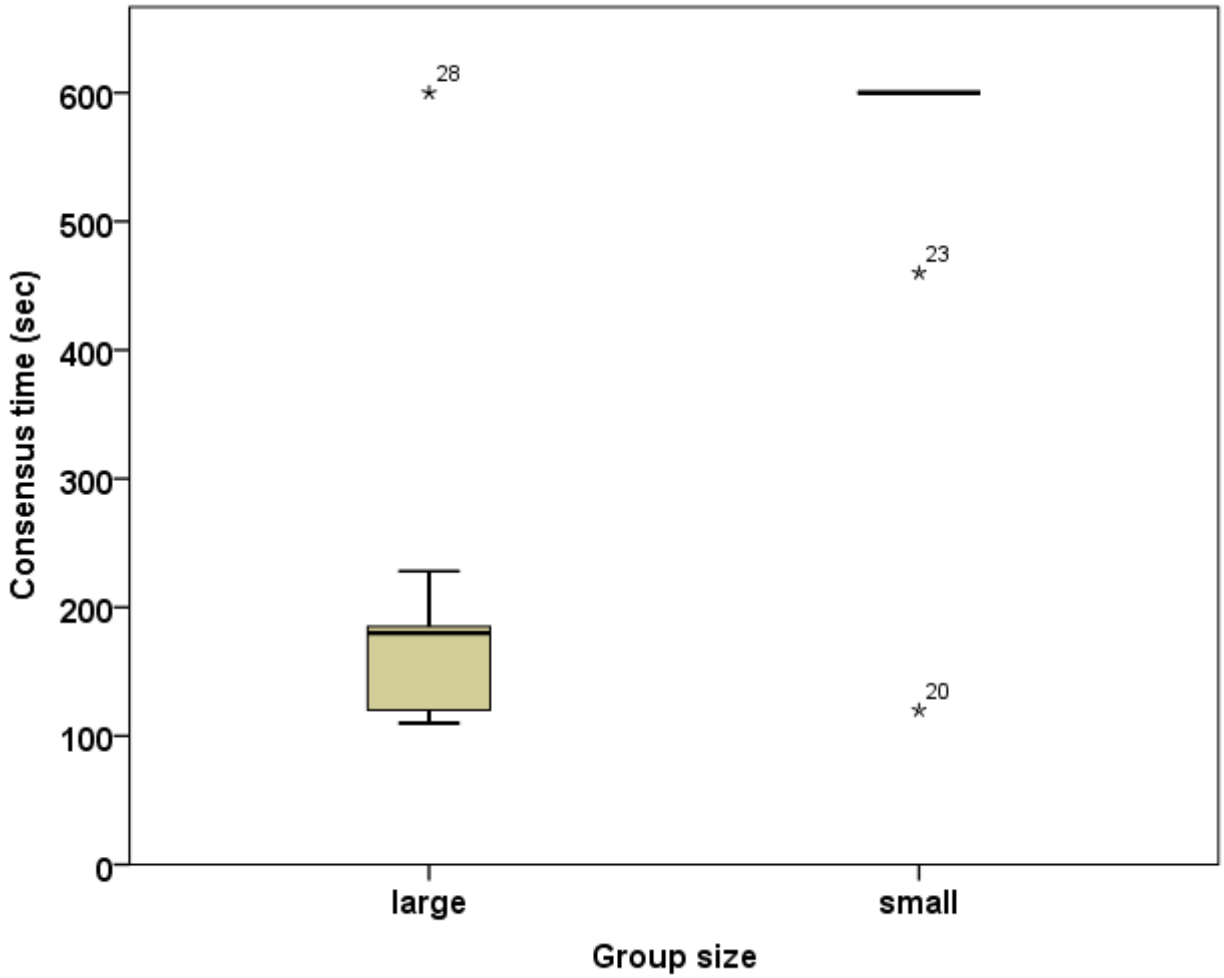

Figure 2.3. Group consensus time (i.e. point at which unanimous agreement was reached on the single best foraging patch) for small and large group sizes. Whisker ends represent the 95% range of all values of the data.

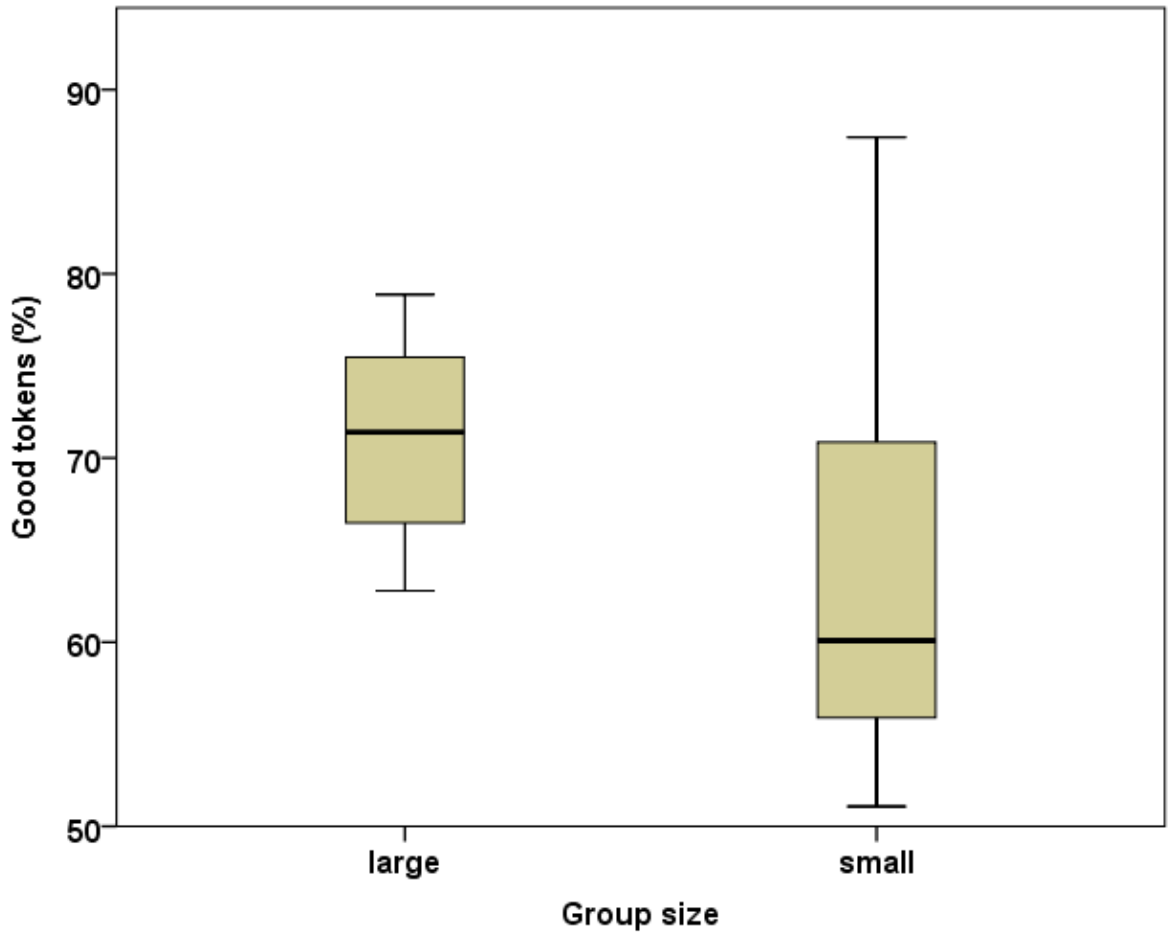

Figure 2.4. Group performance measured as the ratio of 'good' (red) to 'bad' (yellow) tokens for small and large group sizes. Whisker ends represent the 95% range of all values of the data.

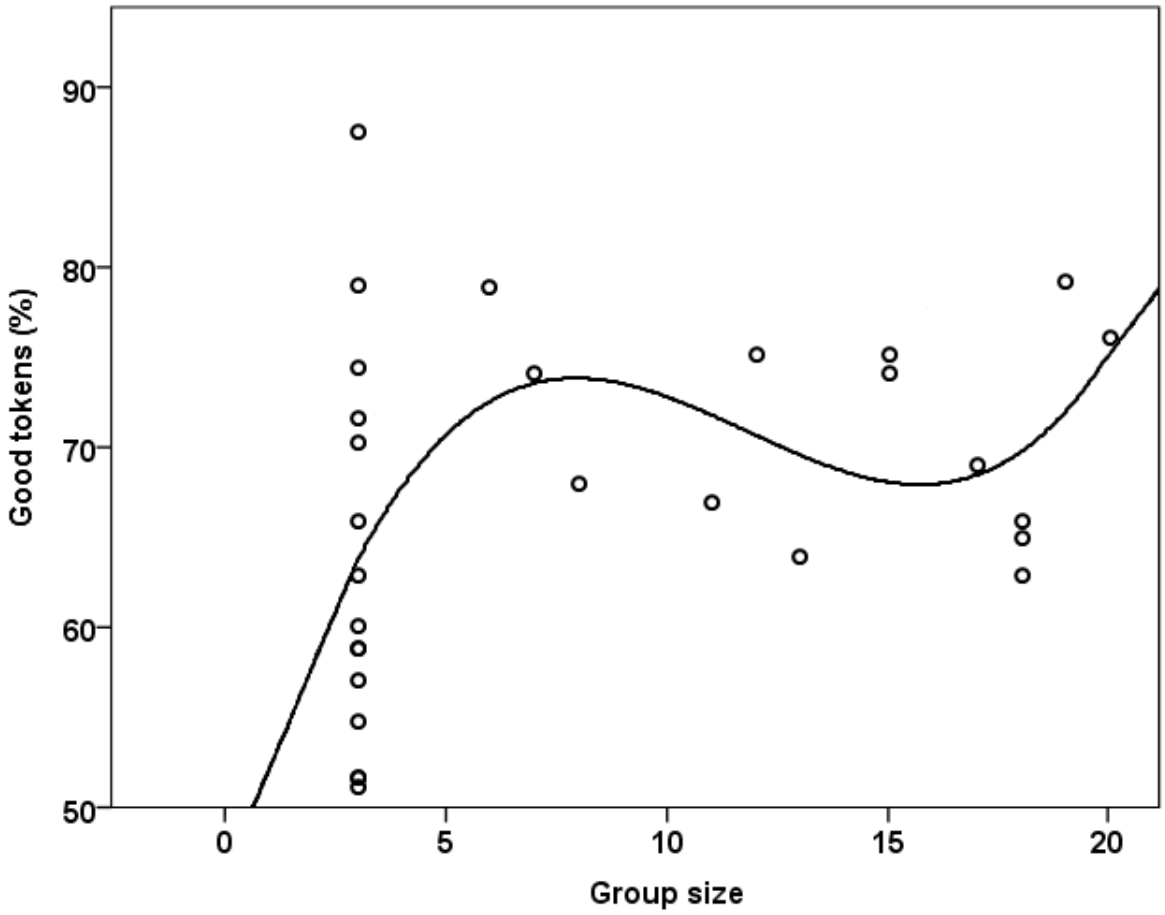

Figure 2.5. The relationship between percentage of 'good' (red) tokens and group size. The equation for the best cubic function is $y = 0.025^3 – 0.087x^2 + 9.15x + 43.54$.

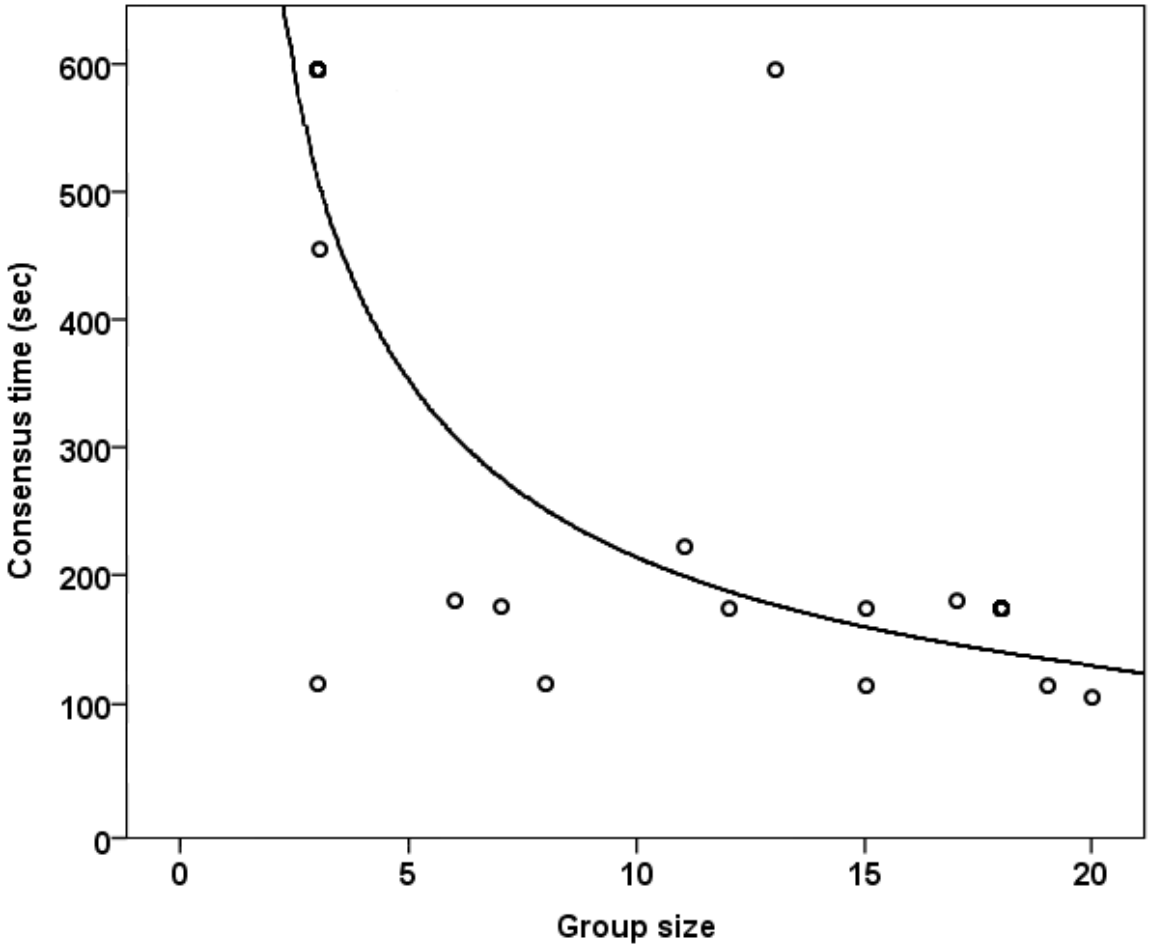

Figure 2.6. The relationship between consensus time (seconds) and group size. The equation for the best fit power function is $y = 1089x^{-0.697}$.

2.3.2 Self-report questionnaire items

Self-report measures of group processes and group familiarity were analysed to determine if they predicted social foraging group performance. Group performance was measured according to the two dependent variables: percentage of 'good' tokens collected and mean group consensus time.

Pairwise correlation analyses of both mean group consensus time and percentage of 'good' tokens collected as dependent variables, and the twenty-five questionnaire items as independent variables, were used to determine the influence of self-rated group processes and relationships on both performance criteria. Each of the questionnaire items was averaged for each group separately, to give an average group score, and these were then correlated against

the two dependent variables (percentage of 'good' tokens and consensus time) (Table 2.1). Significant effects were found between percentage of 'good' tokens collected and cooperation received, as well as between percentage of 'good' tokens collected and the request for verbal information. Similarly, significant effects were found between consensus time and how well the participant knew each person in the group, whether the participant knew the names of other people in the group, and how close the participant felt toward each person in the group.

Notably, the positive correlations for cooperation received and verbal information requests would seem to provide some rather weak evidence in support of the interdependence hypothesis. However, considered in tandem with the positive correlations also found for group acquaintanceship, nominal familiarity and social closeness might also suggest slightly better evidence in support of the social bonding hypothesis. Indeed, given that Bonferroni corrections are often regarded as a rather stringent test for statistical significance, it is perhaps worth mentioning that the remaining strongest correlations approaching significance include additional measures of social bonding; namely, between consensus time and how friendly participants felt toward other people in the group.

Table 2.1. Mean group size values and pairwise correlation values between percentage of 'good' tokens and consensus time for each self-report questionnaire item (significant relationships in **bold**)

Item	Item-group*	Rating-large group *M* (SE)	Rating-small group *M* (SE)	Percentage good tokens		Consensus time		Content
				r	p**	r	p**	
q1	C	5.07 (.137)	4.62 (.436)	.438	.017	-.222	.247	How cooperative and helpful were you to your group members?
q2	C	5.31 (.205)	4.53 (.452)	**.611**	**.001**	-.381	.042	How much cooperative and helpful behaviour did the group provide?
q3	L	3.29 (.114)	3.24 (.373)	.335	.075	-.037	.849	How much encouragement did you provide?
q4	L	3.84 (.202)	3.62 (.359)	.287	.131	-.086	.657	How much encouragement did the group provide?
q5	L	4.39 (.233)	4.52 (.378)	.265	.165	-.058	.764	How much did you talk about the task with the group?
q6	L	4.15 (.180)	3.88 (.362)	.427	.021	-.249	.193	How much suggestion or opinion did you provide to the group?
q7	L	3.46 (.183)	3.44 (.345)	**.577**	**.001**	-.225	.241	How much did you ask for opinions or suggestions from the group?
q8	G	3.03 (.170)	2.93 (.375)	.411	.027	-.162	.402	How much do you think you gestured (e.g. hand-waving, pointing, etc.) to the group?
q9	G	2.98 (.167)	2.77 (.334)	.394	.034	-.162	.400	How much did you see others gesturing (e.g. hand-waving, pointing, etc.) to you?
q10	P	3.40 (.224)	3.33 (.359)	.332	.078	-.067	.728	How much did you help plan and organise task duties for the group?
q11	P	4.61 (.292)	3.50 (.362)	.279	.142	-.397	.033	How much did someone else help plan and organise task duties for the group?
q12	D	3.15 (.160)	3.80 (.325)	-.154	.426	.214	.265	How confused were you about how best to complete the task?
q13	D	3.45 (.213)	3.29 (.310)	-.053	.785	-.028	.885	How much do you believe the group was confused about how best to complete the task?
q14	D	2.32 (.230)	1.87 (.270)	.027	.888	-.117	.546	How much did you disagree with the group about how to complete the task?
q15	D	2.24 (.217)	1.76 (.255)	.094	.628	-.128	.508	How much did your group disagree about how to complete the task?
q16	F	1.84 (.133)	5.60 (.396)	-.367	.050	**.642**	**.001**	How well did you know each person in the group?
q17	F	4.24 (.161)	5.91 (.311)	-.183	.341	.409	.028	How much did you trust other people in the group?
q18	F	4.97 (.137)	6.29 (.178)	-.240	.210	.503	.005	How friendly did you feel toward other people in the group?
q19	F	3.39 (.138)	5.91 (.269)	-.252	.188	**.579**	**.001**	How close did you feel toward other people in the group?
q20	F	4.55 (.140)	5.67 (.272)	.008	.966	.364	.052	Do you consider the people in the group to be charitable and kind?
q21	F	2.29 (.172)	5.58 (.481)	-.309	.103	**.552**	**.002**	Did you know the name of other people in your group?
q22	M	4.70 (.165)	5.09 (.343)	.023	.906	.092	.635	How motivated were you to work with the group?
q23	M	4.78 (.110)	5.02 (.275)	.045	.816	.037	.849	How motivated were you to do the task?
q24	M	4.09 (.221)	5.71 (.256)	-.143	.460	.412	.026	How proud would you be to tell others you were "happy to be a part" of the group?
q25	M	3.14 (.155)	3.00 (.355)	-.046	.813	-.010	.958	How much social pressure did you feel from the group to do well?

* Item-group — L: Linguistic communication; G: Gestural communication; F: Group friendliness and familiarity; D: Group disagreement and confusion; C: Group cooperation; M: Group motivation; P: Group planning and organization. Each of the questionnaire items was averaged for each group separately, providing an average group score, which was then correlated against the two dependent variables, percent of good tokens and mean consensus time (N = 29). **Pairwise correlations were Bonferroni corrected to yield significant analyses at the level of p = 0.002.

2.3.3 Communication and group size effects

I next seek to determine the relationship between communication patterns, group performance and group size. Firstly, I wanted to establish whether large groups involved both more gesturing and more speaking participants, compared to small groups. Next, I wanted to establish whether the number of speaking and gesturing participants was positively related to group performance. Group performance was here measured according to the same two dependent variables: percentage of 'good' tokens collected and mean group consensus time. Finally, I wanted to determine whether gesturing or speaking was the more important means of communication as group size increased. The analyses in this section comprised linear regression and multiple regression analyses.

A linear regression indicated that group size explained 71% of the variance in the number of gesturing subjects, $F(1, 25) = 61.018$, $r^2 = 0.71$, $p < 0.001$ (Figure 2.7). Similarly, group size explained 43% of the variance in the number of speaking subjects, $F(1, 25) = 18.734$, $r^2 = 0.43$, $p < 0.001$, although in this case, a cubic regression gave a significantly improved fit, $r^2 = 0.69$ (Figure 2.8), perhaps suggesting an upper limit on the number of individuals that can be involved in verbal exchanges.

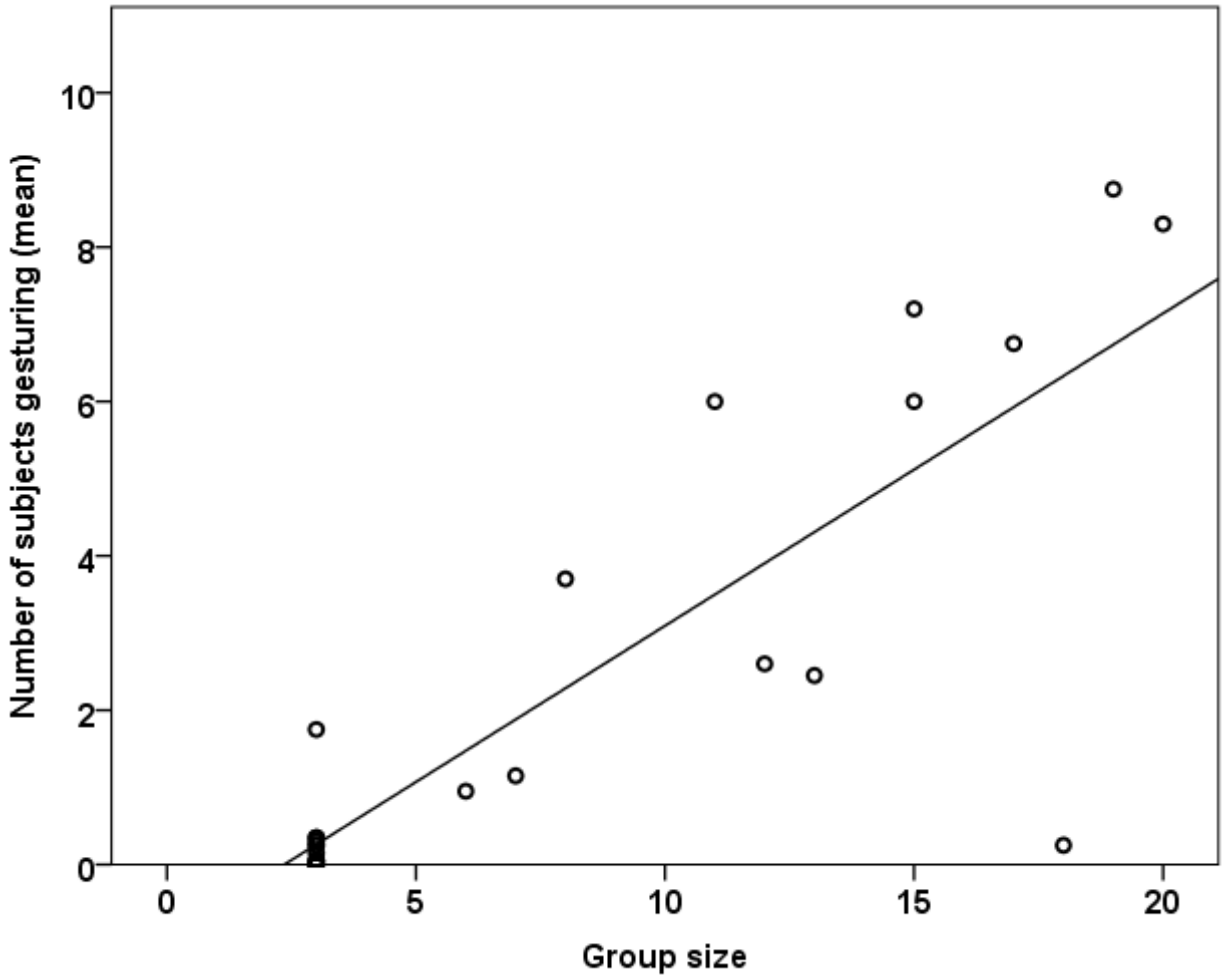

Figure 2.7. The relationship between the mean number of subjects gesturing at each 30 second interval and group size for each 10 minute trial. The equation for the best fit linear regression line is $y = .40x - .95$.

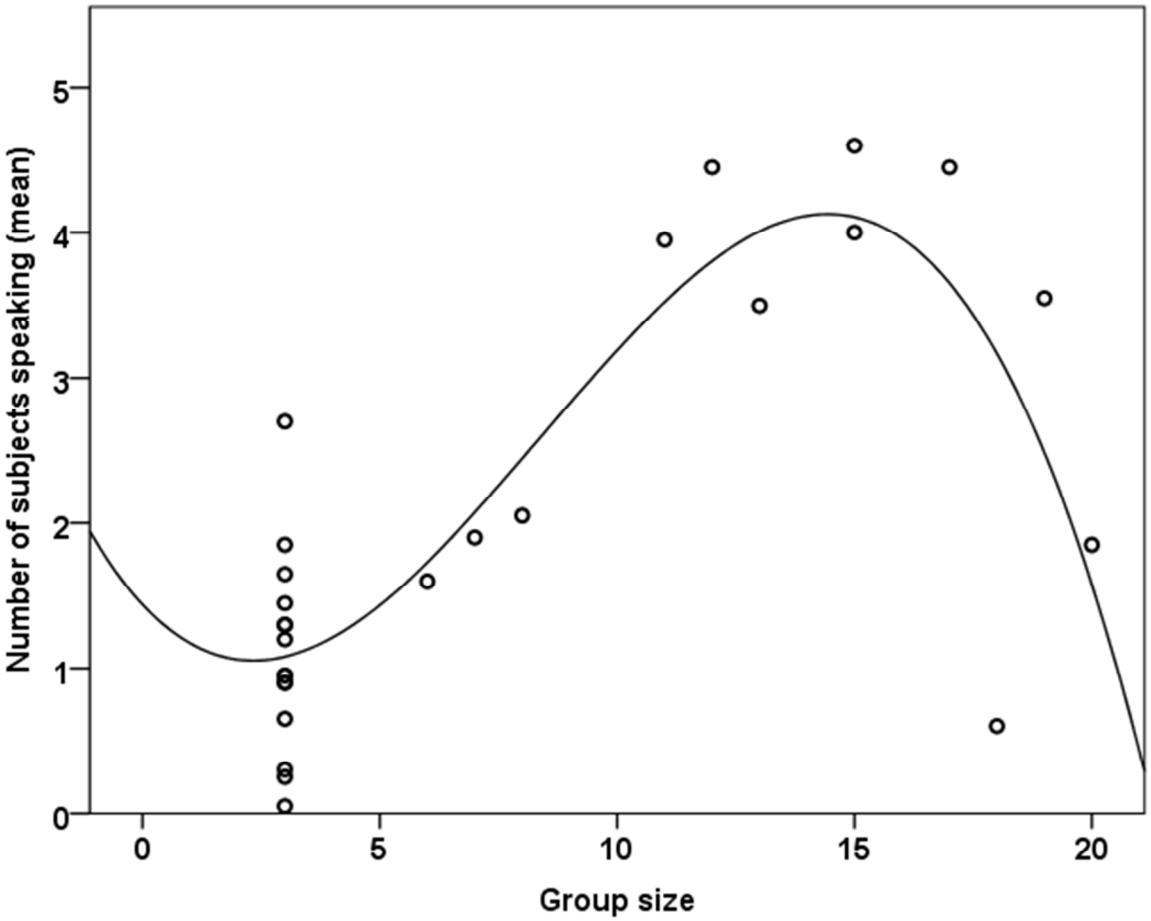

Figure 2.8. The relationship between the mean number of subjects speaking at each 30 second interval and group size for each 10 minute trial. The equation for the best fit cubic function is $y = .00347x^3 + .09x^2 - .35x + 1.44$.

Similarly, linear regression analysis revealed that the number of gesturing subjects explained 49% of the variance in mean consensus time, $F(1, 25) = 24.447$, $r^2 = 0.49$, $p < 0.001$, although a power function gives a slightly improved fit, $r^2 = 0.524$ (Figure 2.9), while the number of speaking subjects explained 38% of the variance in mean consensus time, $F(1, 25) = 15.103$, $r^2 = 0.38$, $p < 0.001$, although a quadratic function provides a slightly improved fit, $r^2 = 0.44$ (Figure 2.10).

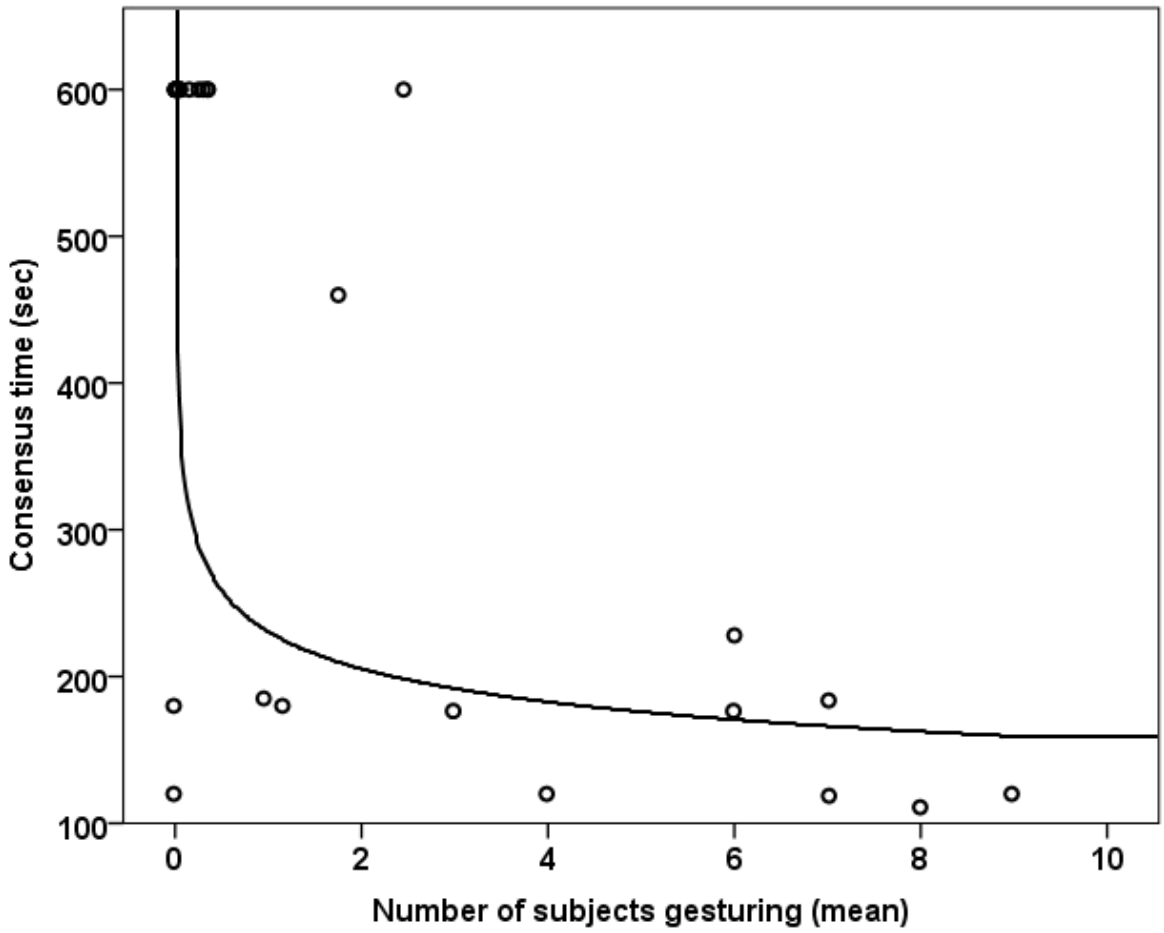

Figure 2.9. The relationship between the mean number of subjects gesturing at each 30 second interval and group consensus time for each 10 minute trial. The equation for the best fit power function is $y = 233.85x^{-0.167}$.

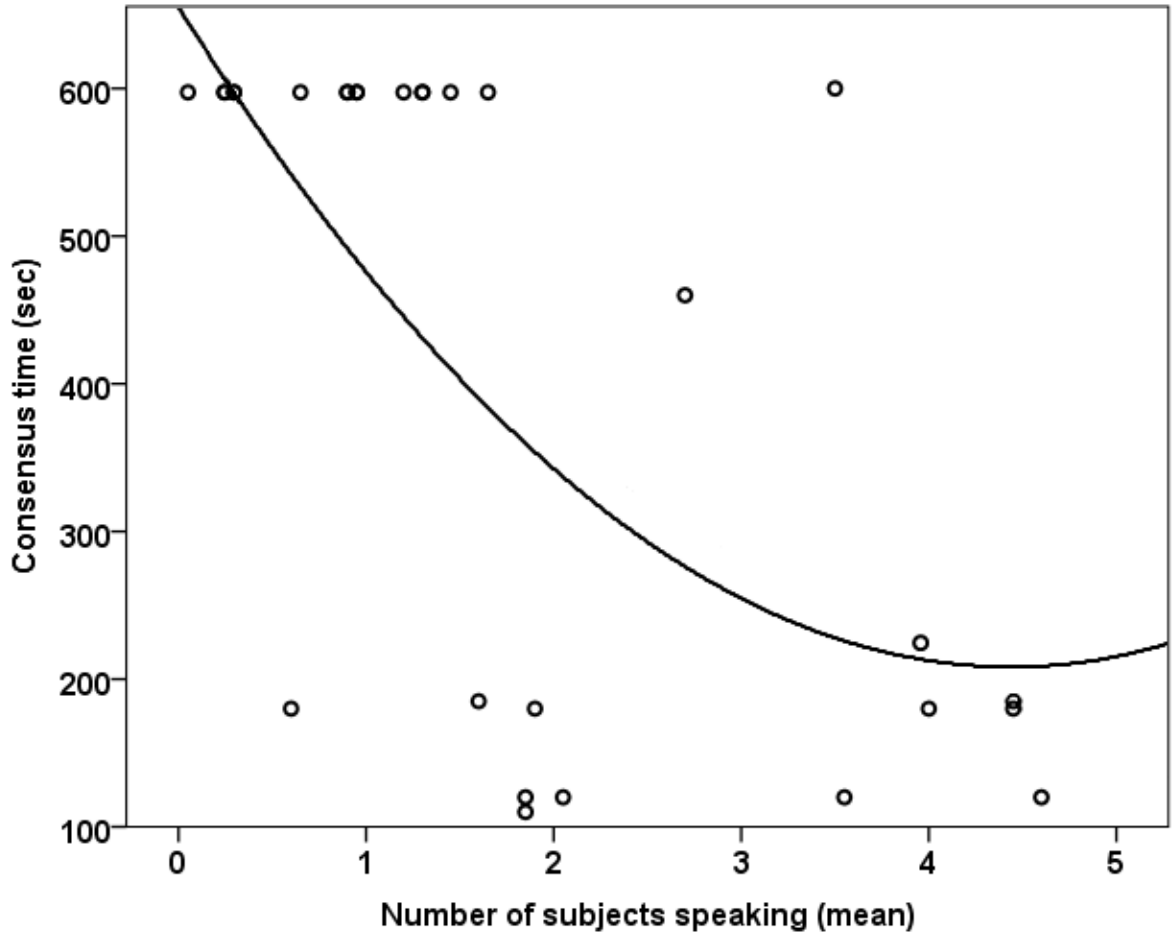

Figure 2.10. The relationship between the mean number of subjects speaking at each 30 second interval and group consensus time for each 10 minute trial. The equation for the best fit quadratic function is $y = 22.51x^2 - 202.89x + 667.87$.

In contrast, the results for number of good tokens collected were much weaker. The number of gesturing subjects explained 19% of the variance in percentage of 'good' tokens collected, $F(1, 25) = 5.675$, $r^2 = 0.19$, $p = 0.025$, with a cubic function yielding a significantly better fit, $r^2 = 0.32$ (Figure 2.11). Similarly, the number of speaking subjects explained 19% of the variance in percentage of 'good' tokens collected, $F(1, 25) = 5.711$, $r^2 = 0.19$, $p = 0.025$, with a cubic function providing an improvement in fit, $r^2 = .25$ (Figure 2.12). In all these cases, the data certainly suggest an asymptotic relationship, as indicated by the improved fits given by nonlinear regressions.

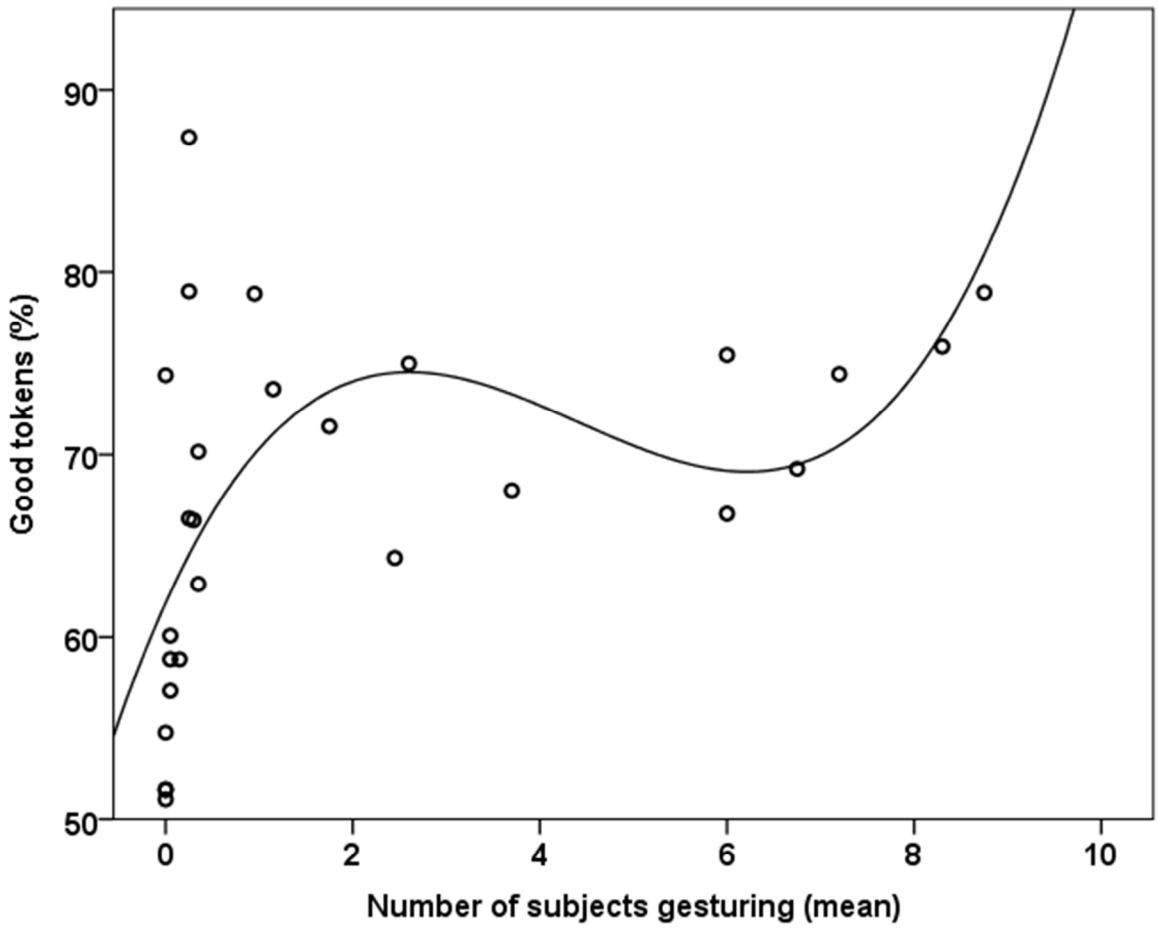

Figure 2.11. The relationship between the mean number of subjects gesturing at each 30 second interval and percentage of 'good' tokens collected for each 10 minute trial. The equation for the best fit cubic function is $y = .002x^3 - .03x^2 - .11x + .62$.

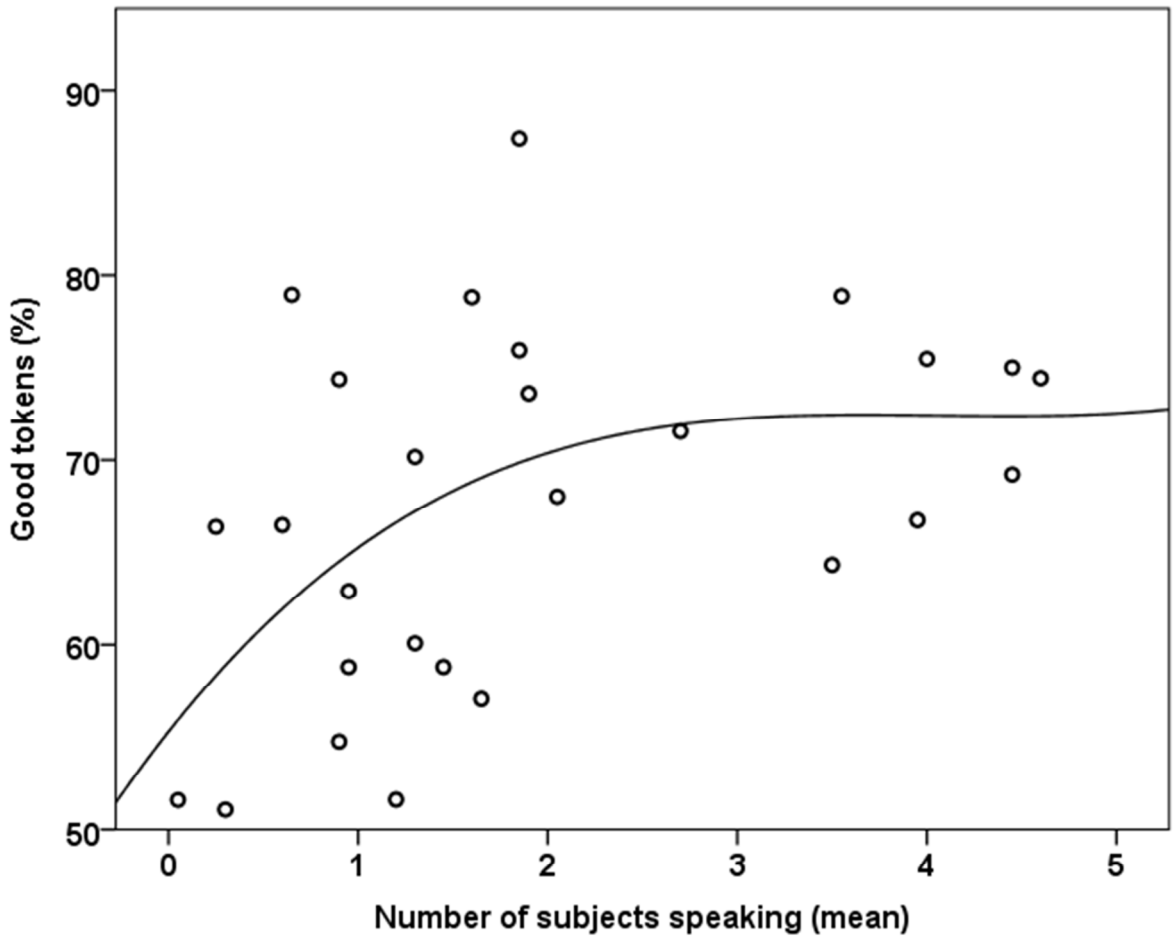

Figure 2.12. The relationship between the mean number of subjects speaking at each 30 second interval and percentage of 'good' tokens collected for each 10 minute trial. The equation for the best fit cubic function is $y = .003x^3 - .03x^2 + .13x + .55$.

To determine whether communication had an effect independently of group size on both mean consensus time and percentage of 'good' tokens collected, a multiple regression analysis was conducted to explore the relationship between mean consensus time as the dependent variable, with mean gesture, mean language and group size as independent variables (with all variables except group size being ln-transformed in the light of the previous results) (Table 2.2). In both cases, the overall model is significant, although only group size is a significant predictor in the case of consensus time and none of the predictor variables are individually significant in the case of percentage of 'good' tokens collected.

Table 2.2
Multiple linear regression analysis for variables predicting (a) ln-transformed mean consensus time and (b) ln-transformed percentage of 'good'tokens collected

a) ln-mean consensus time

	B	*SE B*	*β*	*t*	*p*	*df1*	*df2*
Constant	7.045	.370		19.065	.000	3	23
Mean gesture ln	-.273	.220	-.327	-1.242	.227	3	23
Mean language ln	.080	.301	.055	.264	.794	3	23
Group size	-.780	.290	-.566	-2.692	.013	3	23

Overall model: R=0.813
R^2=0.66
F=14.912, p<0.001

b) ln-percentage of 'good' tokens collected

	B	*SE B*	*β*	*t*	*p*	*df1*	*df2*
Constant	-.562	.112		-4.999	.000	3	23
Mean gesture ln	.036	.067	.203	.533	.599	3	23
Mean language ln	.070	.092	.229	.762	.454	3	23
Group size	.044	.088	.150	.494	.626	3	23

Overall model: R=0.537
R^2=0.288
F=3.103, p=0.046

2.4 Discussion

2.4.1 Summary of results

Chapter II introduced a novel experimentally-controlled human social foraging method to determine whether language indeed facilitates cooperation, as predicted by the interdependence hypothesis. Previous studies of this hypothesis had not experimentally tested prosociality, communication and behavioural cooperation in the context of a human social foraging paradigm. As such, this constitutes an important contribution of the present thesis.

The results of this study were consistent with the first prediction that performance would improve in larger groups as a result of more independent sources of information, in accordance with the wisdom of the crowd effect (James, 2004; Vul and Pashler, 2008). The large groups had faster overall group consensus times, though not necessarily a higher percentage of 'good' tokens collected. However, as indicated by the multiple regression analyses, group performance was not linearly related to either the number of speaking or gesturing participants. This suggests that the faster consensus time in large groups might be best explained by the opportunity provided for casual observational learning or behavioural mimicry.

The second prediction that performance would be greater in groups with strong social ties than in groups with weaker social ties was also supported. Indeed, as shown in Table 2.1, both group cooperation and the request for verbal information were positive predictors of group performance, at least in the case of the percentage of 'good' tokens collected; moreover, both group friendliness and familiarity were positive predictors of mean consensus time. Indeed, studies have shown that although group cohesion is in general positively related to group productivity (Beal et al. 2003; Robbins and Judge, 2013), the relationship between group cohesion and productivity depends on the existence of performance-related norms

established by the group (Beal et al. 2003; Robbins and Judge, 2013). That is, if performance-related norms are high, a cohesive group will be more productive than a less cohesive group; but if a cohesive group has low-performance norms, productivity will be low (Robbins and Judge, 2013). Accordingly, given group familiarity, closeness, and cooperation were among the main significant relationships found here does seem to indicate the significant impact close social relationships can have on group cooperation.

The third prediction was that language and gesture would be used more frequently in large groups to facilitate coordination than in small groups. As shown in Figures 2.7 and 2.8, both gesture and language use were positively related to increasing group size (albeit asymptotically). Note that while the number of gesturing subjects seemed to increase with no limit, the number of speaking participants seems to asymptote at around four individuals. Indeed, analysis of the data for large groups only reveals the mean number of gesturing subjects ($M = 4.51$, $SD = 3.01$) was above the mean number of speaking subjects ($M = 3.04$, $SD = 1.36$). This result may have been largely to do with the fact that, as group size increased, it gradually became more difficult for listeners to perceive and understand multiple subjects speaking at the same time (Argyle et al. 1968; Cohen, 1971; Leavitt and Mueller, 1955).

2.4.2 Implications

Inspection of the slope in Figure 2.8 suggests an approximate ratio of one speaker for every three listeners, as has been found by previous studies to be the limit on conversation group size (Dunbar et al. 1995; Waller et al. 2011; Dezecache and Dunbar, 2012). Consequently, there may be a critical limit on the number of conversation participants a social group can maintain at around four (Dunbar, 1993; Dunbar et al. 1995; Waller et al. 2011), although this does not necessarily seem to impede the effective transfer of

information. This may, in part, reflect the fact that individuals can resort to gesture, which does not suffer the same limitations from noise interference as auditory communication. However, the fact that the number of speakers seems to drop dramatically in groups of more than 15 members might be indicative of an escalating interference effect. It may also be that, in larger groups, individuals pay closer attention to each other, and thus monitor each other's success rates more effectively.

It is noteworthy that, on both performance criteria, performance does not seem to improve when there is more than one individual gesturing (Figures 2.9 and 2.11). This suggests that there may be an upper limit on the number of individuals that can productively gesture at any one time. In contrast, performance did seem to improve with more speakers (Figures 2.10 and 2.12), possibly reflecting the fact that it may be possible to maintain several independent conversations simultaneously when the group is large enough. If so, this may have important implications for the debate over the gestural versus vocal origin of language: speech would appear to have a clear advantage in group-based tasks. In sum, Figures 2.9-2.12 suggest that gesturing is less efficient than speaking; although lots of participants gestured, they didn't actually coordinate each other's behaviour as well as they did with speech.

An alternative interpretation of the results, following the social complexity hypothesis (Dunbar, 2003, 2004; Freeberg, Dunbar and Ord, 2012) is that language is more functionally critical for facilitating social cohesion than for the transfer of information. Groups that are well-coordinated and cohesive, as a result of language, may simply function better, because they can pay closer attention to what each other are doing and be better at interpreting each others' non-verbal signals, intentional or otherwise. A further possibility is that language is more critical for cultural learning (Baumeister et al. 2004; De Backer and Gurven, 2006), although the cultural learning hypothesis makes no explicit predictions concerning the effects of group size on language use or group performance. In many ways, the cultural learning

hypothesis predicts language should be primarily important for vocal information transmission and less so for other forms of cultural learning such as observation or imitation that might actually be important in the kinds of tasks considered here. Our results suggest that group size was directly related both to communication complexity (indexed as the number of speaking and gesturing subjects) and group performance, just as the social complexity hypothesis would predict (Freeberg, Dunbar and Ord, 2012). However, the limit to conversation group size at four suggests that large groups cannot be coordinated through language alone. As implied by the positive effect of familiarity on convergence time (Table 2.1), it may require formal or informal rules of coordination based on prior experience and 'intentional' familiarity with one's fellow group members. Such rules depend, of course, on the use of language.

A potential objection to the results found here is that larger groups will inevitably do better than smaller groups on the grounds that there are simply more hands to collect tokens. However, the evidence on group research indicates that smaller groups are generally faster and more efficient than larger groups when the task goals are obvious and clearly defined (Robbins and Judge, 2013; Seijts and Latham, 2000). On the other hand, if the group is engaged in problem solving, larger groups consistently do better than smaller groups, especially when the goal of the group is related to fact-finding and gaining diverse input (Robbins and Judge, 2013). The results here are therefore consistent with previous studies which find that larger groups tend to exchange more information (likely due to observational learning and behavioural mimicry) than smaller groups when diverse information is critical to group consensus and problem solving.

The results of this study were also consistent with the prediction that performance would be greater in groups with strong social ties than groups with weaker social ties, at least in so far as consensus time is concerned. This also seemed to be the case for proportion of

'good' tokens collected, which correlated positively with cooperation received and the request for verbal information. It is likely that this was because performance-related norms were high for some reason in large groups in this particular case (cf. Robbins and Judge, 2013). Nevertheless, Robbins and Judge (2013) note that group performance may be enhanced in such cases by making groups smaller, encouraging agreement with group goals, increasing the time participants spend together, increasing the status of the group and the perceived difficulty of attaining membership in the group, stimulating competition with other groups, giving rewards to the group rather than individual members, and physically isolating the group. Future research may need to ensure, wherever possible, that the majority of these guidelines are consistently adhered to across all groups.

In summary, the findings here showed little evidence for the interdependence hypothesis of human language function and evolution; more specifically, for the claim that language and cooperation coevolved together according to an interdependent process, whereby language was critical for human cooperation. Although the first two predictions were supported, namely, that language would be used more frequently in large group coordination than in small group coordination and that performance would increase in larger groups, the multiple regression analyses indicated that neither verbal nor gestural communication were critical for overall group performance. Indeed, the human social foraging groups seemed to cooperate well in spite of the limitations imposed on gestural and vocal communication, ostensibly relying much more on imitation and behavioural mimicry. Lastly, the potential social bonding and cooperation effects, as measured by the self-report prosociality index questions, were also evident in the human social foraging experiment. Namely, the positive correlations indicated that groups that were subjectively more cooperative and familiar with one another performed better on the human social foraging

task. In conclusion, this experiment did not provide results consistent with the hypothesis that language or gesture facilitates cooperation within an ecologically valid experiment.

2.4.3 Limitations

There are several potential limitations to the experiment conducted for this study. First, it is possible that the group sizes tested were simply not large enough to detect significant differences in performance due to vocal communication. Indeed, language has been proposed as a mechanism by which humans effectively bond socially with more than one individual at a time, thereby facilitating the development of larger social networks beyond the limit of 50 individuals allowed by dyadic primate strategies such as manual grooming (Dunbar, 2004). The group sizes tested in this experiment did not go above 20 individuals, so it is possible that language is only needed to facilitate cooperation and group cohesion with very much larger human group sizes. If this is the case, then future studies should perhaps investigate larger group size limits, associated with the potentially cooperative effects of language, as well as different types of cooperation and coordination. Nevertheless, this possible interpretation seems unlikely for reasons already mentioned; namely, consistent with a variety of other studies, group performance typically plateaus for language-using participants at around four individuals – the natural size of human conversation groups.

A second limitation of the experiment could have been related to the lack of a direct quantification of cooperation. Although cooperation, in the human social foraging experiment, was ostensibly measured in terms of the percentage of good tokens collected and mean group consensus time, it is conceivable that neither of these measures are an accurate metric of cooperation. Although it seems likely, based on the results, that participants worked mostly independently, and converged on a group consensus through observation and

behavioural mimicry, it is also possible that, had the measure of cooperation been more valid, the results would have been consistent with the interdependence hypothesis. For instance, the direct quantification of cooperation could have potentially been achieved through the direct coding of the video files by multiple naïve coders and/or through motion energy analysis software, which allows for continuous objective quantification of body positions and nonverbal behaviour. Nevertheless, even in the best of cases where advanced technology and methods could have been used, it is not obvious that this would generate a better index of cooperation than the ones I used. As such, while it is not intuitively obvious how a better measure of cooperation could be designed for this experiment (especially since the human social foraging experiment is itself a novel contribution), future studies should consider whether the measure of cooperation tested in this experiment is valid.

A final limitation of the experiment may have been related to participant recruitment. More specifically, there may in fact have been confounds due to gender and/or random assignment of participants. For instance, potential limitations may have been present due to the fact that only females were tested for these series of experiments (primarily to reduce confounds due to mixed sex interactions). Future studies should perhaps replicate this experiment on all male and mixed sex groups. However, previous research has shown, in accordance with expectation states theory, that men will have an advantage over women in mixed sex groups, other things being equal (Wittenbaum, 1998). Accordingly, this could have quite drastic consequences on group performance, and may be undesirable for testing functional hypotheses of the kind considered for this experiment. As such, all male groups may be the ideal alternative.

Moreover, there may not have been truly random assignment of participants to group size conditions. Although earnest efforts were made for random assignment at the beginning of each trial, the nonsignificant negative correlations found between group performance and

several measures of group cooperation and familiarity may suggest otherwise. Future research should attempt to better control for this possible confound, by testing participants for acquaintanceship prior to the experiment, to ensure no prior affiliation of any kind. In conclusion, future studies should replicate this experiment on all male groups, address the challenges of potential variability in prior acquaintanceship of participants, and standardise indices for measuring cooperation and prosociality.

Chapter III

Honest signalling theory and human language: An experimental test of four hypotheses

3.1 Introduction

Given that the results from Chapter II indicated that language did not seem to aid or facilitate group cooperation, I elected to test several other alternative hypotheses of human language evolution. In contrast to the interdependence or cooperation hypothesis for human language function, it has been suggested that language might be used to influence the behaviour of conspecifics in other ways, under the rubric of honest signalling theory. One version of this hypothesis is couched in terms of advertising strategies in mate choice contexts (Burling, 1986; Locke and Bogin, 2006; Miller, 2000), while the other emphasizes the skilled use of deception and manipulation of fellow group members (Bliege Bird and Smith, 2005; Donath, 2008; Greengross and Miller, 2008; Lachmann et al. 2001).

Sexual advertisement, in this sense, is defined as an honest signal of genetic quality or social status which increases the probability of reproductive success (Burling, 1986; Miller, 1999, 2000). Deception, in this sense, refers to the skilled use of self-deprecation, which might also signal moral virtues such as prosociality and humility (Greengross and Miller, 2008; Miller, 2007), as well as imperviousness to danger and social repercussions (Donath, 2008; Bliege Bird and Smith, 2005). Indeed, it has been suggested that the high prevalence of linguistically honest information in the everyday use of language (and potential reputation-based social costs of lying and deception) may have evolved as a socially or sexually selected handicap (Bliege Bird and Smith, 2005; Donath, 2008; Lachmann et al. 2001). Chapter III

therefore takes an experimental approach to determine whether aspects of human language can be characterised in terms of honest signalling theory.

Honest signalling theory seeks to explain what keeps communication honest and reliable in the face of so many incentives for deception (Donath, in press; Maynard-Smith and Harper, 2003). Evolutionary biologists have agreed that honest signals generally fall into two basic categories: receiver-dependent costs (i.e. those that reflect agonistic, mating and social costs) and receiver-independent costs (i.e. those that reflect underlying biology including production, development, and maintenance costs) (Searcy and Nowicki, 2005). In many human social interactions, there are motivations for deception, as people want to make the best possible impression (to appear important, dominant, creative, and popular) while others want to know if they actually possess these qualities (Donath, 2008; DePaulo and Kashy, 1998).

As human language is perhaps the most salient and elaborate of human signalling systems, many recent attempts have been made to characterize aspects of human language according to honest signalling theory (Bliege Bird and Smith, 2005; Lachmann et al. 2001; Miller, 1999). For example, some have proposed that vocabulary size may have evolved as an honest signal of human intelligence (as a receiver-independent cost), and should therefore be attractive to prospective mates (Miller, 1999, 2000; Rosenberg and Tunney, 2008). Alternatively, others have suggested that self-disclosure of revealing or culpable personal information, a kind of social risk-taking (as a receiver-dependent cost), may be a form of social advertisement, while also signalling imperviousness to danger and social repercussions (Donath, 2008; Bliege Bird and Smith, 2005). However, many have considered these proposals speculative and unconvincing, as well as lacking in substantive empirical evidence (Briscoe, 2008; Fitch, 2005; Hurford, 2007; Zahavi and Zahavi, 1997).

An alternative view is that honest signalling may have become a secondary or derived selection pressure for human language evolution, once it had already evolved in response to the primary selection pressure that originally brought it into existence (Dunbar, 2009; Millikan, 1984; Origgi and Sperber, 2000). The original selection pressure, it has been argued, is that language originally evolved for social communication, thereby facilitating social bonding and group cohesion, in large human communities (Dunbar, 1993, 1996). In the context of human language evolution, this distinction is often referred to as the difference between *direct* and *derived* functions (Millikan, 1984; Origgi and Sperber, 2000), whereby sexual selection may have been one of the more significant secondary selective mechanisms (Miller, 1999).

In the current study, I tested two related hypotheses in Experiments I and II, that a large and elaborate vocabulary evolved as an honest signal of human intelligence (as a receiver-independent cost), and should therefore be attractive to prospective mates. In contrast, Experiments III and IV test two related hypotheses that self-disclosure of revealing or culpable personal information, as a kind of social risk-taking (as a receiver-dependent cost), may be a form of social advertisement through signalling imperviousness to danger and/or the social repercussions of disclosure. The relevant background and details for these four sets of experiments are outlined below.

3.1.1 Vocabulary size as an honest signal

Human vocabulary size has been noted to differ quite dramatically among people, is at least 60 percent genetically heritable, and has about an 80 percent correlation with general intelligence (Jensen, 1998; Plomin et al. 1997; Stromswold, 2001). Moreover, intelligence has been established as one of the single most important traits desired in a long-term partner for both males and females (Buss, 1989; Buss, Abbott and Angleitner et al. 1990).

In addition, studies have also found that people in long-term relationships usually have vocabularies of similar sizes (Mascie-Taylor, 1988). Further evidence for this hypothesis has shown that males do in fact use more low-frequency vocabulary when in the presence of attractive females, whereas females seem unmotivated to speak in this respect when in the presence of attractive males (Rosenberg and Tunney, 2008). According to this hypothesis, human vocabulary size may have evolved through sexual selection as an honest signal, in much the same way as bird song repertoire size evolved in song birds (Miller, 2000).

On the other hand, some authors have cast doubt on this interpretation, arguing that fluent low-frequency vocabulary users are often judged as *less* intelligent than more plain-spoken individuals (Oppenheimer, 2006). However, this particular study did not specifically investigate use of low-frequency vocabulary in the context of mate attraction.

> *Experiment I therefore tests the reciprocal hypothesis of the supporting results found by Rosenberg* and *Tunney (2008), that use of low-frequency vocabulary is more sexually attractive to the opposite sex than the use of high-frequency words.*

Additionally, as it may be 'flowery' language rather than simply novel words that provide the cues for mate choice, I also ran a variant of Experiment I that used more poetic words rather than merely rare words. Indeed, given that poetic words are just as rare as low-frequency words in everyday conversation, and have been similarly argued to be an effective courtship strategy (Miller, 2000) I further elected to test this unique hypothesis.

> *Experiment II therefore tests the hypothesis that use of more poetic vocabulary is more sexually attractive to the opposite sex than more common vocabulary.*

3.1.2 Self disclosure as an honest signal

On the other hand, recent work by Donath (2008) and others has argued that human honest signalling theory should emphasize status and influence, with concurrent costs including the risk of unfavorable social reputation in an information-based society (Donath, in press; McCracken, 1998; Thornton, 1996). For example, at least one social anthropologist has argued that nowhere would this phenomenon be more salient among traditional societies than in the context of oratory and public speaking (Burling, 1986). Indeed, according to charismatic leadership theory, several characteristics are commonly attributed to extraordinary leadership ability, including having an appealing vision and a willingness to take personal risks to achieve that vision (Robbins and Judge, 2007). A range of traits have further been identified which tend to correlate with group leadership, including extraversion, sociability, and assertiveness, as well as verbal and social intelligence (Van Vugt, 2006), suggesting the salience brought by public speaking may be an honest signal of these traits (with concurrent costs of social embarrassment and unfavourable social reputation).

Moreover, evolutionary biologists have demonstrated that punishment and the enforcement of strict social sanctions are relatively common occurrences in many social species (Clutton-Brock and Parker, 1995). Badges of status, of direct relevance here, are a subset of honest signals that reveal information about an individual's genetic quality, including body size and social dominance (Searcy and Nowicki, 2005; Tibbetts and Dale, 2004). For example, Tibbetts and Dale (2004) found that, in staged contests between pairs of unfamiliar wasps, subordinates with experimentally enhanced badges of status (i.e. variable facial patterns indicating social dominance), received considerably more aggression from dominants than did sham controls (indicating facial patterns are signals and dishonest signalling imposes social costs).

An alternative possibility arises from the apparent ubiquity of deception throughout

the animal kingdom. Some theorists have argued that animal communication has evolved solely and primarily (as opposed to secondarily as a derived function) for the purposes of deception. For instance, in a classic series of articles by Dawkins and Krebs (1978) and Krebs and Dawkins (1984), it was argued that animal communication, and by extension, human language ought to be viewed as an evolutionary arms race in which signallers evolve to become better at manipulating receivers, while receivers evolve to become more resistant to manipulation. Moreover, at least one prominent evolutionary linguist has voiced a very similar assertion, namely that human language evolved primarily and specifically for the function of mind-reading and manipulation of group conspecifics (Scott-Phillips, 2006). According to this hypothesis, human language is viewed as a fundamentally competitive method of communication.

However, although deception may accurately characterise certain types of animal communication in certain contexts, there are strong reasons to doubt it can be identified as the primary function for the majority of animal communication systems, including human language. Firstly, behavioral ecologists have argued on theoretical grounds that Dawkins and Krebs' (1978, 1984) signal evolution models (which argue that signallers evolve to manipulate receivers, while receivers evolve to resist manipulation) are in fact critically flawed; if there is, on average, no information of benefit to the receiver of a signal, then receivers should evolve to ignore the signal (in which case signalling no longer has any benefit to the signaller and the communication system collapses) (Searcy and Nowicki, 2005: 8). Granted, a partial solution to this paradox was proposed by Zahavi (1975), but was limited specifically to signals with respect to mate choice or aggressive contexts; as males might be inclined to deceive or manipulate females (or other male rivals) with respect to their genetic superiority, such males were actually 'handicapped', thereby ensuring the honesty of the signal (Searcy and Nowicki, 2005: 9). Clearly, then, such models apply only in a relatively

narrow range of contexts, and it is furthermore unclear whether such animal models can be generalised to human language (as there are currently no existing data to support or indicate that they do in fact apply to human language).

In addition, analyses of actual human conversational content reveal that social relationship topics overwhelmingly dominate most everyday conversations, accounting for about two-thirds of total conversation time (Bischoping, 1993; Dunbar et al. 1997; Haviland, 1991). Indeed, an increasing amount of evidence currently suggests that language primarily evolved for facilitating social bonding and group cohesion (Cohen, 2012; Dunbar, 2003, 2004; Freeberg, Dunbar and Ord, 2012; Ireland et al. 2011; Nettle and Dunbar, 1997; Pietraszewski and Schwartz, 2014; Roberts, 2010, 2013; Tomasello, 2008; Weaver and Bosson, 2011). Moreover, as in any biological system that has significant time and energetic costs as well as adaptive benefits (Barrett, Dunbar and Lycett, 2002), it certainly begs the question as to why we spend most of our time in the social use of language, if language reputedly evolved primarily for manipulation and deception.

In addition, studies have further shown that in humans, lying and deception are relatively rare occurrences in the general population (DePaulo et al. 1996; Hancock, Thom-Santelli and Ritchie, 2004; Serota, Levine and Boster, 2010). For example, Serota, Levine and Boster (2010) found, in a sample of 1,000 Americans, the average number of lies told per day was 1.65. Moreover, the distribution was highly skewed; 22.7% of all lies were told by one percent of the sample, and half of all lies were told by 5.3% of the sample. Studies conducted in controlled laboratory environments have obtained similar results (Abeler, Becker and Falk, 2012). Given so many clear incentives to deceive, often in the interest of manipulating rivals (Hauser, 1992; Dawkins and Krebs, 1978; Whiten and Byrne, 1988), many investigating the evolution of language have questioned how linguistic communication can remain primarily honest in the everyday use of language (Fitch, 2010; Knight, 1998;

Nettle, 1999). One possible explanation, as noted above, is the risk of unfavourable social reputation and implementation of strict social sanctions against lying and deception (Bliege Bird and Smith, 2005; Donath, 2008; Lachmann et al. 2001).

Furthermore, previous findings have also revealed that individuals seem to be much more willing to deceive when the potential stakes are quite high, as for example when attempting to advertise to prospective mates, suggesting a context-sensitive signalling device (Ellison, Heino and Gibbs, 2006; Hall, Park, Song and Cody, 2010; Toma, Hancock and Ellison, 2008). For instance, one study of online daters in New York City found that 87% of men and 76% of women lied about at least one attribute in their personal profiles; men tended to increase their height, whereas women tended to lower their weights (Toma, Hancock and Ellison, 2008). Deviations from truth were generally quite small in magnitude, however, suggesting that individuals were at least partially aware of the possible social repercussions from being discovered as deceitful (Toma, Hancock and Ellison, 2008).

Deception and active signal falsification have further been found within animal communication to manipulate mates for the purpose of reproductive benefits (Brø-Jorgensen and Pangle, 2010). For example, male topi antelopes have been found to alarm snort deceptively in order to retain receptive females within the males' territory (Brø-Jorgensen and Pangle, 2010). As such, deceptive alarm snorting would appear to be a secondary or derived function, rather than the primary function (as previously noted for human language) (Millikan, 1984; Origgi and Sperber, 2000).

Finally, it is important to note that social psychologists typically characterise the degree of human deception according to four different types of lying (including various other subcategories): 1) *prosocial,* lying to protect someone, or to benefit or help others 2) *self-enhancement,* lying to save face, avoid embarrassment, disapproval or punishment or gain an advantage (i.e. lies not intended to hurt anyone, rather they benefit the self) 3) *selfish,* lying to

protect oneself at the expense of another and/or to conceal a misdeed 4) *antisocial,* lying to hurt someone else intentionally (DePaulo et al. 1996; Iñiguez et al. 2014). In this study, I distinguish between two general types of lies: prosocial lies of type 1 and antisocial lies of types 2-4. Of these two general types, it is widely recognised that antisocial lies are destructive of relationships, because they are selfish (the liar gains fitness at the expense of the target of the lie), whereas it has been claimed that prosocial lies often help to keep relations in good condition (DePaulo and Kashy, 1998; Nyberg, 1993). For present purposes, I further distinguish an additional subcategory of prosocial lies, *self-deprecation,* as various studies have shown it often makes the deceiver seem modest, accessible and human, while simultaneously making others feel comfortable by downplaying the differences in their abilities (Bornstein et al. 1996; Schlenker and Leary, 1982).

In Experiments III-IV, I aimed to investigate: 1) whether reliable indicators of intelligence and social status may lead to increased attractiveness ratings, and 2) whether strict social sanctions may partially ensure honest linguistic communication in humans.

> *Experiment III tests the prediction that public speaking may constitute an honest signal of intelligence, status, and influence as evidenced by increased attractiveness ratings (with concurrent potential costs of unfavorable social reputation).*
>
> *Experiment IV tests the prediction that prosocial courtship deception leads to increased attractiveness ratings, while antisocial courtship deception leads to decreased attractiveness ratings (as a direct test of both increased social status and social sanctions for deception within the context of mate attraction).*

In sum, the current study aimed to test several competing, but not necessarily mutually exclusive, hypotheses characterising human language in terms of honest signalling theory.

All experiments used a vignette-based approach, hosted on Amazon Mechanical Turk (AMT), an Internet application that provides instant access to thousands of potential participants for questionnaire-based psychology experiments (among various other cognitive tasks) for a small monetary payment. Several studies have demonstrated the validity of online surveys and shown them to be as reliable as traditional paper-and-pencil laboratory questionnaires (Buchanan and Smith, 1999; Epstein et al. 2001; Gosling et al. 2004; Metzger et al. 2003; Preckel and Thiemann, 2003; Salgado and Moscoso, 2003; Smith and Leigh, 1997).

Previous studies have had mixed success with the vignette-based approach. A study on the attractiveness of personal chat-up lines, for example, found good agreement among raters about the relative effectiveness of each questionnaire item, supporting a vignette-based approach (Bale et al. 2006). Haselton and Miller (2006) also found success in using a vignette-based approach in testing for attractiveness ratings between creative and uncreative males. However, the authors encountered a problem in content bias; the difficulty of creating vignettes with sufficient realism introduced details that were unrelated to the intended variable being measured. Another recent study on attractiveness found that autobiographical statement vignettes were a successful methodology for measuring attractiveness ratings for humour (Bressler and Balshine, 2006). However, these vignettes were generally one-line statements that typically incorporated few extraneous details as a contextual backdrop for each vignette. Perhaps because of this, the authors did not seem to encounter similar problems with item content as a secondary influence on raters' judgments. Despite these potential limitations, vignettes constructed as author-written dialogues were chosen for the

clear advantage of more control over vignette content than perhaps would be reasonable in a real-time experimental context.

3.2 Experiment I: Effect of low-frequency vocabulary vignettes on heterosexual attractiveness ratings

Experiment I manipulated the vocabulary used in the vignettes, in terms of rare versus common words, and tested three specific predictions. Firstly, the use of uncommon low-frequency vocabulary will succeed in making an individual appear more intelligent to the opposite sex. This was interpreted as a manipulation check to ensure that the vocabulary manipulation did affect perceived intelligence. Second, vignettes with low-frequency vocabulary will be judged as more sexually attractive and more desirable as a date by the opposite sex, relative to vignettes with simpler more straightforward vocabulary. Third, if the use of low-frequency vocabulary words is in fact unsuccessful, the failure of the strategy will be explained, at least in part, by simultaneous lower ratings on several personality dimensions included in the study, such as warmth, pleasantness and dominance and higher ratings on dimensions such as pretentiousness.

3.2.2 Methods

3.2.2.1 Participants

An online questionnaire was distributed to participants via AMT. The study was advertised as being about a "male (female) personality impression questionnaire," and all responses were anonymous, voluntary, and restricted to individuals over 18 years of age. All participants provided informed consent before participation and were paid $0.50 for completing the 20 minute survey.

Subjects were 180 female and 200 male undergraduates, postgraduates and working professionals from a range of backgrounds from the United States (81%), Canada (6%), the United Kingdom (6%), Australia (2%), the Czech Republic (1%), Ireland (1%), Spain (1%), Trinidad and Tobago (1%) and industrialised nations with available Internet access. High school education was completed by 100% of participants, with 60% having attained a diploma or some college experience, and 33% having completed a bachelor's degree or higher. All participants were native English first language speakers. To avoid developmental effects and confounds due to declining social engagement in old age, participants were restricted to the age range 18–65 years (M = 22.0, SD = 6.3).

3.2.2.2 Materials

A short descriptive vignette, containing low-frequency vocabulary, and depicting a relaxing vacation holiday (i.e. a prospective date describing a beach scene) was presented to half of the participants to read (version 1). A second short descriptive vignette, containing normal vocabulary, and depicting the same vacation holiday was presented to the remaining half of the participants to read (version 2). Each vignette was approximately 70 words long and was matched for content, number of words, sentences, and overall complexity, as these factors have been shown to affect participant ratings (Bale et al. 2006).

The low-frequency vocabulary words were selected based on rare English word frequency lists found in the *Brown Corpus* compiled by Francis and Kučera (1967). The normal frequency vocabulary words were selected based on the simplest equivalent synonym for each low-frequency vocabulary word. The study used a between-subjects design, where each participant was presented with either vignette version 1 or version 2. After each vignette had been read, participants answered questions about the individual who had reportedly authored the vignette.

Female subjects read a description of male protagonist Todd using low-frequency vocabulary (version 1):

> Hi- I'm Todd and I want to take a long sojourn this year to Cocos Island. It's unearthly - balmy weather, sparse tourists, and enchanted beaches. I love to swim and lay out in the sand - it's the consummate place to relax. But there's also lots of things to do like scuba dive and surf. I'd really like to find someone venturesome to accompany me this year.

or normal vocabulary (version 2):

> Hi- I'm Todd and I want to go on break this year to Cocos Island. It's really nice - warm weather, not too many tourists, and really pretty beaches and stuff. I love to swim and lay out in the sand - it's the perfect place to relax. But there's also lots of things to do like scuba dive and surf. I'd really like to find someone fun to come with me.

Male subjects read descriptions that were identical except that the target person was named Barbara in the first instance or Caroline in the second case.

3.2.2.3 Procedure

Participants were first asked to provide information about their age, sex, ethnicity, education, income, and whether they were native English speakers. Half of the pool of participants was then presented with a short descriptive vignette to read, containing low-frequency vocabulary (version 1). The remaining half of the participants was presented with a very similar, but slightly modified short descriptive vignette, not containing rare vocabulary (version 2). After the vignette had been read, subjects were then presented with instructions informing them that the experimenters are interested in their "ability to form accurate impressions of others when one has only very limited information to go on", before evaluating the protagonist in the vignette. Participants were then asked to rate the stimulus person in the vignette on several 7-point bipolar scales, referred to as a "personality impression rating", which included the dependent variables and a number of filler items (adapted from Sadalla, Kenrick and Vershure, 1987).

3.2.2.4 Experimental design

The following 9 bipolar adjectives were used as dependent variables in the study: dominant-nondominant, sexually attractive- unattractive, masculine-unmasculine, warm-cold, pleasant-unpleasant, pretentious-unpretentious, intelligent-unintelligent, tough-tender, desirable-undesirable as a date. Each bipolar dimension was used as a 7-point scale. The intelligent-unintelligent dimension constituted a manipulation check. Two items measured attraction and constituted the central dependent variables: sexually attractive- unattractive and desirable-undesirable as a date. The remaining items were used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent evidence for the effectiveness of the experimental manipulation. The study used a 2 (sexually attractive/desirable as a date) within-subjects x 2 (sex) x 2 (low-frequency/normal vocabulary) between-subjects factorial ANOVA.

3.2.3 Results

3.2.3.1 Manipulation check

A one-way ANOVA, with intelligence rating as the dependent variable and high vs low frequency vocabulary as the independent variable, clearly supported the effectiveness of the manipulation, with targets who displayed low-frequency vocabulary rated as significantly more intelligent ($M = 5.40$, $SE = .093$) than targets who displayed normal-frequency vocabulary ($M = 4.36$, $SE = .090$). Table 3.1 gives the results of the analysis of variance for the ratings on intelligence, with vocabulary and sex as the independent variables, for the results shown in Figure 3.1.

Table 3.1. ANOVA results summary (dependent variable: intelligence)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	101.480	101.480	64.315	.000	.146
Sex	1	9.924	9.924	6.290	.013	.016
Vocabulary * sex	1	50.435	50.435	31.964	.000	.078
Error	376	593.270	1.966			
Total	380	9803.000				

An independent samples *t*-test confirmed that male targets who displayed low-frequency vocabulary were rated as significantly more intelligent than male targets who displayed normal-frequency vocabulary ($M = 5.60$ vs. 3.83), $t(178) = 8.950$, $p < .001$, Cohen's $d = 1.342$. In contrast, although the difference was in the same direction, female targets who displayed low-frequency vocabulary were not rated as significantly more intelligent than female targets who displayed normal vocabulary ($M = 5.19$ vs. 4.89), $t(198) = 1.810$, $p = .072$, Cohen's $d = 0.257$ (Figure 3.1).

In addition, male targets were rated significantly higher on the following dimensions in the normal-frequency vocabulary condition: pleasant ($M = 5.26$ vs. 4.70), $t(178) = 2.643$, $p = .009$, Cohen's $d = 0.396$, masculine ($M = 5.59$ vs. 4.64), $t(178) = 5.451$, $p < .001$, Cohen's $d = 0.817$, and tough ($M = 4.84$ vs. 4.24), $t(178) = 3.241$, $p < .001$, Cohen's $d = 0.486$. Female targets were rated significantly higher in the low-frequency vocabulary condition only for pretentiousness ($M = 4.46$ vs. 3.98), $t(198) = 2.159$, $p = .032$, Cohen's $d = 0.307$. For the remaining traits, the difference between high and low frequency vocabulary was not significant: pleasant ($M = 5.29$ vs. 5.27), $t(198) = .110$, $p = .913$, Cohen's $d = 0.016$, masculine ($M = 3.74$ vs. 3.50), $t(198) = .953$, $p = .342$, Cohen's $d = 0.135$, tough ($M = 4.22$ vs. 3.83), $t(198) = 1.896$, $p = .059$, Cohen's $d = 0.269$, warm ($M = 5.22$ vs. 5.06), $t(198) = .924$, $p = .356$, Cohen's $d = 0.131$, and dominant ($M = 4.44$ vs. 4.13), $t(198) = 1.489$, $p = .138$, Cohen's $d = 0.212$.

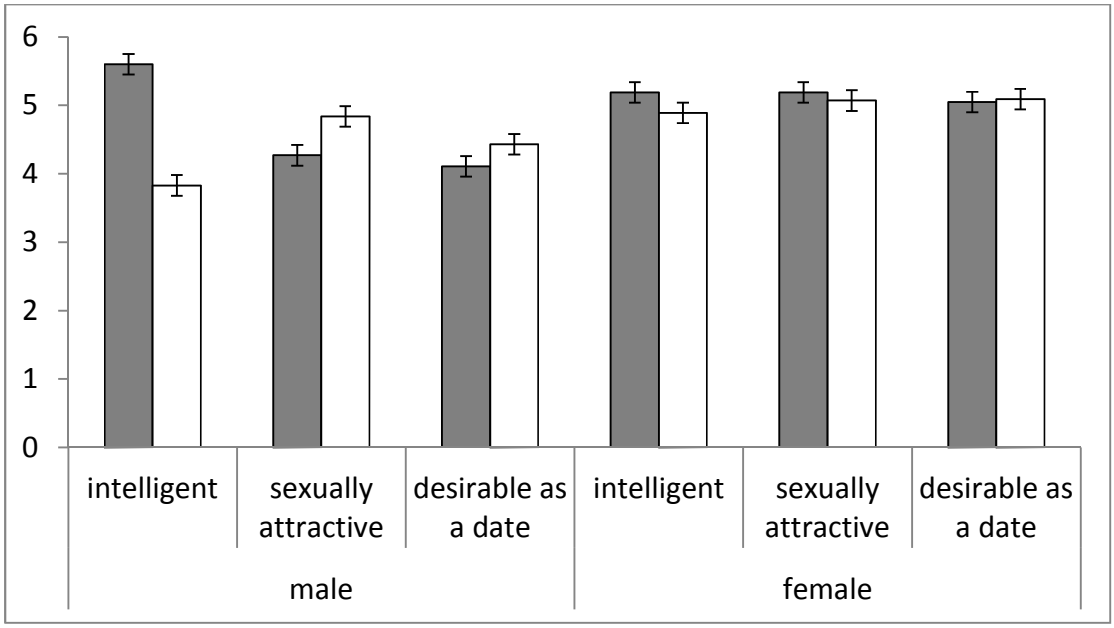

Figure 3.1. Mean values of major dependent variables for male and female targets. Shaded bars depict low-frequency vocabulary ratings, while white bars indicate normal frequency vocabulary ratings. Higher ratings on dependent variables indicate that subjects perceived target persons as more intelligent, sexually attractive, and desirable as a date. Error bars denote standard error of the mean.

3.2.3.2 Attraction measures

Table 3.2 gives the results of the analysis of variance for the ratings on sexual attractiveness as the dependent variable, with vocabulary and sex as the independent variables, for the results shown in Figure 3.1. Vocabulary did not have a significant effect on sexual attractiveness ratings, although sex did with females being rated as more attractive than males; the sex x vocabulary interaction was also significant. Note that the effect in males was the reverse of that predicted: males who used low-frequency vocabulary were rated as *less* attractive than males who used high-frequency vocabulary.

Table 3.2. ANOVA results summary (dependent variable: sexually attractive)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	4.777	4.777	2.587	.109	.007
Sex	1	31.133	31.133	16.856	.000	.043
Vocabulary * sex	1	11.774	11.774	6.375	.012	.017
Error	376	694.480	1.847			
Total	380	9699.000				

Table 3.3 gives the results of the analysis of variance for the ratings on dating desirability as the dependent variable, with vocabulary and sex as the independent variables, for the results shown in Figure 3.1. Vocabulary did not have a significant effect on dating desirability ratings, although sex did with females being rated as more desirable as a date than males; the sex x vocabulary interaction was not significant. Note that the effect in males was the reverse of that predicted: males who used low-frequency vocabulary were rated as *less* desirable as a date than males who used high-frequency vocabulary.

Table 3.3. ANOVA results summary (dependent variable: desirable as a date)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	3.095	3.095	1.432	.232	.004
Sex	1	60.701	60.701	28.081	.000	.069
Vocabulary * sex	1	1.886	1.886	.873	.351	.002
Error	376	812.785	2.162			
Total	380	9254.000				

3.2.4 Discussion

The results of Experiment I supported prediction 1 that users of low-frequency vocabulary are viewed as being more intelligent. However, the results did not support prediction 2 that use of low-frequency vocabulary enhances the sexual attractiveness or dating desirability of males or females. In contrast to the predictions of the sexual selection hypothesis, the direction of the effect was opposite to that predicted: targets that used normal-frequency vocabulary were rated as significantly more sexually attractive (though not more desirable as a date) than targets that used low-frequency vocabulary, but only significantly so in the case of males.

In addition, males that used normal-frequency vocabulary were also rated higher on the dimensions of pleasant, masculine, and tough, suggesting that males that used low-

frequency vocabulary were perceived to be less masculine, pleasant and tough than those who used every day common language. Females that used low-frequency vocabulary were rated significantly higher on pretentiousness, suggesting that females that used low-frequency vocabulary were perceived to be more pretentious than those who used more common vocabulary. Between them, these findings clearly support prediction 3. Lastly, the finding that females were rated significantly more sexually attractive and more desirable as dates than were males would seem to be consistent with evolutionary psychology theory predicting that females should be choosier than males for both short and long-term relationships (Pawlowski and Dunbar, 2001; Voracek, Hofhansl and Fisher, 2005): these results can be interpreted as suggesting that females were less easy to please than males.

In spite of the fact that the data for males did not support predictions made from the sexual selection hypothesis for human vocabulary size (Miller, 2000), these results might be open to alternative explanations. Although the manipulation check suggests male targets that used low-frequency vocabulary were successful in conveying greater intelligence than those that used normal-frequency vocabulary, it is possible that the wording in these vignettes may have simultaneously displayed other traits in the low-frequency vocabulary condition that conveyed less attractive qualities.

Indeed, given the lower ratings on masculinity, pleasantness, and toughness, for males in the low-frequency vocabulary condition, and the higher ratings on pretentiousness for females in the low-frequency vocabulary condition, this seems justified. Therefore, it is conceivable that male targets able to use low-frequency vocabulary, without simultaneously communicating negative qualities, might gain higher attractiveness ratings than male targets using normal-frequency vocabulary. However, given the magnitude of the effect found in our sample, it seems unlikely the fine-tuned verbal manipulation necessary for this effect could produce a result essentially opposite to the results found here.

3.3 Experiment II: Effect of poetic vocabulary vignettes on heterosexual attractiveness ratings

Experiment I tested the hypothesis that use of rare, low-frequency words might be sexually attractive, but did not find the predicted effect. However, it is possible that what makes language skills attractive is not a large vocabulary, including many rare words, but the ability to use language in a more poetic, romantic way, especially in the context of mate choice. It has been suggested that universal aspects of poetry, such as poetic rhythm, meter and rhyme may have evolved as a system of sexually selected handicaps (Miller, 2000). Experiment II tests this possibility.

Experiment II tested three specific predictions. Firstly, use of poetic vocabulary will succeed in making an individual appear more intelligent to the opposite sex. Again, this functioned as a manipulation check. Second, use of poetic vocabulary will also increase an individual's attractiveness ratings by the opposite sex, relative to use of simpler more straightforward vocabulary. Third, if the use of poetic vocabulary is in fact unsuccessful, the failure of the strategy will be explained, at least in part, by simultaneous lower ratings on several personality dimensions included in the study, such as warmth, pleasantness and dominance and higher ratings on dimensions such as pretentiousness.

3.3.1 Methods

3.3.1.1 Participants

Participants were recruited via AMT. Subjects were 172 female and 163 male undergraduates, postgraduates and working professionals from a range of backgrounds from India (64.7%), North America (27.3%), Macedonia (0.6%), Qatar (0.6%), the United Kingdom (0.3%), Australia (0.3%), Italy (0.3%), Kenya (0.3%), Spain (0.3%), Australia

(0.3%), Russia (0.3%) and other industrialised nations with available Internet access. High school education was completed by 100% of participants, with 88.2% having attained a diploma or some college experience, and 79.6% having completed a bachelor's degree or higher. All participants were native English first language speakers. Participants were restricted to the age range 18–65 years (M = 30.6, SD = 8.7).

3.3.1.2 Materials

A short descriptive vignette, containing poetic vocabulary, and depicting a relaxing vacation holiday (i.e. a prospective date describing a beach scene) was presented to half of the participants to read (version 1). A second short descriptive vignette, containing normal vocabulary, and depicting the same vacation holiday was presented to the remaining half of the participants to read (version 2). Each vignette was approximately 75 words long and was matched for content, number of words, sentences, and overall complexity, as these factors have been shown to affect participant ratings (Bale et al. 2006).

The poetic vocabulary word vignettes were adapted from a poem entitled, "Light: A Narrative Poem," by nineteenth century American poet Joaquin Miller (Miller, 1907). The normal frequency vocabulary words were selected based on the simplest equivalent synonym for each poetic vocabulary word. The study used a between-subjects design, where each participant was presented with either vignette version 1 or version 2. After each vignette had been read, participants answered questions about the individual who had reportedly authored the vignette.

Female subjects read a description of male protagonist Joaquin using poetic vocabulary (version 1):

> Hello- I am Joaquin and departing soon for the Galapagos. You will not soon forget – white sea-gulls glisten in the sun, ten thousand if a single one. The sea is as an opal sea, inlaid with scintillating light, yet close about and left and right. The sea lies banked and bossed in night, as black as every night may be. The sundown sea is shone, mobile, translucent, flaming molten steel; I eagerly await our trip with zeal**.**

or normal vocabulary (version 2):

> Hello- I am Joaquin and leaving soon for the Galapagos. It is a nice place and I won't forget about it very soon – lots of birds are flying around. You can see the sun reflected in the water. The ocean is very dark and always moving about quite a lot. The waves crash on the beach. The sunsets are pretty and make the sky look nice. I would really like to go with someone fun.

Male subjects read descriptions that were identical except the target person was named Caroline.

3.3.1.3 Procedure

Participants were first presented with an online questionnaire and asked to provide information about their age, sex, ethnicity, education, income, and whether they were native English speakers. The participants were then presented with one of the two vignettes, with half being given vignette version 1 and the other half vignette version 2. After the vignette had been read, subjects were then presented with instructions informing them that the experimenters are interested in their "ability to form accurate impressions of others when one has only very limited information to go on", before evaluating the protagonist in the vignette. Participants were then asked to rate the stimulus person in the vignette on several 7-point bipolar scales, referred to as a "personality impression rating", which included the dependent variables and a number of filler items (adapted from Sadalla, Kenrick and Vershure, 1987).

3.3.1.4 Experimental design

As in Experiment I, the following 9 bipolar adjectives were used as dependent variables in the study: dominant-nondominant, sexually attractive-unattractive, masculine-unmasculine, warm-cold, pleasant-unpleasant, pretentious-unpretentious, intelligent-unintelligent, tough-tender, desirable-undesirable as a date. Each bipolar dimension was presented as a 7-point scale. The intelligent-unintelligent dimension constituted a manipulation check. Two items measured attraction and constituted the central dependent

variables: sexually attractive-unattractive and desirable-undesirable as a date. The remaining items were used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent evidence for the effectiveness of the experimental manipulation. The study used a 2 (sexually attractive/desirable as a date) within-subjects x 2 (sex) x 2 (poetic/normal vocabulary) between-subjects factorial ANOVA.

3.3.2 Results

3.3.2.1 Manipulation check

The one-way ANOVA results failed to support the effectiveness of the manipulation; in contrast to the prediction, targets who displayed normal vocabulary were not rated as significantly less intelligent (M = 5.27, SE = .085) than targets who displayed poetic vocabulary (M = 5.10, SE = .107). Table 3.4 gives the results of the analysis of variance for the ratings on intelligence, with vocabulary and sex as the independent variables, for the results shown in Figure 3.2. Vocabulary did not have a significant effect on intelligence ratings, although sex did with females being rated as more intelligent than males; the sex x vocabulary interaction was not significant. Note that the effect in females was the reverse of that predicted: females who used poetic vocabulary were rated as *less* intelligent than females who used normal frequency vocabulary. Although female targets that used poetic vocabulary were rated as less intelligent than those who used normal vocabulary, the difference failed to reach statistical significance in the case of male targets.

Table 3.4. ANOVA results summary (dependent variable: intelligence)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	3.087	3.087	2.068	.151	.006
Sex	1	10.224	10.224	6.847	.009	.020
Vocabulary * sex	1	1.574	1.574	1.054	.305	.003
Error	331	494.293	1.493			
Total	335	9526.000				

An independent samples *t*-test indicated that female targets who displayed normal vocabulary were rated as significantly more intelligent than female targets who displayed poetic vocabulary, $t(161) = 1.974$, $p < .05$, Cohen's $d = 0.311$, but male targets who displayed normal vocabulary were not rated significantly more intelligent than male targets who displayed poetic vocabulary, $t(153) = .262$, $p = .794$, Cohen's $d = 0.042$ (Figure 3.2). In addition, male targets were rated significantly higher in the normal vocabulary condition for pleasantness ($M = 5.07$ vs. 4.56), $t(169) = 2.348$, $p = .02$, Cohen's $d = 0.361$. However, the differences between the normal and poetic vocabulary conditions on all of the following dimensions were not significantly different: masculine ($M = 4.72$ vs. 4.38), $t(169) = 1.598$, $p = .112$, Cohen's $d = 0.246$, pretentious ($M = 4.55$ vs. 4.71), $t(169) = .689$, $p = .492$, Cohen's $d = 0.106$, warm ($M = 5.06$ vs. 4.81), $t(169) = 1.272$, $p = .205$, Cohen's $d = 0.196$, and dominant ($M = 4.56$ vs. 4.51), $t(169) = .275$, $p = .783$, Cohen's $d = 0.042$.

Female targets were also rated significantly higher in the normal vocabulary condition on: pleasantness ($M = 5.64$ vs. 5.11), $t(161) = 2.626$, $p = .009$, Cohen's $d = 0.414$, but not significantly different between the two vocabulary conditions on all of the following dimensions: masculine ($M = 4.33$ vs. 3.78), $t(161) = 1.969$, $p = .051$, Cohen's $d = 0.310$, pretentious ($M = 4.70$ vs. 4.95), $t(161) = 1.135$, $p = .258$, Cohen's $d = 0.179$, warm ($M = 5.30$ vs. 4.99), $t(161) = 1.777$, $p = .078$, Cohen's $d = 0.280$, and dominant ($M = 4.87$ vs. 4.71), $t(161) = .773$, $p = .441$, Cohen's $d = 0.122$.

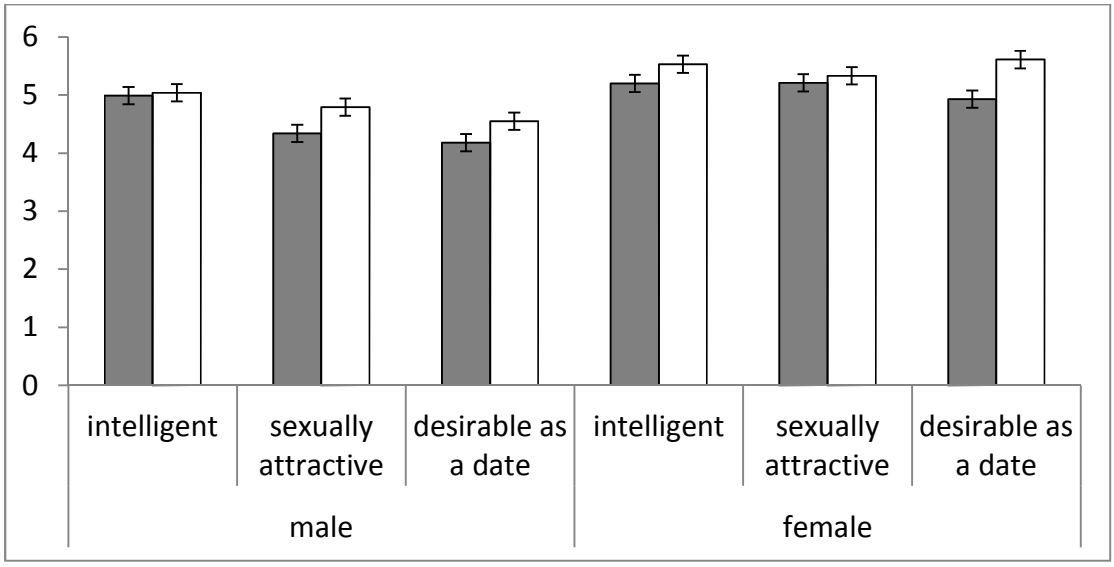

Figure 3.2. Mean values of major dependent variables for male and female targets. Shaded bars depict poetic vocabulary ratings, while white bars indicate normal vocabulary ratings. Higher ratings on dependent variables indicate that subjects perceived target persons as more intelligent, more sexually attractive, and more desirable as a date. Error bars denote standard error of the mean.

3.3.2.2 Attraction measures

Table 3.5 gives the results of the analysis of variance for the ratings on sexual attractiveness, with vocabulary and sex as the independent variables, for the results shown in Figure 3.2. Vocabulary did not have a significant effect on sexual attractiveness ratings, although sex did with males being rated as more attractive than females; the sex x vocabulary interaction was not significant. Note that the effect in males was the reverse of that predicted: males who used poetic vocabulary were rated as *less* attractive than males who used normal vocabulary.

Table 3.5. ANOVA results summary (dependent variable: sexually attractive)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	6.633	6.633	3.374	.067	.010
Sex	1	41.405	41.405	21.065	.000	.060
Vocabulary * sex	1	2.391	2.391	1.217	.271	.004
Error	331	650.615	1.966			
Total	335	8806.000				

Table 3.6 gives the results of the analysis of variance for the ratings on dating desirability, with vocabulary and sex as the independent variables, for the results shown in Figure 3.2. Vocabulary had a significant effect on dating desirability ratings, as did sex, with females being rated as more desirable as a date than males; the sex x vocabulary interaction was not significant. Note that, once again, the effect in females was the reverse of that predicted: females who used poetic vocabulary were rated as *less* desirable as a date than females who used normal vocabulary.

Table 3.6. ANOVA results summary (dependent variable: desirable as a date)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Vocabulary	1	23.159	23.159	10.673	.001	.031
Sex	1	68.185	68.185	31.425	.000	.087
Vocabulary * sex	1	2.183	2.183	1.006	.317	.003
Error	331	718.204	2.170			
Total	335	8586.000				

3.3.3 Discussion

The results of Experiment II did not support prediction 1 that users of poetic vocabulary are viewed as being more intelligent than users of normal vocabulary. In fact, the direction of the effect was opposite to that predicted: females that used poetic vocabulary were rated as *less* intelligent than females that used normal vocabulary, although not significantly so in the case of males. The results also did not support prediction 2 that use of poetic vocabulary enhances the sexual attractiveness or dating desirability of targets. In contrast to the predictions of the sexual selection hypothesis, the direction of the effect was opposite to that predicted: targets that used normal vocabulary were rated as significantly more desirable as a date than targets that used poetic vocabulary, though only significantly so in the case of females.

In addition, female targets that used normal vocabulary were also rated higher on the pleasant-unpleasant dimension, demonstrating that female targets that used poetic vocabulary were perceived to be less pleasant than those who used every day common language. This finding offers tentative support for prediction 3. Lastly, the finding that female targets were rated significantly more sexually attractive and more desirable as dates than were male targets is, again, consistent with evolutionary psychology theory predicting that females should be choosier than males for both short and long-term relationships (Pawlowski and Dunbar, 2001; Voracek, Hofhansl and Fisher, 2005).

Although the data did not support predictions made from the sexual selection hypothesis for poetic courtship (Miller, 2000), these results might be open to alternative explanations. The first point to note is that the manipulation check suggested that the manipulation did not work in the way expected: targets that used poetic vocabulary were not rated as more intelligent than those that used normal vocabulary. This might suggest that the wording used in the vignettes was not realistic enough. It is also possible that the vignette wordings might have displayed other traits in the poetic vocabulary condition that conveyed less attractive qualities. Indeed, the fact that the ratings were lower on the pleasant-unpleasant dimension offers support for this conclusion. Given this, it is conceivable that poetic vignettes that lacked these negative signals might have been more attractive than normal vocabulary vignettes. Nonetheless, given the very similar results for each condition found in this sample, it seems unlikely that fine-tuning the verbal manipulation would produce a result essentially opposite to that found here.

3.4 Experiment III: Effect of public speaking vignettes on heterosexual attractiveness ratings

Given that Experiments I-II failed to support the sexual selection hypothesis for human vocabulary, I next considered the alternative hypothesis that oratory and public speaking skills might be a more honest signal (as a receiver-dependent cost) of social status in the context of human mate choice. The basic suggestion is that public speaking in large groups may have evolved to function as a fitness indicator, as a socially risky honest signal not only of general intelligence, but also of social status - in effect, a more general test for attractiveness of verbal fluency. Indeed, it has been suggested that the universal prevalence of oratory may have evolved as a sexually selected handicap (Burling, 1986). A similar conclusion is suggested by the finding that men, but not women, adjust their conversation to discuss academic topics in the presence of the opposite sex (Dunbar et al. 1995) and that women, but not men spend less time speaking as the number of males in the group increases (Dunbar et al. 1997).

Experiment III tested three specific predictions. Firstly, evidence of the ability to do effective public speaking will cause the individual to be rated as more intelligent by the opposite sex. Second, women will rate men higher in attractiveness judgements if they can provide evidence of effective public speaking, whereas men will not rate women differently (as men don't seem to pay attention to women speaking in normal conversation). Third, if the strategy is in fact only successful when speaking with social relevance, the lack of success in socially awkward contexts will be explained, at least in part, by lower ratings on several personality dimensions included in the study, such as warmth, pleasantness and dominance. In effect, the failure of public speaking as a mate choice cue will be at least partially explained according to costly social blunders such as awkwardness or embarrassment.

3.4.1 Methods

3.4.1.1 Participants

As before, participants were recruited via AMT. Subjects were 310 female and 276 male undergraduates, postgraduates and working professionals from a range of backgrounds from India (52.5%), North America (39.8%), Macedonia (1.33%), Mexico (0.5%), Austria (0.5%), Bangladesh (0.5%), Italy (0.33%), Singapore (0.33%), the United Kingdom (0.16%) and other industrialised nations with available Internet access. High school education was completed by 100% of participants, with 90% having attained a diploma or some college experience, and 77% having completed a bachelor's degree or higher. All participants were native English first language speakers. Subjects were restricted to the age range 18–65 years (M = 33.6, SD = 10.9).

3.4.1.2 Materials

A series of four short descriptive vignettes, depicting a protagonist speaking to a small group of people, was presented to half of the participants to read (version 1). A second series of four short descriptive vignettes, depicting the same protagonist speaking to a large audience of people, was presented to the remaining half of the participants to read (version 2). Each vignette was approximately 80 words long and was matched for content, number of words, sentences, and overall complexity, as these factors have been shown to affect participant ratings (Bale et al. 2006).

Each set of four author-written descriptive vignettes were comprised of a protagonist in a socially relevant and a socially awkward context; two additional filler vignettes were also included, depicting a protagonist in a socially clever and a socially shy context (Appendix A.3.1). The filler vignettes were used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent

evidence for the effectiveness of the experimental manipulation.

The study used a between-subjects design, where each participant was presented with the set of four vignettes from either version 1 or version 2 (Table 3.7). After each vignette had been read, participants answered questions about the vignette.

Table 3.7. Small group (version 1) and large group (version 2) public speaking vignettes

Version 1: Small group public speaker	***Version 2: Large group public speaker***
Steve: socially awkward (test item)	Steve: socially awkward (test item)
Tim: socially shy (filler item)	*Tim: socially shy (filler item)*
John: socially clever (filler item)	*John: socially clever (filler item)*
Steve: socially relevant (test item)	Steve: socially relevant (test item)

3.4.1.3 Procedure

Participants were first presented with an online questionnaire and asked to provide information about their age, sex, ethnicity, education, income, and whether they were native English speakers. Each participant was then presented with a series of four short descriptive vignettes to read in their entirety. Two of the four descriptive vignettes were presented to the protagonist depicting either an embarrassing context or relevant context, while the remaining two vignettes were presented as filler items. Female subjects read a description of male protagonist Steve, while male subjects read identical descriptions, apart from the target named Susan (Appendix A.3.1). After each vignette had been read, participants answered questions about each vignette.

Subjects then rated the target on a "personality impression rating" that included the same bipolar personality dimensions used in Experiments I and II. Subjects were presented with instructions informing them that the experimenters are interested in their "ability to form accurate impressions of others when one has only very limited information to go on", before evaluating the protagonist in the vignette.

The order of the vignettes within each set (i.e. version 1 or version 2) was counterbalanced across subjects, and the study was administered according to a between-subjects design, where each participant was presented with vignettes from either version 1 or version 2.

3.4.1.4 Experimental design

The same 9 bipolar adjectives used in Experiment I and II - in addition to two further dimensions, awkward-unawkward and publically prominent-not publicly prominent - were used as dependent variables for Experiment III. Each bipolar dimension was presented as a 7-point scale. The intelligent-unintelligent, awkward-unawkward and publically prominent-not publicly prominent dimensions constituted a manipulation check. Two items measured attraction and constituted the central dependent variable: sexually attractive-unattractive and desirable-undesirable as a date. The remaining items were used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent evidence for the effectiveness of the experimental manipulation.

The study used a 2 (sexually attractive/desirable as a date) x 2 (social relevance/ awkwardness) within-subjects x 2 (sex) x 2 (small/large group) between-subjects factorial ANOVA.

3.4.2 Results

The data for this experiment were analysed using a substantive number of ANOVAs and *t*-tests concerning two different group sizes each within two different social contexts. Consequently, due to the number of permutations of conditions involved in this experiment, the formal statistical results as presented below are rather complicated, and perhaps,

unavoidably confusing. To alleviate difficulty of interpretation, the results are summarised in diagrammatic form in Table 3.13 at the end of the Results section.

3.4.2.1 Manipulation check

The one-way ANOVA results clearly supported the effectiveness of the group size manipulation, with targets who spoke publicly in large groups rated as significantly more publicly prominent (M = 5.25, SE = .082) than targets who spoke publicly in small groups (M = 4.91, SE = .082) (Figure 3.3).

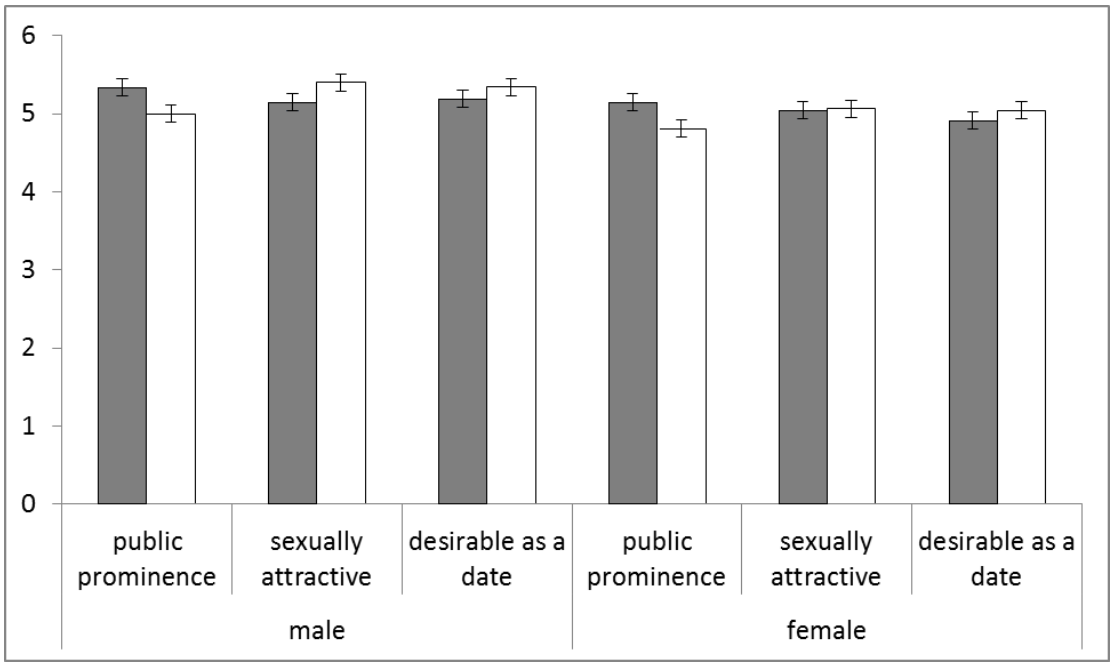

Figure 3.3. Mean values of major dependent variables, for male and female targets, as a function of group size. Shaded bars depict large group ratings, while white bars indicate small group ratings. Higher ratings on dependent variables indicate that subjects perceived target persons as more publicly prominent, sexually attractive, and desirable as a date. Error bars denote standard error of the mean.

Table 3.8 gives the results of the analysis of variance for the ratings on public prominence, with group size and sex as the independent variables, for the results shown in Figure 3.3. Group size had a significant effect on public prominence ratings, although sex did not and the sex x group size interaction was not significant.

Table 3.8. ANOVA results summary (dependent variable: public prominence)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Group size	1	20.384	20.384	8.710	.003	.007
Sex	1	.619	.619	.264	.607	.000
Group size * sex	1	1.483	1.483	.634	.426	.001
Error	1168	2733.413	2.340			
Total	1172	28436.000				

An independent samples *t*-test indicated that male targets who spoke publicly in large groups were rated as significantly more publicly prominent than male targets who spoke in small groups, $t(308) = 2.289$, $p = .023$, Cohen's $d = 0.261$. However, female targets who spoke publicly in large groups were not rated as significantly more publicly prominent than female targets who spoke publicly in small groups, $t(274) = 1.846$, $p = .066$, Cohen's $d = 0.223$.

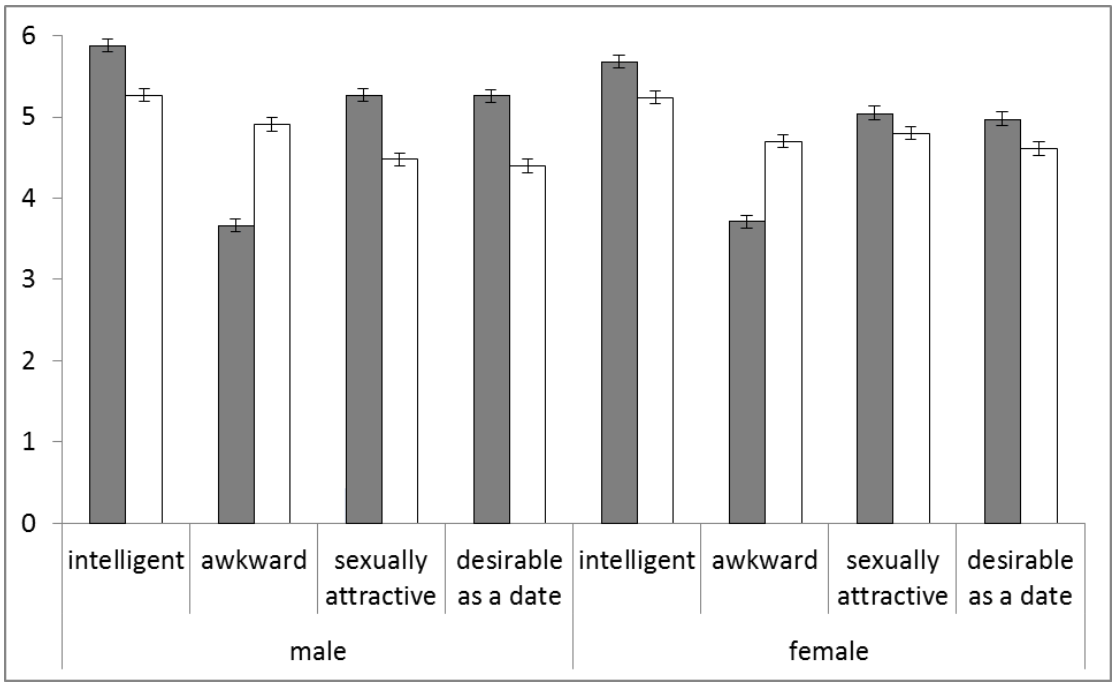

Figure 3.4. Mean values of major dependent variables, for male and female targets, as a function of the speaker's performance. Shaded bars depict socially relevant vignette ratings, while white bars indicate socially awkward vignette ratings. Higher ratings on dependent variables indicate that subjects perceived target persons as more intelligent, more awkward, more sexually attractive, and more desirable as a date. Error bars denote standard error of the mean.

The one-way ANOVA results also clearly supported the effectiveness of the social context manipulation (Figure 3.4), with targets who spoke awkwardly rated significantly more awkward (M = 4.80, SE = .064) than targets who spoke relevantly (M = 3.69, SE = .064), independent of group size. Table 3.9 gives the results of the analysis of variance for the ratings on awkwardness, with social context and sex as the independent variables, for the results shown in Figure 3.4. Social context had a significant effect on awkwardness ratings, although sex did not have a significant effect, and the sex x social context interaction was also not significant.

Table 3.9. ANOVA results summary (dependent variable: awkwardness)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Social context	1	361.158	361.158	151.304	.000	.115
Sex	1	2.400	2.400	1.005	.316	.001
Social context * sex	1	4.680	4.680	1.961	.162	.002
Error	1168	2787.988	2.387			
Total	1172	24340.000				

In addition, a relevant speaker was rated as more intelligent than an awkward speaker (M = 5.78, SE = .046 vs. M = 5.26, SE = .046), independent of group size. Table 3.10 gives the results of the analysis of variance for the ratings on intelligence, with social context and sex as the independent variables, for the results shown in Figure 3.4. Social context had a significant effect on intelligence ratings, although sex did not, and the sex x social context interaction was not significant.

Table 3.10. ANOVA results summary (dependent variable: intelligence)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Social context	1	79.208	79.208	62.955	.000	.051
Sex	1	4.030	4.030	3.203	.074	.003
Social context * sex	1	1.983	1.983	1.576	.210	.001
Error	1168	1469.554	1.258			
Total	1172	37274.000				

3.4.2.2 Attraction measures

Table 3.11 gives the results of the analysis of variance for the ratings on sexual attractiveness, with group size, social context and sex as the independent variables, for the results shown in Figure 3.4. Social context and group size had a significant effect on sexual attractiveness ratings, although sex did not have a significant effect; the sex x social context interaction was significant. However, the social context x group size, group size x sex and social context x sex x group size interactions were not significant.

Table 3.11. ANOVA results summary (dependent variable: sexually attractive)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Social context	1	78.507	78.507	45.351	.000	.038
Group size	1	7.469	7.469	4.315	.038	.004
Sex	1	.666	.666	.385	.535	.000
Social context * sex	1	21.569	21.569	12.460	.000	.011
Social context * group size	1	.125	.125	.072	.788	.000
Group size * sex	1	3.811	3.811	2.202	.138	.002
Social context * sex * group size	1	.006	.006	.004	.952	.000
Error	1164	2014.980	1.731			
Total	1172	30255.000				

Table 3.12 gives the results of the analysis of variance for the ratings on dating desirability, with group size, social context and sex as the independent variables, for the results shown in Figure 3.4. Social context and group size had a significant effect on dating desirability ratings, although sex did not have a significant effect; the sex x social context interaction was significant. The social context x group size, group size x sex and social context x sex x group size interactions were not significant.

In summary, targets who spoke with social relevance had higher ratings on both sexual attractiveness and dating desirability, relative to socially awkward targets.

Table 3.12. ANOVA results summary (dependent variable: desirable as a date)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Social context	1	110.899	110.899	54.099	.000	.044
Group size	1	12.814	12.814	6.251	.013	.005
Sex	1	.452	.452	.220	.639	.000
Social context * sex	1	17.991	17.991	8.776	.003	.007
Social context * group size	1	1.570	1.570	.766	.382	.001
Group size * sex	1	1.090	1.090	.532	.466	.000
Social context * sex * group size	1	1.570	1.570	.766	.382	.001
Error	1164	2386.112	2.050			
Total	1172	29669.000				

In addition, an independent samples *t*-test indicated both male and female targets were rated significantly *higher* on the awkward-unawkward dimension ($M = 4.81$ vs. 3.69), $t(1170) = 12.398$, $p < .001$, Cohen's $d = 0.725$, in the awkward social context condition, but significantly *lower* between the awkward and relevant conditions on the dimensions: pleasant ($M = 4.92$ vs. 5.59), $t(1170) = 9.748$, $p < .001$, Cohen's $d = 0.570$, warm ($M = 4.94$ vs. 5.24), $t(1170) = 4.377$, $p < .001$, Cohen's $d = 0.256$, and dominant ($M = 3.74$ vs. 4.60), $t(1170) = 9.506$, $p < .001$, Cohen's $d = 0.556$.

Table 3.13 provides a simplified and complete tabularised summary of the ANOVA results, for significant relationships with respect to each personality dimension, measured for both small and large group public speakers, within socially relevant and socially awkward contexts. Note the consistent pattern of personality dimension ratings across all protagonists, despite their difference in social context and group size condition. With the exception of the awkwardness dimension (which is always rated higher), socially awkward vignettes are rated low on all dimensions (save public prominence in large group public speaking). In contrast, socially relevant vignettes are rated high on all dimensions (again save the awkwardness dimension which is rated lower), excepting the public prominence rating, which is rated lower in small group public speaking.

	publicly prominent	awkward	intelligent	sexually attractive	desirable as a date	warm, pleasant, dominant
Large group public speaker						
socially relevant	↑	↓	↑	↑	↑	↑
socially awkward	↑	↑	↓	↓	↓	↓
Small group public speaker						
socially relevant	↓	↓	↑	↑	↑	↑
socially awkward	↓	↑	↓	↓	↓	↓

Table 3.13. Complete tabularised summary of the ANOVA results for significant relationships with respect to each personality dimension, measured for both small and large group public speakers, within socially relevant and socially awkward contexts. Ratings that are *higher,* relative to the alternative social context, are indicated by an upward arrow. Ratings that are *lower,* relative to the alternative social context, are indicated by downward arrow. Public prominence ratings indicate comparisons made between thelarge group public speakers and the small group public speakers.

3.4.3 Discussion

The results of Experiment III did not support the prediction that large group public speaking enhances the sexual attractiveness or dating desirability of targets who engage in this strategy. In other words, speaking in large public gatherings is not an honest signal of social status. Firstly, use of language in the context of public speaking did not succeed in making an individual appear more intelligent to the opposite sex, contradicting prediction 1. Secondly, relative to small group speaking, speaking in large groups did not increase an individual's attractiveness ratings by the opposite sex for either males or females, contradicting prediction 2. Nonetheless, the lower ratings on warmth, pleasantness and dominance, among those who spoke awkwardly, suggests there are social costs among those who do not speak eloquently, regardless of whether in the context of small or large groups.

Both sexual attractiveness and dating desirability were influenced by the sex of the target and the social context, indicating that social relevance increased the attractiveness of targets relative to those who spoke awkwardly, but only males were

significantly different. Moreover, speakers who spoke with social relevance were also rated as more intelligent, honest, pleasant, warm and dominant, as well as less awkward, relative to those who spoke awkwardly. In effect, the failure of public speaking as a mate choice cue seems at least partially explained by costly social blunders, offering support for prediction 3. These results are therefore consistent with previous studies which found that males can use group conversations as advertising 'leks' (Dunbar et al. 1997).

3.5 Experiment IV: Effect of intentionally deceptive vignettes on heterosexual attractiveness ratings

Experiments I-III did not support predictions made from the sexual selection hypothesis for human vocabulary size, poetic use of language, or oratory and public speaking (albeit there was support for the suggestion that relevant speakers were rated more attractive than awkward speakers, independent of group size). Nonetheless, it could be that more subtle aspects of conversational behavior are more important. One example of such behavior might be the skilled use of self-deprecation (Greengross and Miller, 2008) and prosocial behavior (Bornstein et al. 1996; Schlenker and Leary, 1982), which might also signal moral virtues such as humility (Greengross and Miller, 2008; Miller, 2007), as well as imperviousness to danger and social repercussions (Donath, 2008; Bliege Bird and Smith, 2005).

Indeed, it has been suggested that the high prevalence of linguistically honest information in the everyday use of language (and potential reputation-based social costs of lying and deception) may have evolved as a socially or sexually selected handicap (Bliege Bird and Smith, 2005; Donath, in press; Donath, 2008; Lachmann et al. 2001). In this light, truth-telling and self-deprecation might be a better honest signal (as a receiver-dependent cost) of humility, prosociality and social status with respect to human mate choice.

Experiment IV tested three specific predictions. First, antisocial deception will lead to lower attractiveness ratings, as a direct test of social sanctions for deception within the context of mate attraction. The specific type of antisocial deception examined here was *self-enhancement* deception (boastful exaggeration of desirable qualities). Second, prosocial deception will lead to higher attractiveness ratings, as an indirect test of humility, intelligence and prosocial behavior. The specific type of prosocial deception examined here was *self-deprecating* deception (modest understating of desirable qualities). Third, self-enhancement deception will lead to lower ratings on pleasantness, warmth and intelligence, while self-deprecation deception will lead to higher ratings on pleasantness, warmth and intelligence. In effect, self-deprecation should be an effective strategy to justify understating desirable qualities, while self-enhancement should be an ineffective strategy, due to lower ratings on various personality dimensions included in the study.

3.5.1 Methods

3.5.1.1 Participants

Participants were undergraduates, postgraduates and working professionals from a range of backgrounds recruited from AMT and included 172 females and 192 males from India (49.4%), the United States (44.7%), Macedonia (1.3%), Canada (1.1%), Romania (0.53%), Egypt (0.27%), Ethiopia (0.27%), Hungary (0.27%) and industrialised nations with available Internet access. High school education was completed by 100% of participants, with 82.89% having attained a diploma or some college experience, and 72.19% having completed a bachelor's degree or higher. All participants were native English first language speakers. Participants were restricted to the age range 18–65 years (M = 32.1, SD = 9.9).

3.5.1.2 Materials

A series of two short descriptive vignettes, each comprised of a before-and-after description of two potential romantic protagonists, was presented to half of the participants to read (version 1). A second series of two short descriptive vignettes, each comprised of a before-and-after description of the same potential romantic protagonists, but described slightly differently, was presented to the remaining half of the participants to read (version 2). Each before-and-after vignette was approximately 75 words (before description) and 285 words (after description) long and was matched for content, number of words, sentences, and overall complexity, as these factors have been shown to affect participant ratings (Bale et al. 2006).

The first set of author-written descriptive vignettes depicted a before-and-after description of a boastful 'self-enhancing' attorney and a before-and-after description of a boastful 'self-enhancing' janitor protagonist (version 1). The second set of descriptive vignettes depicted a before-and-after description of a modest 'self-deprecating' attorney and a before-and-after description of a modest 'self-deprecating' janitor protagonist (version 2) (Appendix A.3.2). The study used a between-subjects design, where each participant was presented with the set of two vignettes from either the boastful version 1 or the modest version 2 (Table 3.14). After each vignette had been read, participants answered questions about the vignette.

Table 3.14. Boastful (version 1) and modest (version 2) protagonist vignettes

Version 1: Boastful protagonist vignettes
self-enhancing attorney before-and-after description
self-enhancing janitor before-and-after description

Version 2: Modest protagonist vignettes
self-deprecating attorney before-and-after description
self-deprecating janitor before-and-after description

3.5.1.3 Procedure

Participants were first presented with an online questionnaire and asked to provide information about their age, sex, ethnicity, education, income, and whether they were native English speakers. The participants were then presented with one of the two vignette versions, with half being given the boastful vignette version 1 (self-enhancing) and the other half given modest vignette version 2 (self-deprecating). In both versions, participants were presented to read a before-and-after description of an attorney and a before-and-after description of a janitor; thus, each subject had four descriptions to read in total. In vignette version 1, female subjects first read an ostensibly truthful description of a male protagonist, followed by a second description which revealed that the protagonist had in fact deceptively *exaggerated* desirable qualities in his initial description. In vignette version 2, female subjects read the same initial description, followed by a second description which deceptively *understated* desirable qualities in his initial description. Male subjects were shown vignettes that were identical, except for respective pronouns which were modified to reflect the opposite gender. After each vignette had been read, participants answered questions about each vignette.

Subjects then rated the target on a "personality impression rating" that included the same personality dimensions described in Experiment III. Subjects were presented with instructions informing them that the experimenters are interested in their "ability to form accurate impressions of others when one has only very limited information to go on", before evaluating the protagonist in the vignette. The order of the vignettes was counterbalanced across subjects; half received the attorney before-and-after description first while the other half received the janitor before-and-after description first. The study was administered according to a between-subjects design, where each participant was presented with vignettes from either version 1 (an overstated attorney and janitor) or version 2 (an understated attorney and janitor), this being defined as the vignette's truth value.

3.5.1.4 Experimental design

Subjects were given same 11 bipolar adjectives to rate on a 7-point scale as used in Experiment III. As before, the honest-dishonest dimension constituted a manipulation check, while sexually attractive-unattractive and desirable-undesirable as a date were the central dependent variables. The remaining items were used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent evidence for the effectiveness of the experimental manipulation. The study used a 2 (sexually attractive/desirable as a date) x 2 (janitor/attorney profession) within-subjects x 2 (sex) x 2 (overstated/ understated deception condition) between-subjects factorial ANOVA.

3.5.2 Results

The data for this experiment were analysed using a substantive number of ANOVAs and *t*-tests concerning two different protagonists with each protagonist having two different deception conditions. Consequently, due to the number of permutations of character and conditions involved in this experiment, the formal statistical results as presented below are rather complicated, and perhaps, unavoidably confusing. As an aid to the reader, the results are summarised in diagrammatic form in Table 3.26 at the end of the Results section.

3.5.2.1 Manipulation check

The one-way ANOVA results clearly supported the effectiveness of the truth value manipulation, with the overstated janitor rated significantly less honest (M = 4.13, SE = .143 vs. M = 5.15, SE = .092) compared to the initial description, following revealed deception. Table 3.15 gives the results of the analysis of variance for the ratings on honesty, with truth

value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on honesty ratings, as did sex with females rated as more honest than males; the sex x truth value interaction was not significant.

Table 3.15. ANOVA results summary (dependent variable: honesty overstated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	94.392	94.392	35.063	.000	.088
Sex	1	15.991	15.991	6.026	.015	.016
Truth value * sex	1	2.877	2.877	1.084	.298	.003
Error	360	955.273	2.654			
Total	364	8869.000				

The understated janitor was also rated as significantly less honest (*M* = 5.01, *SE* = .120 vs. *M* = 5.57, *SE* = .112) compared to the initial description, following revealed deception. Table 3.16 gives the results of the analysis of variance for the ratings on honesty, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on honesty ratings, although sex did not have a significant effect; the sex x truth value interaction was also not significant.

Table 3.16. ANOVA results summary (dependent variable: honesty understated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	28.509	28.509	11.552	.001	.031
Sex	1	7.304	7.304	2.968	.086	.008
Truth value * sex	1	.059	.059	.024	.877	.000
Error	360	886.014	2.461			
Total	364	10739.000				

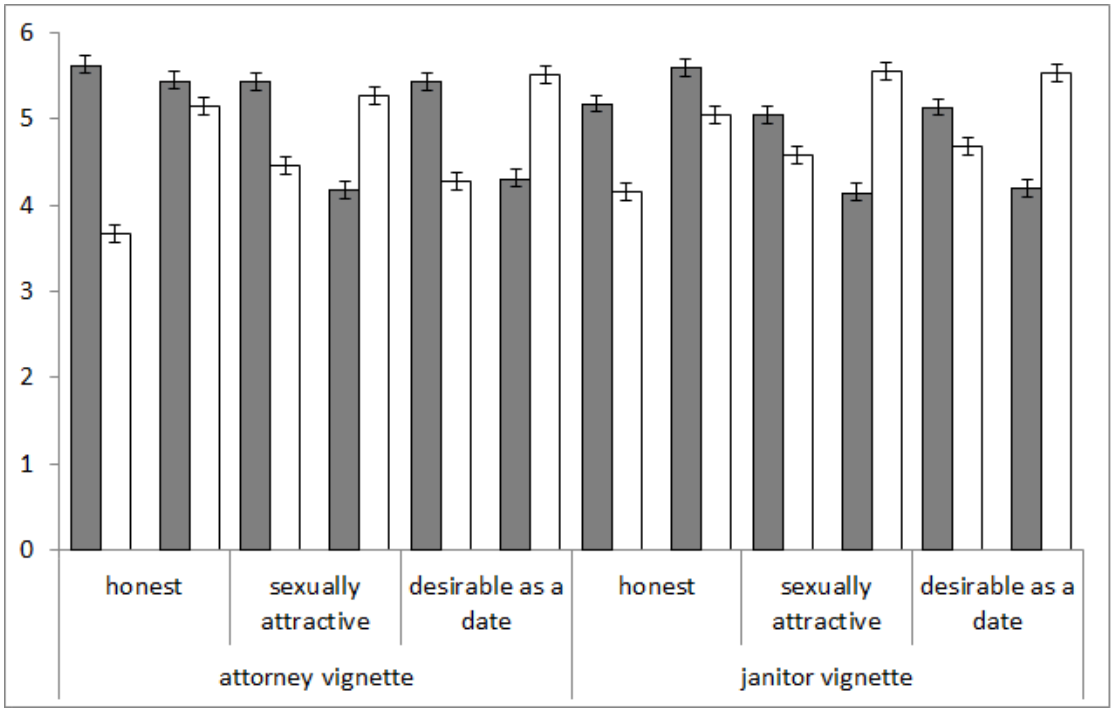

Figure 3.5. Mean values of major dependent variables for the attorney and janitor vignettes, as a function of truth value, when subjects rated the before-and-after description for each protagonist. Shaded bars depict before deception; white bars indicate after deception ratings. The first two bars for each variable indicate overstated deception; the next two adjacent bars indicate understated deception. Higher ratings on dependent variables indicate subjects perceived target persons as more honest, more sexually attractive, and more desirable as a date. Error bars denote standard error of the mean.

In similar fashion, the overstated attorney was rated significantly less honest (M = 3.64, SE = .143 vs. M = 5.60, SE = .086), compared to the initial description, following revealed deception. Table 3.17 gives the results of the analysis of variance for the ratings on honesty, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on honesty ratings, although sex did not have a significant effect; the sex x truth value interaction was also not significant.

Table 3.17. ANOVA results summary (dependent variable: honesty overstated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	348.310	348.310	133.921	.000	.270
Sex	1	8.676	8.676	3.378	.067	.009
Truth value * sex	1	7.801	7.801	3.037	.082	.008
Error	360	924.559	2.568			
Total	364	8970.000				

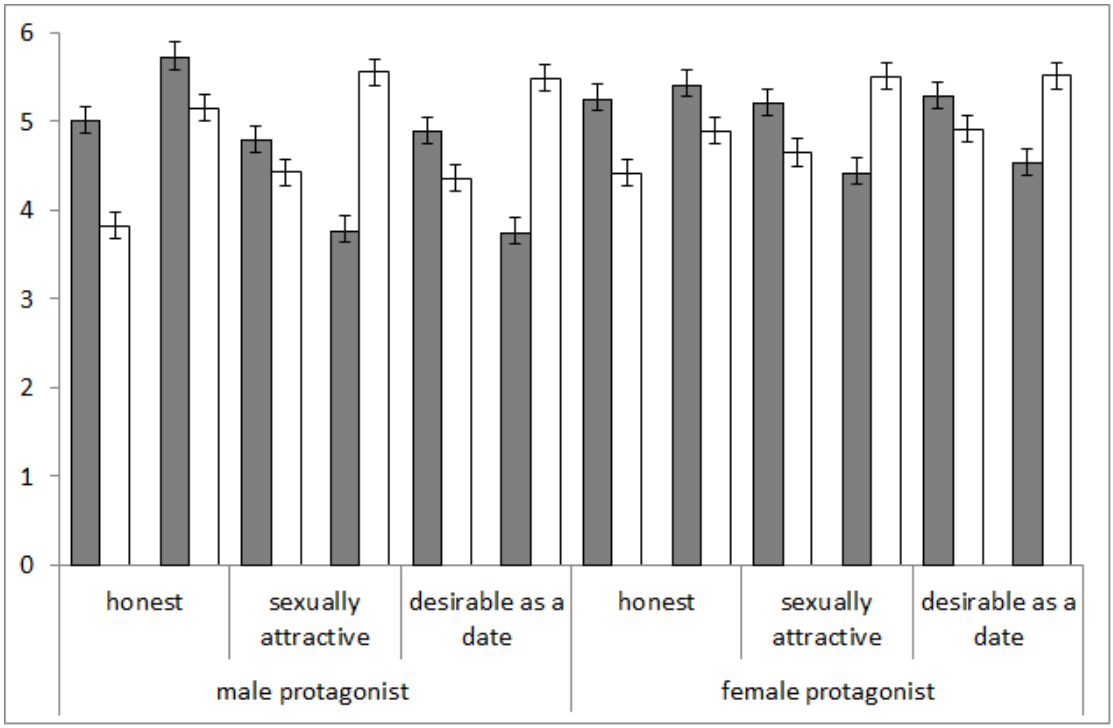

Figure 3.6. Mean values of major dependent variables for male and female targets for the janitor vignettes. Shaded bars depict before deception; white bars indicate after deception ratings. The first two bars for each dependent variable indicate overstated deception; the next two adjacent bars indicate understated deception. Higher ratings on dependent variables indicate subjects perceived target persons as more honest, sexually attractive, and desirable as a date. Error bars denote standard error of the mean.

The understated attorney was also rated significantly less honest (M = 5.12, SE = .105 vs. M = 5.42, SE = .090), compared to the initial description, following revealed deception. Table 3.18 gives the results of the analysis of variance for the ratings on honesty, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on honesty ratings, although sex did not have a significant effect; the sex x truth value interaction was also not significant.

Table 3.18. ANOVA results summary (dependent variable: honesty understated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	8.519	8.519	4.859	.028	.013
Sex	1	1.169	1.169	.666	.415	.002
Truth value * sex	1	1.589	1.589	.905	.342	.003
Error	360	631.795	1.755			
Total	364	10739.000				

In summary, all four post-deception vignette ratings were significantly lower than their corresponding pre-deception vignette ratings, indicating that deceptive vignettes decreased the honesty ratings of each protagonist relative to their ostensibly honest first impression.

3.5.2.2 Attraction measures

The understated janitor protagonist was rated significantly more sexually attractive (M = 5.53, SE = .089 vs. M = 4.13, SE = .123), compared to the initial description, following revealed deception. Table 3.19 gives the results of the analysis of variance for the ratings on sexual attraction, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on sexual attraction ratings, as did sex with females rated as more sexually attractive than males; the sex x truth value interaction was also significant.

Table 3.19. ANOVA results summary (dependent variable: sexually attractive understated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	184.422	184.422	90.996	.000	.202
Sex	1	8.198	8.198	4.045	.045	.011
Truth value * sex	1	11.252	11.252	5.552	.019	.015
Error	360	729.613	2.027			
Total	364	9496.000				

Similarly, the understated janitor protagonist was also rated significantly more desirable as a date (M = 5.51, SE = .099 vs. M = 4.18, SE = .123), compared to the initial description, following revealed deception. Table 3.20 gives the results of the analysis of variance for the ratings on dating desirability, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on dating desirability ratings, as did sex, with females rated as more desirable as a date than males; the sex x truth value interaction was significant.

Table 3.20. ANOVA results summary (dependent variable: desirable as a date understated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	165.520	165.520	76.284	.000	.175
Sex	1	15.245	15.245	7.026	.008	.019
Truth value * sex	1	13.371	13.371	6.162	.014	.017
Error	360	781.121	2.170			
Total	364	9576.000				

In summary, as shown by the means in Figures 3.5 and 3.6, deceptive vignettes *increased* both the sexual attractiveness and dating desirability of the *understated* janitor relative to their ostensibly honest first impression.

In addition, the understated janitor was rated significantly *higher* on the intelligent-unintelligent dimension (M = 5.93 vs. 4.72), $t(362) = 9.803$, p = .000, Cohen's d = 1.030, compared to the initial description, following revealed deception, as well as significantly *higher* on all of the following dimensions: pleasant (M = 5.87 vs. 5.11), $t(362) = 5.863$, p < .001, Cohen's d = 0.616, warm (M = 5.62 vs. 5.28), $t(362) = 2.870$, p < .004, Cohen's d = 0.302, and dominant (M = 4.55 vs. 3.67), $t(362= 5.333$, p < .001, Cohen's d = 0.561.

However, the overstated janitor was rated significantly less sexually attractive (M = 4.55, SE = .108 vs. M = 5.02, SE = .106), compared to the initial description, following revealed deception. Table 3.21 gives the results of the analysis of variance for the ratings on sexual attraction, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on sexual attraction ratings, as did sex with females rated as more sexually attractive than males; the sex x truth value interaction was not significant.

Table 3.21. ANOVA results summary (dependent variable: sexually attractive overstated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	19.963	19.963	9.714	.002	.026
Sex	1	9.702	9.702	4.721	.030	.013
Truth value * sex	1	.926	.926	.451	.503	.001
Error	360	739.806	2.055			
Total	364	9079.000				

Similarly, the overstated janitor protagonist was also rated significantly less desirable as a date (*M* = 4.65, *SE* = .120 vs. *M* = 5.11, *SE* = .109), compared to the initial description, following revealed deception. Table 3.22 gives the results of the analysis of variance for the ratings on dating desirability, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.6. Truth value had a significant effect on dating desirability ratings, as did sex with females rated as more desirable as a date than males; the sex x truth value interaction was not significant.

Table 3.22. ANOVA results summary (dependent variable: desirable as a date overstated janitor)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	19.760	19.760	8.354	.004	.023
Sex	1	20.147	20.147	8.517	.004	.023
Truth value * sex	1	.540	.540	.228	.633	.001
Error	360	851.556	2.365			
Total	364	9538.000				

In summary, as shown by the means in Figures 3.5 and 3.6, deceptive vignettes *decreased* both the sexual attractiveness and dating desirability of *overstated* janitor protagonists relative to their ostensibly honest first impression.

In addition, the overstated janitor was rated significantly *lower* on the intelligent-unintelligent dimension (*M* = 5.07 vs. *M* = 5.44), t(362) = 2.802, p = .005, Cohen's *d* = 0.29, compared to the initial description, following revealed deception, as well as lower on

pleasantness (M = 4.89 vs. M = 5.36), t(362) = 3.142, p < .002, Cohen's d = 0.33 and dominance (M = 4.35 vs. M = 5.03), t(362) = 4.829, p < .001, Cohen's d = 0.51, but not significantly different on warmth (M = 5.04 vs. M = 5.06), t(362) = .184, p = .854, Cohen's d = 0.02. Moreover, the overstated janitor was rated significantly *higher* on the pretentious-unpretentious dimension than the understated janitor, including both before (M = 5.15 vs. M = 3.40), t(352) = 10.347, p = .000, Cohen's d = 1.103, and after (M = 4.73 vs. M = 3.87), t(372) = 4.673, p = .000, Cohen's d = 0.485, revealed deception.

In summary, in addition to the attractiveness ratings, deceptive vignettes *decreased* the intelligence, pleasantness and dominance ratings of the overstated janitor protagonist, compared to their ostensibly honest first impression.

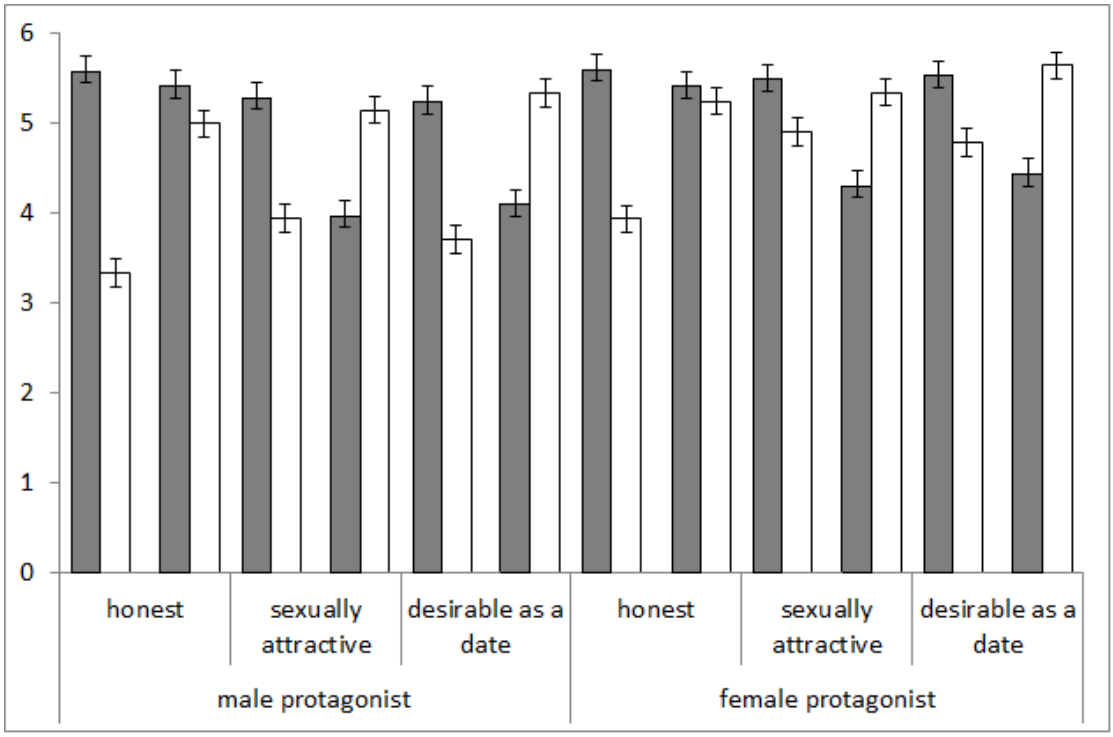

Figure 3.7. Mean values of major dependent variables for male and female targets for the attorney vignettes. Shaded bars depict before deception; white bars indicate after deception ratings. The first two bars for each variable indicate overstated deception; the next two adjacent bars indicate understated deception. Higher ratings on dependent variables indicate subjects perceived target persons as more honest, sexually attractive, and desirable as a date. Error bars denote standard error of the mean.

Consistent with the results of the overstated janitor protagonist, the overstated attorney protagonist was also rated significantly less sexually attractive (M = 4.44, SE = .120 vs. M = 5.41, SE = .088), compared to the initial description, following revealed deception. Table 3.23 gives the results of the analysis of variance for the ratings on sexual attractiveness, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on sexual attractiveness ratings, as did sex with females rated as more sexually attractive than males; the sex x truth value interaction was also significant.

Table 3.23. ANOVA results summary (dependent variable: sexually attractive overstated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	87.008	87.008	44.704	.000	.110
Sex	1	30.124	30.124	15.477	.000	.041
Truth value * sex	1	12.930	12.930	6.643	.010	.018
Error	360	700.669	1.946			
Total	364	9612.000				

Additionally, the overstated attorney protagonist was also rated significantly less desirable as a date (M = 4.26, SE = .132 vs. M = 5.41, SE = .101), compared to the initial description, following revealed deception. Table 3.24 gives the results of the analysis of variance for the ratings on dating desirability, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on dating desirability ratings, as did sex with females rated as more desirable as a date than males; the sex x truth value interaction was also significant.

Table 3.24. ANOVA results summary (dependent variable: desirable as a date overstated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	122.337	122.337	50.921	.000	.124
Sex	1	42.437	42.437	17.664	.000	.047
Truth value * sex	1	14.120	14.120	5.877	.016	.016
Error	360	864.887	2.402			
Total	364	9494.000				

In summary, as shown by the means in Figures 3.5 and 3.7, deceptive vignettes *decreased* both the sexual attractiveness and dating desirability of *overstated* attorney protagonists relative to their ostensibly honest first impression.

In addition, the overstated attorney was rated significantly *lower* on the intelligent-unintelligent dimension ($M = 5.00$ vs. $M = 5.93$), $t(362) = 6.751$, $p < .001$, Cohen's $d = 0.710$, compared to the initial description, following revealed deception, as well as lower on pleasantness ($M = 4.88$ vs. $M = 5.81$), $t(362) = 6.426$, $p < .001$, Cohen's $d = 0.68$, warmth ($M = 4.97$ vs. $M = 5.56$), $t(362) = 4.697$, $p < .001$, Cohen's $d = 0.49$ and dominance ($M = 4.34$ vs. $M = 4.77$), $t(362) = 3.099$, $p < .002$, Cohen's $d = 0.33$. Moreover, the overstated attorney was rated significantly *higher* on the pretentious-unpretentious dimension than the understated attorney, including both before ($M = 4.71$ vs. $M = 4.01$), $t(352) = 4.398$, $p < .001$, Cohen's $d = 0.469$, and after ($M = 5.04$ vs. $M = 4.32$), $t(372) = 4.409$, $p < .001$, Cohen's $d = 0.457$, revealed deception.

In summary, in addition to the attractiveness ratings, deceptive vignettes *decreased* the intelligence, pleasantness, warmth and dominance ratings of the *overstated* attorney protagonist, compared to their ostensibly honest first impression.

However, consistent with the results of the understated janitor protagonist, the understated attorney protagonist was rated significantly more sexually attractive ($M = 5.24$, $SE = .095$ vs. $M = 4.16$, $SE = .111$), compared to the initial description, following revealed

deception. Table 3.25 gives the results of the analysis of variance for the ratings on sexual attractiveness, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on sexual attractiveness ratings, although sex did not have a significant effect; the sex x truth value interaction was also not significant.

Table 3.25. ANOVA results summary (dependent variable: sexually attractive understated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	107.355	107.355	56.121	.000	.135
Sex	1	6.735	6.735	3.521	.061	.010
Truth value * sex	1	.453	.453	.237	.627	.001
Error	360	688.647	1.913			
Total	364	8892.000				

Similarly, the understated attorney protagonist was rated significantly more desirable as a date ($M = 5.49$, $SE = .092$ vs. $M = 4.29$, $SE = .118$), compared to the initial description, following revealed deception. Table 3.26 gives the results of the analysis of variance for the ratings on dating desirability, with truth value and sex as the independent variables, for the results shown in Figures 3.5 and 3.7. Truth value had a significant effect on dating desirability ratings, as did sex with females rated as more desirable as a date than males; the sex x truth value interaction was not significant.

Table 3.26. ANOVA results summary (dependent variable: desirable as a date understated attorney)

Source of Variation	df	Sum of Squares	Mean Squares	F-ratio	Significance Level	η^2
Truth value	1	132.002	132.002	65.881	.000	.155
Sex	1	9.734	9.734	4.858	.028	.013
Truth value * sex	1	.010	.010	.005	.943	.000
Error	360	721.309	2.004			
Total	364	9626.000				

In summary, as shown by the means in Figures 3.5 and 3.7, deceptive vignettes *increased* both the sexual attractiveness and dating desirability of *understated* attorney protagonists relative to their ostensibly honest first impression.

In addition, the understated attorney was rated significantly *higher* on the intelligent-unintelligent dimension ($M = 5.73$ vs. $M = 4.77$), $t(362) = 7.275$, $p = .000$, Cohen's $d = 0.765$, compared to the initial description, following revealed deception,
as well as higher on pleasantness ($M = 5.70$ vs. $M = 4.72$), $t(362) = 7.583$, $p < .001$, Cohen's $d = 0.80$, warmth ($M = 5.36$ vs. $M = 5.02$), $t(362) = 2.748$, $p < .006$, Cohen's $d = 0.29$ and dominance ($M = 4.78$ vs. $M = 4.05$), $t(362) = 4.786$, $p < .001$, Cohen's $d = 0.50$. Moreover, the understated attorney was rated significantly *lower* on the pretentious-unpretentious dimension than the overstated attorney, including both before ($M = 4.01$ vs. $M = 4.71$), $t(352) = 4.399$, $p < .001$, Cohen's $d = 0.47$, and after ($M = 4.32$ vs. $M = 5.04$), $t(372) = 4.409$, $p < .001$, Cohen's $d = 0.46$, revealed deception.

In summary, in addition to the attractiveness ratings, deceptive vignettes *increased* the intelligence, pleasantness, warmth and dominance ratings of the *understated* attorney protagonist, compared to their ostensibly honest first impression.

Table 3.27 provides a simplified and complete tabularised summary of the ANOVA results for significant relationships with respect to each personality dimension measured for all four protagonists. Note the consistent pattern of personality dimension ratings across both protagonists, despite their difference in social standing and condition. With the exception of the honesty dimension, which is always rated lower for all protagonists, overstated vignettes are rated low on all dimensions, save pretentiousness, while understated vignettes are rated high on all dimensions, again save pretentiousness, which is rated lower.

	honest	sexually attractive	desirable as a date	intelligent	warm, pleasant, dominant	pretentious
self-enhancing (overstated) janitor	↓	↓	↓	↓	↓ (except warm)	↑
self-enhancing (overstated) attorney	↓	↓	↓	↓	↓	↑
self-deprecating (understated) janitor	↓	↑	↑	↑	↑	↓
self-deprecating (understated) attorney	↓	↑	↑	↑	↑	↓

Table 3.27. Complete tabularised summary of the ANOVA results for significant relationships with respect to each personality dimension, measured for all four protagonists. Ratings that are *higher* relative to the initial description (after participants had read the post-deception vignette) are indicated by an upward arrow. Ratings that are *lower* relative to the initial description (after participants had read the post-deception vignette) are indicated by downward arrow. Pretentiousness ratings indicate comparisons made between the overstated protagonist relative to the understated protagonist.

3.5.3 Discussion

The results of Experiment IV supported the hypothesis that the high prevalence of linguistically honest information in the everyday use of language (and potential reputation-based social costs of lying and deception) may have evolved as a socially or sexually selected handicap. In other words, truth-telling and self-deprecation may be an honest signal (as a receiver-dependent cost) of humility, prosociality and social status with respect to human mate choice.

Firstly, *self-enhancement* deception (as a specific type of antisocial deception) was predicted to be an ineffective strategy. Consistent with prediction 1, the exaggerated attorney and janitor vignettes were both rated as less attractive following *overstated* revealed deception (in comparison to the initial description), indicating that lies which inflated the traits of a potential romantic interest were costly in terms of lower attractiveness ratings.

Secondly, *self-deprecation* deception (as a specific type of prosocial deception) was predicted be an effective strategy to justify understating desirable qualities, when later revealed as an actually attractive mate prospect (in effect, revealing a humble temperament and general imperviousness to social repercussions in the context of mate choice). Consistent with prediction 2, the self-deprecating janitor and attorney vignettes were both rated as more attractive, following *understated* revealed deception, indicating that lies which 'under sold' the traits of a potential romantic interest were beneficial in terms of attractiveness ratings.

Thirdly, self-enhancement deception led to lower ratings on pleasantness, warmth and intelligence, and higher ratings on pretentiousness, while self-deprecation deception led to higher ratings on pleasantness, warmth and intelligence, and lower ratings on pretentiousness, confirming prediction 3. In conclusion, self-deprecation was an effective strategy to justify understating desirable qualities, while self-enhancement was shown to be an ineffective strategy, due to lower ratings on various personality dimensions included in the study.

These results are consistent with previous studies which have shown that self-deprecation garners increased attractiveness ratings in the context of social displays of humor (Greengross and Miller, 2008). Further, inspection of the results for intelligence, pleasantness, warmth and dominance ratings reveal that the understated janitor and attorney vignettes were also rated higher, in comparison to the overstated janitor and attorney vignettes, in spite of the fact the janitor and attorney vignettes involved a certain degree of self-deprecating deception. Lastly, inspection of the means for pretentiousness suggest that the understated janitor and attorney vignettes were rated significantly *lower* in comparison to the overstated vignettes, likely due largely to the fact the understated vignettes involved a certain degree of self-deprecating deception (Bornstein et al. 1996; Greengross and Miller, 2008; Schlenker and Leary, 1982).

3.5.4 General discussion

3.5.4.1 Summary of results

The results from Experiments I and II did not support the prediction that use of low-frequency or poetic vocabulary enhances the sexual attractiveness of either males or females in the way predicted by the sexual selection hypothesis (Miller, 2000).

In contrast to the predictions of the sexual selection hypothesis, the direction of the effect was opposite to that predicted, for both experiments. In Experiment I, targets that used normal-frequency vocabulary were rated as significantly more sexually attractive than targets that used low-frequency vocabulary, but only significantly so in the case of males. In Experiment II, targets that used normal vocabulary were rated as significantly more desirable as a date than targets that used poetic vocabulary, but only significantly so in the case of females.

In addition, in Experiment I, males that used normal-frequency vocabulary were also rated higher on the dimensions of pleasantness, masculinity, and toughness, suggesting that males that used low-frequency vocabulary were perceived to be less masculine, pleasant and tough than those who used every day common language. In Experiment II, female targets that used normal vocabulary were also rated higher on the pleasant-unpleasant dimension, demonstrating that female targets that used poetic vocabulary were perceived to be less pleasant than those who used every day common language.

The results of Experiment III did not support the prediction that charisma, as reflected in being able to speak at large public gatherings, is attractive to the opposite sex (as suggested by Burling, 1986). However, the significant sex x social context interaction with respect to both sexual attractiveness and dating desirability indicates that social relevance does increase the attractiveness of a target relative to those who speak awkwardly, but only males were significantly different. Moreover, speakers who spoke with social relevance were rated as

more intelligent in comparison to those who spoke awkwardly. These results thus provide some support for predictions from evolutionary leadership theory, which emphasises the importance of traits such as extraversion, assertiveness, and verbal and social intelligence (Van Vugt, 2006), as well as the costly risk of social embarrassment and unfavourable social reputation in an information-based society (Donath, 2008; McCracken, 1998; Thornton, 1996).

The results of Experiment IV supported the prediction that revealed courtship deception leads to significant modulations in attractiveness ratings, for both males and females. The overstated attorney and janitor vignettes were rated as less attractive following revealed *self-enhancing* deception (in comparison to the initial description), indicating that lies which inflated the traits of a potential romantic interest were costly in terms of attractiveness ratings. On the other hand, the understated janitor and attorney vignettes were rated as more attractive, following revealed *self-deprecating* deception, indicating that lies which 'under sold' the traits of a potential romantic interest were beneficial in terms of attractiveness ratings.

Inspection of the ratings for pleasantness, intelligence, warmth and dominance reveal the understated janitor and attorney vignettes were rated higher in comparison to the overstated janitor and attorney vignettes in spite of the fact (or perhaps more likely due to the fact) that the janitor and attorney vignettes involved a certain degree of self-deprecating deception. Perhaps most notably, the respective professions of each protagonist did not seem to dramatically affect attractiveness ratings; both attorneys and janitors were negatively affected by boastful deception, while positively affected by self-deprecation.

3.5.4.2 Implications

Experiments I-III found little support for the sexual selection hypothesis for human language function: none of the language manipulations made protagonists attractive as mates, even though most of the manipulations seemed to have worked in terms of making protagonists seem more intelligent. In contrast, Experiment IV found convincing evidence that self-deprecation might be an attractive trait in mate choice.

The results from Experiment IV, however, are consistent with previous studies which have shown increased attractiveness ratings of self-deprecating protagonists in the context of social displays of humor (Greengross and Miller, 2008). As such, self-deprecation (whether in the context of humor or otherwise), may be a way of temporarily faking low-status personality traits in order to better highlight the discrepancy between the faked traits and the traits actually required to convey high status information (Greengross and Miller, 2008), which are furthermore likely mediated via contrast effects (Moskowitz, 2005). Nonetheless, given that pretentiousness ratings in Experiment IV were significantly lower (indicating a more modest and therefore likable temperament) (Bornstein et al. 1996; Schlenker and Leary, 1982), and intelligence ratings significantly higher for the self-deprecating janitor and attorney vignettes, it might still be plausible to suggest that honest linguistic communication evolved, at least in part, to function as a fitness indicator.

In effect, skilled use of self-deprecation may be a kind of social risk-taking (as a receiver-dependent cost), as an honest signal not only of general intelligence (Greengross and Miller, 2008) and prosocial behavior (Bornstein et al. 1996; Schlenker and Leary, 1982), but also of moral virtues such as humility (Greengross and Miller, 2008; Miller, 2007), while concurrently signalling imperviousness to danger and potentially costly social sanctions (Bliege Bird and Smith, 2005; Donath, 2008; Lachmann et al. 2001). The results here are also consistent with computational modelling experiments which have shown that antisocial lying

causes social networks to become increasingly fragmented, whereas prosocial 'white' lies are often beneficial in smoothing the flow of social interactions so as to facilitate larger, more integrated networks (Iñiguez et al. 2014; Rauwolf, Mitchell and Bryson, 2014). Self-deprecation may be an example of the kind of prosocial lies that Iñiguez et al. (2014) found to be highly beneficial in terms of maintaining social network cohesion. More broadly, perhaps, these results also highlight previous research on the psychology of deception and self-deception (Ekman, 2009; Trivers, 2014), and may help to provide an important experimental foundation for future studies in this area.

In summary, I find suggestive evidence from the results of Experiment IV that the high prevalence of linguistically honest information in the everyday use of language (and potential reputation-based social costs of lying and deception) may have evolved as a socially or sexually selected handicap. In other words, truth-telling and self-deprecation may be an honest signal (as a receiver-dependent cost) of humility, prosociality and social status with respect to human mate choice. Although the data here were collected within the context of mate choice, it seems intuitively reasonable to suppose these results might also apply to communicative honest signalling in social contexts more broadly, such as those that involve status, reputation effects, and prosocial behavior. In conclusion, given the high prevalence of linguistically honest information in the everyday use of language, and possible reputation-based social costs of lying and deception (Bliege Bird and Smith, 2005; Donath, in press; Donath, 2008; Lachmann et al. 2001), self-deprecation may have evolved as a socially or sexually selected handicap (Greengross and Miller, 2007).

3.5.4.3 Limitations

The experiments in Chapter III had several limitations in the design that should be clarified. For example, in Experiments I and II, although the manipulation check suggests

targets that used low-frequency vocabulary were successful in conveying greater intelligence (albeit not in the poetic vocabulary experiment) than those that used normal-frequency vocabulary, it is possible that the targets may have simultaneously displayed other traits in the low-frequency (or poetic) vocabulary condition that conveyed less attractive qualities. A possible solution to this potential limitation might be to use several different carefully-controlled vignettes, pre-rated for attractiveness ratings (prior to the inclusion of low-frequency or poetic vocabulary words), in a variety of suitable contexts.

Second, the author-written vignettes which tested for vocabulary attractiveness in Experiments I-II also examined a very narrow range of low-frequency vocabulary. As such, it is certainly conceivable that some of the participants found the particular low-frequency words used in these particular series of vignettes as unattractive, pretentious, or in some other way unappealing, or even that these were words they simply didn't understand. A possible solution to this potential limitation might be to use several different vignettes, identical in all respects, apart from several different low-frequency vocabulary words.

Indeed, the lower ratings on masculinity, pleasantness, and toughness in the low frequency vocabulary experiment (Experiment I), and lower ratings on pleasantness in the poetic vocabulary experiment (Experiment II), can be interpreted as supporting this suggestion. Therefore, it is still conceivable that targets who are able to use either low-frequency or poetic vocabulary without simultaneously communicating negative qualities might be more attractive than targets using normal-frequency vocabulary. This might include not using low-frequency words that are likely to be unfamiliar to the person being addressed, injecting humor in the conversation, overlaying spoken words with appropriately friendly visual signals, such as smiles, etc – none of which were of course addressed in the present study.

Nevertheless, it is critically important to note that in both experiments, the results were the reverse of those predicted by the sexual selection hypothesis: subjects rated low-frequency and poetic vignettes as *less* desirable. It thus seems unlikely that removing any confound would have a big enough effect to reverse these results. Given that, we should perhaps confidently conclude that vocabulary size, whether measured in terms of low-frequency or poetic vocabulary, did not evolve as an honest signal of human intelligence (as a receiver-independent cost) so as to be attractive to prospective mates in the way proposed by the sexual selection hypothesis (Miller, 2000; Rosenberg and Tunney, 2008).

In Experiment III, an important limitation of the results is that the public speaking vignettes did not include depictions of public speakers making grandiose or visionary proclamations: because of this, the current results may not necessarily provide a comprehensive test of the charismatic leadership theory, defined as having both an appealing vision and a willingness to take personal risks to achieve that vision (Robbins and Judge, 2013). Thus, although the results here are consistent with the hazards of social risk-taking hypothesis, exemplified in the lower ratings associated with speaking awkwardly in public gatherings, future studies may need to examine the potential social and sexual benefits from having an appealing vision in the context of charismatic leadership. Second, similar to Experiments I-II, it is possible that the vignette protagonists may have simultaneously displayed other traits that conveyed less attractive qualities. As before, a possible solution to this potential limitation might be to use several different carefully-controlled vignettes, pre-rated for attractiveness ratings, in a variety of auspicious contexts.

Lastly, Experiment IV likely also had several important limitations. First, the author-written vignettes which tested for attractiveness ratings examined a very narrow range of protagonists; namely an attorney and a janitor. As such, it is certainly conceivable that participants found the particular protagonists in these series of vignettes as somehow

particularly endearing or attractive, in spite of their deceptive characteristics. A possible solution to this potential limitation might be to use several different vignettes, with several additional protagonists, in a variety of different contexts. Second, despite the findings which ostensibly supported honest signalling theory, an important limitation of the results (also applicable to Experiments I-III) may have been related to the somewhat artificial testing methodology used for this experiment. More specifically, the results found here were based on self-report measures of opposite sex attractiveness ratings in the context of vignette-based narratives. Consequently, as the attractiveness ratings were based on self-report measures, there may have been some element of subjective bias introduced into these findings, partially due to these self-report measures.

In summary, the results found may have limited generalizability to actual situations which involve real-world people. Indeed, the difference between what individuals say and what individuals actually do in a given context is often considerable. One way of overcoming this subjectivity, for example, might be to use automated pupilometer technology to measure variations in pupil size, since changes in pupil size often index interest, and particularly sexual interest. A second possibility might be to monitor real-world courtship interactions, in areas such as coffee shops, nightclubs or restaurants. In conclusion, future studies should perhaps attempt to test attractiveness ratings in more naturalistic environments and/or using advanced technologies.

Chapter IV
The evolution of grammar in human language: Mentalising performance predicts recursive clause task performance

4.1 Introduction

Given that the results from Chapter II indicated that language did not seem to aid or facilitate group cooperative problem-solving and that the results from Chapter III seemed to support an honest signalling theory approach to language function (in the context of mate choice and social reputation), I elected to test a final hypothesis for the evolution of language form. Indeed, as deceptive signalling of the kind tested in Chapter III necessarily depends on understanding how signals would be interpreted by recipients, the final data chapter elected to evaluate a key mechanism likely to be important in this process, namely mentalising - the ability to read others' minds.

With respect to language evolution, no current consensus exists within evolutionary linguistics as to how and why human language acquired grammatical structure (Hurford, 2011). However, several primatologists and evolutionary psychologists have recently suggested the adult capacity for higher-order intentionality - embedded mindreading beyond theory of mind and thus an intensely social form of cognition - may have played a critical role in the evolution of language, including especially the linguistic capacity for recursive syntax comprehension and production (Cheney and Seyfarth, 2007; Dunbar, 1998; Dunbar, 2009). Chapter IV therefore takes an experimental approach to the problem of whether recursive syntax evolved as an integral feature of human language through a cognitive bootstrapping effect of higher-order intentionality.

Theory of mind, also known as mentalising, metarepresentation, second-order intentionality or mindreading, is the ability to attribute and reflect on the mental states of

others (Premack and Woodruff, 1978; Wellman, Cross and Watson, 2001; Wimmer and Perner, 1983). Developmental psychologists have found that fully developed theory of mind requires several developmental milestones beginning from rudimentary aspects of social cognition to the most sophisticated aspects of theory of mind. For example, during the period between 6 and 15 months, child development is characterised by aspects such as shared attention, directed eye gaze, imitation and monitoring the intentions of others (de Villiers, 2007; Gergely, Bekkering and Kiraly, 2002; Tomasello, 2009). This stage is typically followed, from 15 months to 3.5 years, by more complex forms of social cognition such as understanding desires, simple scripted emotions, and pretence (de Villiers, 2007; Onishi et al. 2007). From approximately 3.5 to 4 years, children begin to understand the relationship of seeing to knowing and the concept of conflicting emotions (de Villiers, 2007; O'Neill and Gopnik, 1991). Finally, from 4 to 5 years, children begin to fully understand false beliefs, emotions and desires based on beliefs (de Villiers, 2007; Wellman, Cross and Watson, 2001). Accordingly, theory of mind is widely held to be critical to the development of the kinds of advanced social cognitions that characterise humans, thus forming the core part of what it means to be socially human (Barrett, Dunbar and Lycett, 2002; Call and Tomasello, 2008).

Importantly, a number of investigators have noted a robust interaction exists between child language development and children's understanding of second-order intentionality. In general, the finding is that language development and theory of mind are relatively strongly related, even when language measures are taken one or two years before children start mastering false belief tasks around age 5, suggesting language acquisition may be more fundamental than theory of mind (Astington and Jenkins, 1999; Dunn et al. 1991; Farrar and Maag, 2002; Gale et al. 1996; de Villiers and de Villiers, 2000; Watson, Painter and Bornstein, 2002). However, more recent investigations using nonverbal spontaneous-response tasks suggest the ability for a representational theory of mind is present in

prelinguistic infants as young as 15-, 13- and even 10-months-old, suggesting that theory of mind may itself be fundamental for the acquisition of language (Baillargeon, Scott and He, 2010; Luo, 2011; Onishi and Baillargeon, 2005; Surian, Caldi and Sperber, 2007).

Accordingly, many have argued that no solid basis currently exists for strong conclusions concerning the direction of influence between language acquisition and theory of mind (de Villiers, 2007; Wellman et al. 2001). However, at least one influential proposal suggests the conceptual developments of early theory of mind may form an essential basis for helping to fix at least word reference (de Villiers, 2007). After age 5, mastery of grammar has been highlighted as a potential representational tool to enable theory of mind, although it remains debated whether the more important factor is general syntactic development or more specific aspects of grammar, such as syntactic complementation (Apperly, Samson and Humphreys, 2009; de Villiers, 2007). In light of these and other recent developments, some have offered the tentative conclusion that the causal interface between language acquisition and theory of mind is bidirectional (de Villiers, 2007; Miller, 2009; Milligan et al. 2007).

Although the ontogeny of theory of mind has been extensively studied over the past few decades, only recently have we begun to understand more concerning the limits of human mentalising ability in adults. For example, several studies have shown that the limits of mentalising ability for normal adults are consistently placed around fifth-order intentionality (i.e. I *believe* that you *suppose* that I *imagine* that you *want* me to *believe* that...) (Kinderman et al. 1998; Stiller and Dunbar, 2007). Further, adult ability develops over a period of time between age 5, when children first acquire theory of mind or second-order intentionality, and early adolescence, when they finally acquire fifth-order adult competency (Henzi et al. 2007). Moreover, beyond the fact that higher-order intentionality abilities have not yet been acquired in children under the age of 5, research on adult theory of mind and higher-order intentionality should be of interest to developmental psychologists for at least

two additional reasons. First, an account of the mature adult system is necessary to know when development is complete. Even after children pass developmentally sensitive theory of mind tasks their abilities are often slower and less flexible than those of adults (Dumontheil et al. 2010; Keulers et al. 2010). Second, an account of the mature adult system provides critical information for understanding why researchers observe developmental relationships between theory of mind and other abilities such as language and executive function (Apperly, Samson and Humphreys, 2009).

Indeed, recent studies suggest that impaired executive function has multiple roles in mature adult theory of mind (Apperly, Samson and Humphreys, 2009), but that severely impaired grammar can leave theory of mind functionally intact (Apperly, Samson and Humphreys, 2009; Varley and Siegal, 2000; Willems et al. 2011). For example, several studies of patients who have significantly impaired grammar on standard language tests, following left hemisphere lesions, nevertheless perform well on nonverbal theory of mind tasks (Apperly, Samson and Humphreys, 2009; Varley and Siegal, 2000; Willems et al. 2011). In other words, grammar in general, and embedded complement clauses in particular, do not seem necessary for mature adult theory of mind competence. In sum, it may be the case that grammar is not an essential part of the mature adult theory of mind capacity, but may, like temporary building scaffolding, be a necessary condition for its successful construction during child language acquisition (Apperly, Samson and Humphreys, 2009).

As such, while most attention has historically been drawn to the study of child language acquisition, more recent studies have pointed to a significant gap in our understanding of the relation between language and adult higher-order intentionality. Moreover, several primatologists and evolutionary psychologists have recently suggested the adult capacity for higher-order intentionality may have played a critical role in the evolution of language, including especially the ability for recursive syntax comprehension and

production (Cheney and Seyfarth, 2007; Dunbar, 1998; Dunbar, 2009). Indeed, mentalising capacity can be understood in terms of intentional states: an individual's understanding of states of mind, typically exemplified by the use of words like *believe*, *intend*, *suppose*, *think*, et cetera. Intentionality defined in this way forms a naturally recursive hierarchy which corresponds to increasingly embedded mindreading (i.e. I *suppose* that you *intend* that I *believe* that you *want* me to *understand* that...). Representing another's mental state is thus inherently, automatically recursive, and orders of intentionality beyond second-order are even more recursive.

In accord with this argument, it is reasonable to assume that when natural selection favored higher-order intentionality, it necessarily also favored recursive thinking, as it is logically necessary to represent the content of another's mental state within the embedded structure of one's own mental state. According to this hypothesis, recursive thinking became the necessary cognitive scaffolding upon which language and recursive syntax could have potentially later bootstrapped (Cheney and Seyfarth, 2007; Dunbar, 1998; Dunbar, 2009 and see Carey, 2004; Carey, 2009; Gentner and Christie, 2010 for further discussion of cognitive bootstrapping). Of course, modern languages exhibit recursion, often apart from whether or not the sentence contains mental state attribution. However, this does not belie the possibility that spoken language, and the relevant psycholinguistic software, were originally adapted for processing social information, and only later co-opted for also processing instrumental information about the environment. Accordingly, as no empirical studies to date have yet examined how higher-order intentionality capacities beyond second-order may influence the acquisition of recursive syntax comprehension, I elected to test this relation in normal adults.

If this hypothesis is correct, then three general predictions follow with respect to adult competence on intentionality, recursive syntax and basic social memory performance. First, intentionality performance and recursive syntax performance should be highly correlated as

evidence of a bootstrapping effect. Second, simple fact-based socially contextualised memory tasks should be more efficient and cognitively less demanding than either equivalent fact-based mentalising or recursive syntax problem-solving. Indeed, in accord with the second general prediction, several recent studies have shown that information about socially intense third-party social relationships is better remembered (Redhead and Dunbar, 2013), as well as transmitted in greater quantity and accuracy than non-social information (Mesoudi et al. 2006), suggesting human cultural communication may be biased for social information transmission. Third, recursive syntax task problem-solving should be more efficient and cognitively less demanding than equivalent intentionality task problem-solving. In accord with this third general prediction, several studies have demonstrated that intentionality is restricted by strict cognitive limits at around fifth-order (Kinderman et al. 1998; Stiller and Dunbar, 2007), involving significant neurological resources and functional connectivity (Klapwijk et al. 2013; Lewis et al. 2011; Powell et al. 2010) and despite full adult competency, often fails to be reliably deployed in every social context (Keysar, Lin and Barr, 2003). However, to date, no studies have yet tested whether socially contextualised cognitive problem-solving tasks are more easily completed when involving recursive syntax comprehension over higher-order intentionality comprehension or whether a significant correlation exists in overall performance.

In the current study, I test the hypothesis that performance on social non-mentalising recursive syntax comprehension tasks equals or exceeds performance on social intentionality comprehension tasks. The socio-cognitive task involved a previously developed (Stiller and Dunbar, 2007; Powell et al. 2010) set of short vignettes modelled on the classic Sally-Ann tasks used to test for theory of mind in young children; the vignettes describe short social events, memory for which is tested with a set of questions that index both mentalising competence at different levels and short-term memory, rather than long-term memory

(although this might in fact be more important for longer term relationships). As a control, a similar index of basic memory task performance was given measuring short-term memory of non-mentalising non-embedded social information; in other words, sentences that did not include either mindreading or recursive syntax statements. A third social non-mentalising recursive sentence task presented an index of recursive syntax comprehension performance, involving the same short vignettes.

4.2 Methods

4.2.1 Participants

A questionnaire was distributed to participants via the online survey and participant recruitment system Amazon Mechanical Turk (AMT), an Internet application that provides instant access to thousands of potential participants for questionnaire-based psychology experiments (among various other cognitive tasks) for a small monetary payment. Several studies have demonstrated the validity of online surveys and shown them to be as reliable as traditional paper-and-pencil laboratory questionnaires (Buchanan and Smith, 1999; Epstein et al. 2001; Gosling et al. 2004; Metzger et al. 2003; Preckel and Thiemann, 2003; Salgado and Moscoso, 2003; Smith and Leigh, 1997). The study was advertised as being about "memory and perspective-taking." All responses were anonymous, voluntary, and restricted to individuals over 18 years of age. All participants provided informed consent before participation and were paid $0.50 for completing the 20 minute survey.

In total, 210 females and 204 male undergraduates, postgraduates and working professionals from a range of backgrounds from India (59.18 %), North America (35.51 %), the Middle East (1.45 %), Macedonia (1.21 %), Australia (0.48 %), the United Kingdom (0.48 %), and other industrialised nations with available Internet access completed the task.

High school education was completed by 98.8 % of participants, with 94.2 % having attained a diploma or some college experience, and 72.0 % having completed a bachelor's degree or higher. All participants were native English first language speakers. To avoid developmental effects and confounds due to declining social engagement in old age, participants were restricted to the age range 18–65 years (M = 33.6, SD = 11.1).

4.2.2 Materials

The Imposing Memory Task (IMT) was designed by Kinderman, Dunbar and Bentall (1998) to assess mentalising ability in adults. Additions and changes to the IMT were then undertaken by Stiller and Dunbar (2007) and Lewis et al. (2011). The IMT consists of a series of stories involving social situations that require the listener to understand the perspective and intentions of the actors. A series of true/false mentalising and memory questions test subjects' memory for the events described in the stories. The mentalising questions require increasingly complex metarepresentational understanding of a given character's perspective of a social situation, from first order intentionality to as high as ninth order intentionality (Stiller and Dunbar, 2007). The IMT has been used as a measure of theory of mind in a number of studies of both normal and clinical adult populations (Blackshaw et al. 2001; Kerr, Dunbar and Bentall, 2003; Kinderman, Dunbar and Bentall, 1998; Lewis et al. 2011; Paal and Bereczkei, 2007; Powell et al. 2010; Randall et al. 2003; Taylor and Kinderman, 2002). It has also been used on typically developing children as a way of generating a more detailed developmental trajectory of children's increasing metarepresentational capacities, and enabling direct comparisons across preschoolers and preadolescents (Henzi et al. 2007).

4.2.3 Procedure

Participants were first asked to provide information about their age, sex, nationality, education, income, whether they were native English speakers, and to give informed consent. Subjects were asked to read a series of three short stories (each approximately 200 words long) depicting a social situation. After each story had been read, participants answered a series of true/false questions about the story (see Appendix A.4.1). For each story, subjects answered a randomly ordered series of three different types of socially situated fact-based memory questions: 1) levels 1-9 intentionality questions, 2) levels 1-9 embedded clause structure questions, and 3) levels 1-9 sequential clause structure questions. Each question consisted of a statement and the participant was asked to decide whether the statement was in fact true or false. Finally, participants were debriefed on the full aims of the study.

4.2.4 Experimental design

The design of the stories and questions used in the present study were based on Stiller and Dunbar (2007) and Kinderman et al. (1998), except that all question series contained between one to nine levels of intentionality (included to ensure that all participants reached the upper limit on their performance). The stories were slightly modified to eliminate unclear or ambiguous wording. In addition, since not all the original stories had the full range of questions, new questions were constructed (including complete sets of recursive syntax, memory and sixth through ninth level intentionality questions) and the others refined to eliminate ambiguities in the originals and improve overall readability.

The first series of questions designed for the current study were composed of recursive syntax true/false statements. Indeed, few empirical studies to date have specifically targeted syntactic comprehension in children (de Villiers and Pyers, 2002; Huttenlocher et al. 2002) and even fewer have specifically targeted recursive syntax comprehension in adults

(but see Apperly, Samson and Humphreys, 2009 on embedded complement clauses in adults as one recent exception). Therefore, embedded syntactic sentences were developed for the current study, which varied in their levels of recursive syntax. The level of complexity for each embedded syntax question was calculated according to the number of embedded clauses for each sentence. For instance, the statement: *Betty is a roommate of Brian,$_1$ who walked into the bedroom of Frank,$_2$ who received £10 from Brian,$_3$ who likes lettuce and tomatoes for a salad,$_4$ that Betty and Brian helped to prepare,$_5$* has five embedded clauses (subscripts count the number of embedded clauses). The full set of all three stories, and subsequent embedded syntactic sentence questions, are provided as supplementary information in Appendix A.4.1.

The next series of questions designed for the current study were composed of social memory true/false statements. The social memory questions served as a control for the subject's comprehension of basic sentences and understanding of the questioning procedure. In addition, they also provided a more general index of basic memory capacity. For these purposes, I opted for the simplest possible measure of memory capacity (i.e. short-term memory for social facts) in order to differentiate as clearly as possible between mentalising ability, complex recursion, and basic memory capacity (with recursive complexity being held constant). As such, the level of complexity for each social memory question was calculated according to the number of simple clauses for each sentence. For instance, the statement: *The friends share a house,$_1$ but Frank is lazy,$_2$ so Brian gives him £10,$_3$ but this makes Frank confused,$_4$ so Brian writes a list,$_5$* has five simple clauses (subscripts count the number of simple clauses). The full set of all three stories, and subsequent social memory sentence questions, are provided as supplementary information in Appendix A.4.1.

The last series of questions designed for the current study were composed of higher-order intentionality true/false statements. Indeed, previous studies have found that for most

normal healthy adults the highest level of mindreading competency is around fifth-order intentionality (Stiller and Dunbar, 2007). Beyond this level of intentionality, most adults fail IMT questions (Kinderman et al. 1998; Stiller and Dunbar, 2007). The level of complexity for each higher-order intentionality question was calculated according to the number of intentions or belief states for each sentence. For instance, the statement: *Simon believes$_1$ that Martin thinks$_2$ that Charlotte supposes$_3$ that Jane knows$_4$ that Simon thinks$_5$ they had a lovely date at the restaurant*, has five levels of intentionality (subscripts count the number of metarepresentations). Moreover, since not all of the original stories had the full range of questions, new intentionality questions were constructed, including complete sets of sixth through ninth level intentionality questions. Therefore, the most complex mentalising questions contained up to nine orders of perspective-taking (including the participant's own perspective, defined as level one). The full set of all three stories, and subsequent higher-order intentionality sentence questions, are provided as supplementary information in Appendix A.4.1.

In summary, each story was followed by a randomly ordered series of three different types of socially situated fact-based memory questions: 1) levels 1-9 intentionality questions, 2) levels 1-9 embedded clause structure questions (containing increasingly complex recursive syntax, but having no aspects of mindreading), and 3) levels 1-9 sequential clause structure questions (containing consecutive single clauses with no syntactic embedding or mindreading). The sequential clause structure questions were intended to target mainly short-term memory and served as a control for the other two series of questions. Each question consisted of a statement about the story that was either true or false. All story questions were counter-balanced with an equivalent number of true or false correct answers to avoid any systematic bias in answering unknown questions (see supplementary information available in Appendix A.4.2).

4.2.4.1 Dependent and independent variables

The recursive clause structure questions required cognitive processing of linguistic structure and constituted the central dependent variable. The higher-order intentionality questions required complex mentalising over a character's perspective in a social situation and constituted the central independent variable. The social memory questions provided a more general index of short-term memory capacity, served as a control for the basic understanding of the stories and constituted a second central independent variable. Further, the social memory questions were also used both to reduce the demand characteristics of the study (by obscuring the central hypothesis under consideration) and to provide some convergent evidence for the effectiveness of the experimental manipulation. The study used a within-subjects design for all three dependent and independent variables.

4.3 Results

The data for the current study were analysed using repeated measures *t*-tests, simple linear regression, multiple-regression and Bayesian network analysis. Overall, the level at which participants passed intentionality questions was approximately normally distributed with a peak at level 5 ($M = 5.26$) and the same was true for recursive syntax questions ($M = 5.64$) (Figure 4.1). Performance on factual memory questions was slightly higher ($M = 5.93$) and noticeably more right-skewed. The mean and variance in intentionality questions were similar to that previously reported (Kinderman et al. 1998; Stiller and Dunbar, 1998).

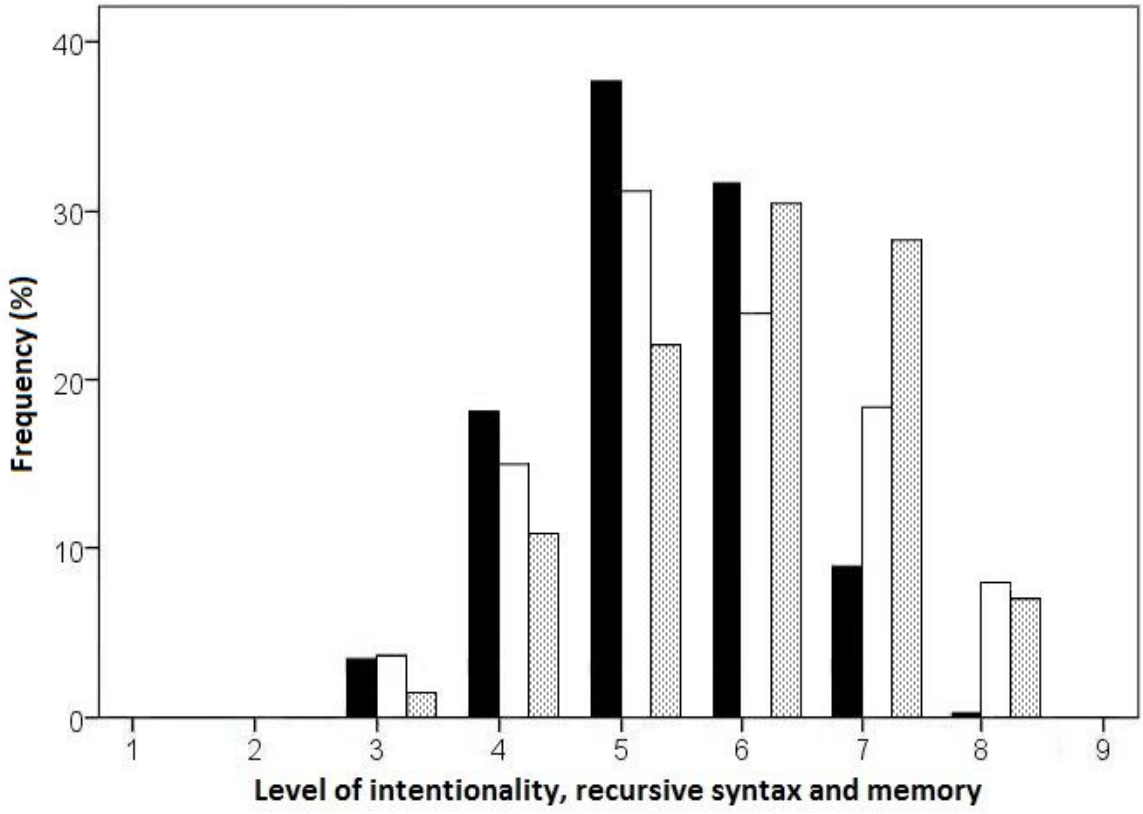

Figure 4.1. The level of intentionality, recursive syntax and memory at which participants passed perspective-taking questions. Intentionality questions are depicted as black bars, recursive syntax questions as white bars and memory questions as shaded bars. Values are rounded here to the nearest whole number.

Repeated measures *t*-tests indicated the mean score of memory performance (M = 5.93) for each subject was significantly greater than recursive syntax performance (M = 5.64), $t(413) = 5.603$, $p < .001$, Cohen's $d = 0.551$, which was in turn significantly greater than intentionality performance (M = 5.26), $t(413) = 7.671$, $p < .001$, Cohen's $d = 0.755$, while the mean score of memory performance for each subject was significantly greater than the mean score of intentionality performance, $t(413) = 13.309$, $p < .001$, Cohen's $d = 1.310$ (Table 4.1). Although there is complete overlap in the ranges of male and female scores, women had significantly higher scores than men on both intentionality performance (M = 5.49 vs. M = 5.03, respectively; $t(412) = 4.996$, $p < .001$, Cohen's $d = 0.4923$), and recursive syntax performance (M = 5.85 vs. M = 5.43, respectively; $t(412) = 3.461$, $p < .001$, Cohen's d =

0.341), but not on memory performance (M = 6.00 vs. M = 5.85, respectively; $t(412) = 1.381$, $p = .168$, Cohen's $d = 0.136$) (Table 4.1).

Table 4.1
Mean values of major independent and dependent variables.
Factual question condition.

	Mean	Male	Female
Intentionality	5.26 (SE=0.05)	5.03 (SE=0.07)	5.49 (SE=0.06)
Recursive syntax	5.64 (SE=0.06)	5.43 (SE=0.08)	5.85 (SE=0.09)
Memory	5.93 (SE=0.06)	5.85 (SE=0.08)	6.00 (SE=0.08)

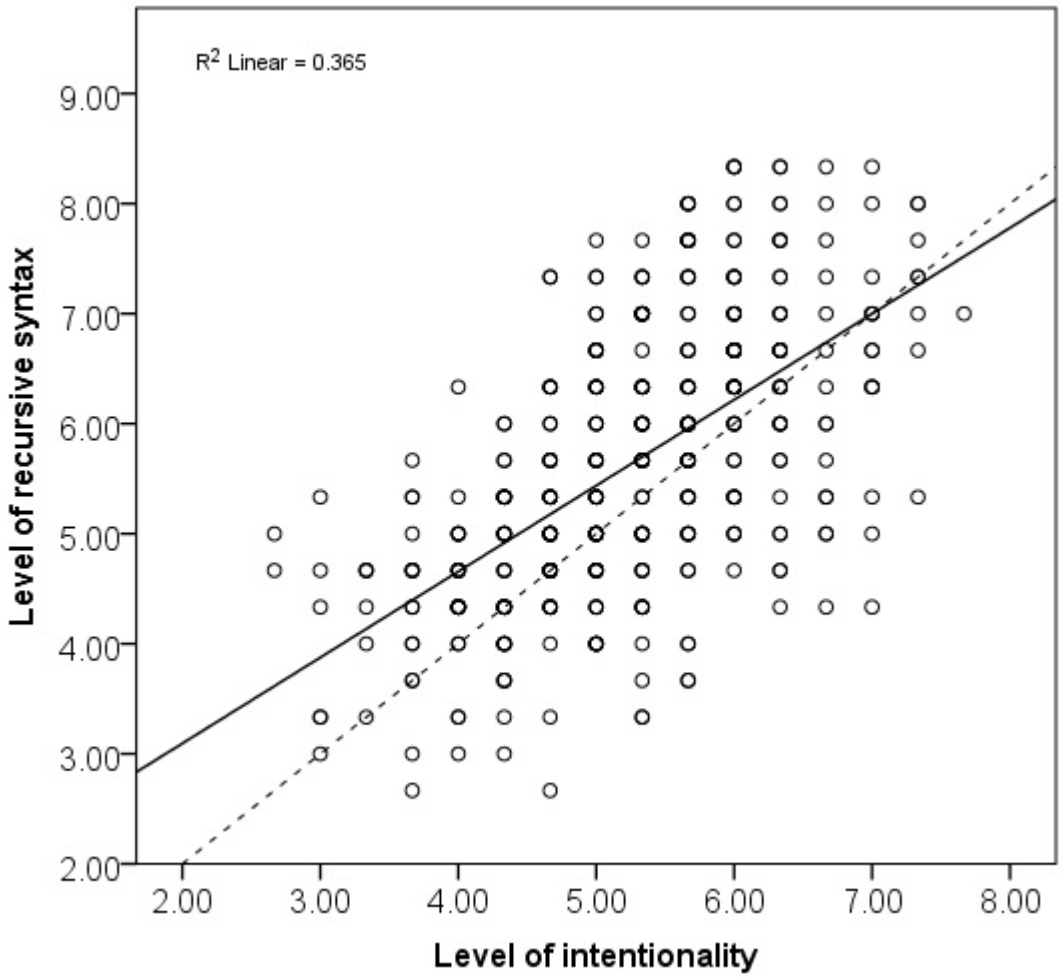

Figure 4.2a. The relationship between performance on recursive syntax and intentionality tasks. The level of recursive syntax measure is plotted as the dependent variable and level of intentionality measure as the independent variable. The line of equality is indicated above as a dashed line. The equation for the best fit liner regression line is $y = 0.78x + 1.53$.

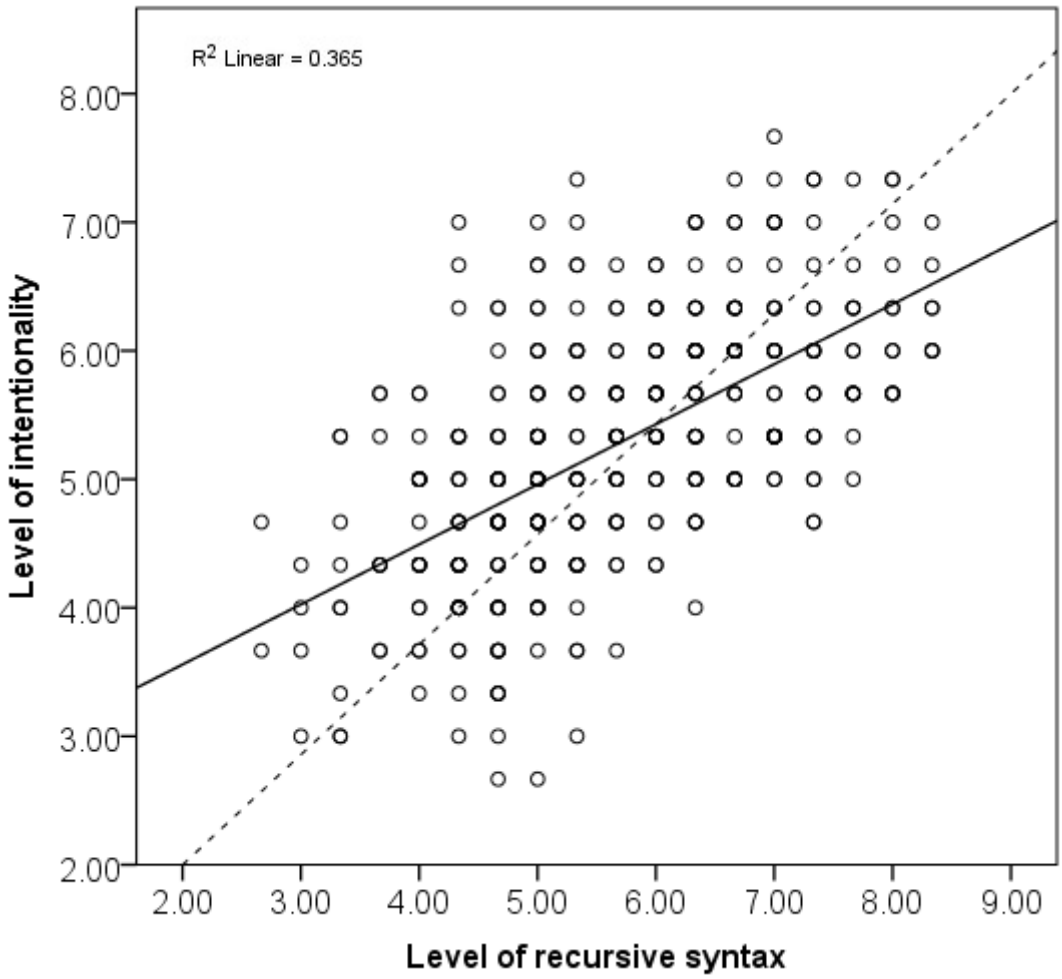

Figure 4.2b. The relationship between performance on intentionality and recursive syntax tasks. The level of intentionality measure is plotted as the dependent variable and level of recursive syntax measure as the independent variable. The line of equality is indicated above as a dashed line. The equation for the best fit liner regression line is $y = 0.47x + 2.62$.

To determine the influence of intentionality performance on recursive syntax performance, I conducted a simple linear regression analysis with recursive syntax as the dependent variable and intentionality as the independent variable. As shown in Figure 4.2a, a significant linear relationship was found ($r^2 = .365$, $F(1, 412) = 236.75$, $p < .001$). The reverse association is shown in Figure 4.2b.

Next, to determine the influence of memory on intentionality and recursive syntax performance, I conducted pairwise regressions with intentionality as the dependent variable and memory as the independent variable, and recursive syntax as the dependent variable and

memory as the independent variable. As shown in Figure 4.3 and Figure 4.4, a significant linear relationship was found between intentionality and memory performance ($r^2 = .283$, $F(1, 412) = 162.23$, $p < .001$), and between recursive syntax and memory performance ($r^2 = .377$, $F(1, 412) = 248.84$, $p < .001$).

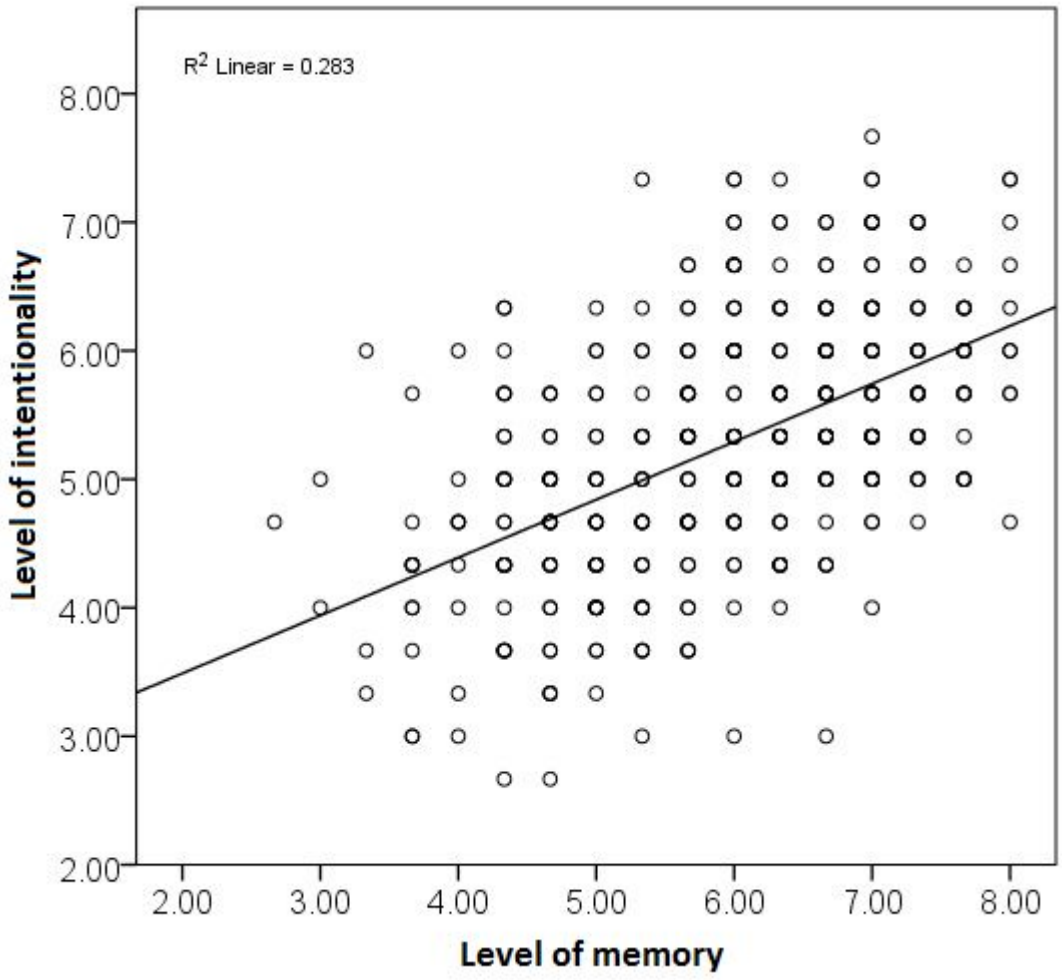

Figure 4.3. The relationship between performance on memory and mindreading tasks. The equation for the best fit liner regression line is $y = 0.45x + 2.59$.

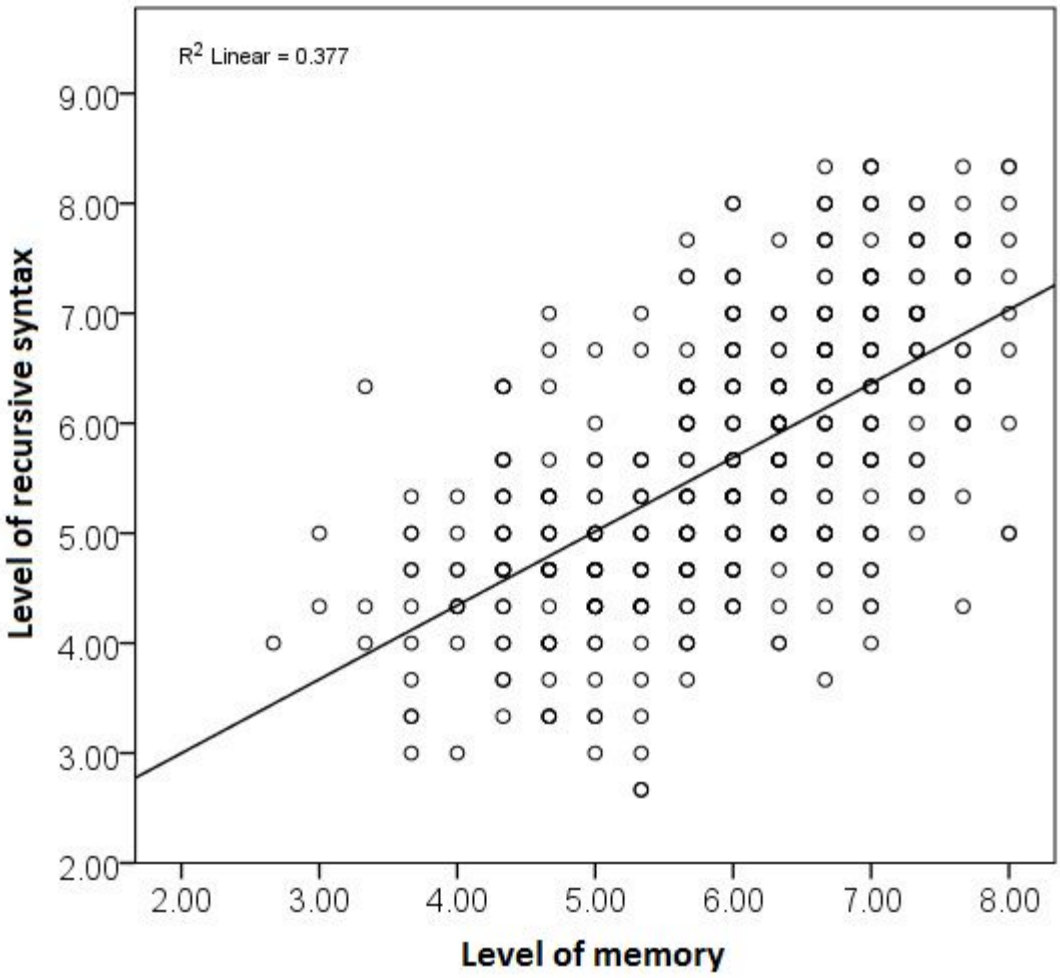

Figure 4.4. The relationship between performance on memory and embedded recursive syntax tasks. The equation for the best fit liner regression line is $y = 0.67x + 1.65$.

To determine if intentionality has an effect independently of memory on recursive syntax performance, a multiple regression analysis was conducted to explore the relationship between recursive syntax as the dependent variable, and memory and intentionality as independent variables. As shown in Table 4.2, including memory in the analysis as an independent variable gives a slightly improved fit. However, the unique contribution of memory to the total R square value only explains 12% of the variance in performance on embedded recursive syntax tasks. Similarly, in order to determine if recursive syntax has an effect independently of memory on intentionality performance, a multiple regression analysis was conducted to explore the relationship between intentionality as the dependent variable,

and memory and recursive syntax as independent variables. The inclusion of memory in the analysis as an independent variable gives an improved fit (Table 4.3). Furthermore, inspection of each respective R square value in Table 4.2 and Table 4.3 reveals that memory and intentionality performance are better predictors of recursive syntax performance (R^2 = .484, $F(1, 411) = 192.917$, $p < .001$), than memory and recursive syntax performance are as predictors of intentionality performance (R^2 = .406, $F(1, 411) = 140.705$, $p < .001$).

Table 4.2
Multiple linear regression analysis for variables predicting level of recursive syntax performance (N = 414)

Variable	*B*	*SE B*	*β*	*t*	*p*	*df1*	*df2*
Constant	.358	.273		1.310	.191	2	411
Intentionality	.501	.054	.387	9.262	.000	2	411
Memory	.447	.046	.408	9.749	.000	2	411
R	.696*						
R^2	.484*						
F-value	192.917*						

*Note: *p < .001*

Table 4.3
Multiple linear regression analysis for variables predicting level of intentionality performance (N = 414)

Variable	*B*	*SE B*	*β*	*t*	*p*	*df1*	*df2*
Constant	2.018	.204		9.894	.000	2	411
Recursive syntax	.345	.037	.446	9.262	.000	2	411
Memory	.219	.041	.258	5.360	.000	2	411
R	.638*						
R^2	.406*						
F-value	140.705*						

*Note: *p < .001*

Finally, I used a probabilistic Bayesian network analysis from the pcalg package implemented in R version 3.0.1 (Colombo et al. 2014) to determine the structural relationships between the three core variables, memory, recursive syntax and intentionality performance. As shown in Figure 4.5, there were robust positive interactions between memory and recursive syntax (r_{partial} (414) = .434, p = .001) and intentionality and recursive

syntax ($r_{partial}$ (414) = .416, p = .001), with a weaker interaction between memory and intentionality ($r_{partial}$ (414) = .255, p = .001).

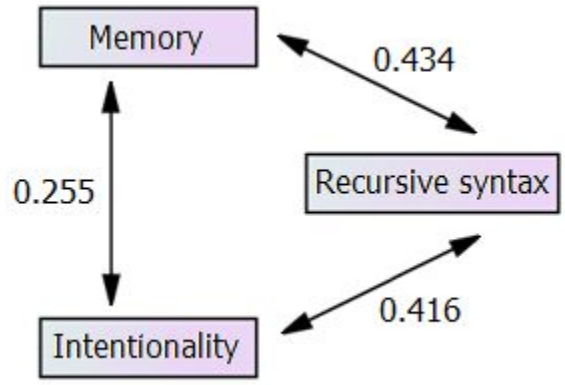

Figure 4.5. Bayesian network analysis of intentionality, recursive syntax and memory as dependent variables depicted as a graphical model.

4.4 Discussion

4.4.1 Summary of results

Chapter IV employed a standardised measure of higher-order intentionality performance, a novel recursive syntax measure and a basic short-term memory measure to determine whether higher-order intentionality capacity is positively related to recursive syntax comprehension, as predicted by the cognitive bootstrapping hypothesis. Previous studies had not experimentally tested higher-order intentionality, recursive syntax and short-term memory capacity in the context of a social vignette methodology. Accordingly, this experiment constitutes an important contribution of the present thesis.

The results confirm previous findings that there is an upper limit on performance on mindreading tasks at the equivalent of fifth-order intentionality (Kinderman et al. 1998; Stiller and Dunbar, 2007). The proportion of all participants correctly answering a question at a given level of intentionality declined precipitously when questions contained more than five

levels of mentalising (Figure 4.1). Several other studies have also found a typical fifth-order limit on mentalising tasks (Lyons, Caldwell and Shultz, 2010; Nettle and Liddle, 2008; Paal and Bereczkei, 2007) and there is evidence to support the claim that mentalising tasks are significantly more demanding of neurological resources and functional connectivity than factual memory tasks (Klapwijk et al. 2013; Lewis et al. 2011; Powell et al. 2010).

Inspection of Figure 4.1 indicates that recursive syntax competences also exceed performance on intentionality, suggesting that they too may be cognitively less demanding (though more demanding than factual memory when facts are not recursively embedded). In addition, inspection of the means in Table 4.1 reveals that females performed significantly better than males on both intentionality performance and recursive syntax performance. As a consequence, it is possible that sex differences in mentalising ability may partially account for sex differences in language ability. Indeed, as this and previous research have shown, women tend to have higher performance than men in both abilities (Hyde and Linn, 1988; Powell et al. 2010; Stiller and Dunbar, 2007).

The major finding of this study, however, is that higher-order intentionality competence correlates with recursive syntax performance independently of short term memory capacity. This result is consistent with prediction 1. Recursive syntax performance was here indexed as the number of embedded clauses within a single given sentence. As shown in Table 4.2 and Table 4.3, the significant relationships revealed by multiple regression analyses between recursive syntax performance and performance on the intentionality task suggests that there is an association between intentionality ability and recursive syntax that is independent of performance on memory tasks. Furthermore, inspection of each respective R^2 value in Table 4.2 and Table 4.3 reveals that memory and intentionality performance are better predictors of recursive syntax performance, than memory and recursive syntax performance are predictors of intentionality performance.

The results are also consistent with prediction 2. The findings revealed a significantly greater performance for memory of socially-situated sequential clause structure questions than for either intentionality questions or embedded clause structure questions. This is in line with the suggestion that, as the more evolutionarily ancient primate cognitive trait (Rushworth, Mars and Sallet, 2013), simple fact-based social memory tasks should be cognitively less demanding than equivalent fact-based mentalising or embedded syntax problem-solving (Kinderman et al. 1998; Stiller and Dunbar, 2007). As previously noted, the index of memory performance was a measure of short-term memory rather than long-term memory (which might be anticipated to be more important for longer term relationships). Moreover, so as to provide a more explicit comparison with mindreading ability, it focused deliberately on memory for basic social facts from the vignettes rather than instrumental facts about the world which were more likely to be general knowledge (and hence reflective of long term memory).

Prediction 3 was that there would be a scaffolding relationship between intentionality and recursive syntax. Although this is implicit in the fact that the two are correlated, deciding which way the causal sequence runs requires examination of the findings from the path analysis in more detail. Accordingly, there are two possible interpretations of the path analysis results: 1) a linear interpretation where memory influences recursive syntax which in turn influences higher-order intentionality and 2) a V-shaped interpretation where memory and higher-order intentionality independently influence recursive syntax.

The first linear interpretation of the path diagram in Figure 4.5, that memory influences recursive syntax which in turn influences higher-order intentionality, can be supported by several findings. Firstly, the mean value for recursive syntax performance was significantly greater than the mean value for intentionality performance (Table 4.1), suggesting that intentionality may be cognitively more demanding than syntactical recursion

per se, suggesting that the former is more likely to be constrained by the latter, rather than vice versa. Secondly, the fact that the regression equation for intentionality regressed on recursive syntax is significantly shallower than the line of equality, might lead us to suppose that recursive syntax scaffolds intentionality, perhaps allowing us to create the higher orders of intentionality via the propositional calculus underpinned by grammatical recursion. This would be in accord with one possible linear interpretation of the results of the Bayesian network analysis: memory constrains recursive syntax, which in turn constrains intentionality.

The second V-shaped interpretation of the path diagram in Figure 4.5, is that memory and intentionality independently determine recursive syntax, with both being independently determined. This interpretation would fit with the neuroanatomical evidence suggesting that memory is primarily associated with the temporal lobes (Eichenbaum, 2001; Gabrieli and Kao, 2007) whereas mentalising competences and language are particularly associated with the frontal cortex (Lewis et al. 2011; Powell et al. 2010; Powell et al. 2014). Discriminating between these alternative interpretations of the Bayesian network analysis is not easy, given that the weightings on each direction are very similar. However, the fact that memory has a weak correlation with intentionality that is independent of its relationship with recursive syntax tends to favour the V-shaped interpretation rather than the linear one. A linear interpretation might be more plausible if memory did not correlate at all with intentionality.

4.4.2 Implications

This study supports the claim that adult higher-order intentionality capacity is positively related to recursive syntax comprehension. Intentionality may thus provide the cognitive scaffolding which recursive syntax utilises for a cognitive bootstrapping effect, perhaps explaining in part how and why so many modern languages exhibit recursive syntax.

Indeed, this is a potentially important finding, as previous theorising on the evolution of language has vociferously argued that syntax could only evolve in the human species via an unexplainable evolutionary 'macromutation' (Bickerton, 1990, 1998), defying conventional wisdom on the ability of phyletic gradualism to explain complex evolved systems. Although this view has now been largely discredited by many scholars (Fitch, 2010; Pinker and Bloom, 1990), the proposal still has yet to be replaced by a more viable hypothesis with supporting evidence.

Nonetheless, an alternative proposal argues for no innate structures or biological constraints to explain the emergence of syntax in modern languages; rather, syntax is viewed as a culturally derived byproduct which develops by the cultural process of grammaticalization, under the constraint of effective communication (Tomasello, 1999). At the opposite extreme, other commentators argue for a rich and complex set of adaptations specific to syntax (Pinker and Jackendoff, 2005). The current study represents a middle-ground between these two extreme perspectives, with an origin for complex syntax in social knowledge and understanding (Cheney and Seyfarth, 2007; Dunbar, 2009).

In addition, higher-order intentionality may be much more critical than previously understood for further explaining many aspects of human nature and defining the difference between human and non-human primate cognition. In fact, several studies have already demonstrated strong correlations between intentionality performance and social network size in humans (Lewis et al. 2011; Stiller and Dunbar, 2007). Moreover, other recent studies have shown associations between intentionality and the appreciation of complex social narratives in literature (Carney et al. 2014). Accordingly, this may inspire further future studies of potential associations between intentionality and additional aspects of human nature, including religion, cooperation and complex social cognition. In summary, understanding the importance and relationship between intentionality and recursive syntax, as well as potential

associations between intentionality and other aspects of human nature, bring us some way toward understanding the mechanisms that have contributed to the complexity of our species.

Nevertheless, intentionality understood as a simple construct, and by extension the relationship between intentionality and recursive syntax, may not necessarily be entirely straightforward. Indeed, it is worth noting that Figure 4.2b suggests a more complex possibility; namely, that intentionality performance exceeds recursive syntax performance from first-order through fifth-order intentionality, but then reverses. The pitch of the slope in Figure 4.2b suggests that recursive syntax abilities are lower than intentionality competences below 5th order, but that the reverse is the case at higher values, indicating that intentionality competences asymptote at a lower value than those for syntax. The plot in Figure 4.2b suggests that there is an absolute upper limit at 7th order intentionality on mentalising skills, and a much more fluid upper limit for recursive syntax. More importantly, the pitch of the slopes suggests that recursive syntax skills develop faster than intentionality. Extrapolating the regression on the left hand side of Figure 4.2b suggests that 3rd order recursion is achieved when individuals acquire 4th order intentionality, while the ability to manage complex embedded sentence structures at 5th order recursion is achieved when individuals have acquired 5th order intentionality. Consistent with this result, several studies of child language development have found child syntactic abilites are typically comparable to those of adults by age 8-11, provided the children are given a task that is cognitively challenging enough to reveal their syntactic competence (Berman, 2004; Nippold et al. 2005; Nippold, Ward-Lonergan and Fanning, 2005; Nippold, 2009), supporting the suggestion that complex cognition is driving the use of complex syntax (Nippold et al 2005; Nippold, 2009).

Intentionality may thus provide the cognitive scaffolding with which embedded syntax utilises for a cognitive bootstrapping effect, at least up until fifth-order intentionality. However, once a given level of language competence has been achieved, recursive syntax

may then provide the cognitive scaffolding for even higher levels of mindreading, enabling some individuals to develop performance above fifth-order intentionality. In summary, while first through fifth-order intentionality may be necessary to assist the processing of simpler syntactic structures, once mentalising ability has been taxed to an uppermost limit, the cognitive scaffolding provided by recursive syntax may be engaged to enable higher-order mentalising beyond fifth-order intentionality.

A potential objection to the results found here is that the strong correlation seen here between recursive syntax and intentionality ability is merely a more general by-product of general intelligence. However, given the generally *lower* correlations (than the one revealed for the current study) typically recorded between the various general intelligence subtests, such as vocabulary, comprehension, arithmetic, picture completion, and digit span (Chabris, 2007), this indicates the unique contribution specifically of intentionality competence on recursive syntax performance. As such, this interpretation is unlikely.

A second potential objection is that the strong correlation seen here between memory and recursive syntax is a potential confound with respect to the unique contribution of intentionality on recursive syntax performance. Certainly, as shown in Figure 4.1, Figure 4.3 and Figure 4.4, memory is in fact a significant predictive variable, especially with respect to recursive syntax performance above fifth-order intentionality. However, for many reasons, this is hardly surprising; previous studies have shown the limit of short term memory to be roughly seven, plus-or-minus two items (Miller, 1994) and our results replicate this, with mean memory score at nearly six, higher than either intentionality or recursive syntax performance.

Moreover, working memory itself has been further shown to be strongly related to performance on the vast majority of complex cognitive tasks, including reading comprehension, problem-solving, and general intelligence (Ackerman, Beier and Boyle,

2005; Conway, Kane and Engle, 2003). It is therefore likely that short-term memory capacity explains a significant proportion of the variance in recursive syntax performance. Nevertheless, given that the correlation between memory and recursive syntax performance is approximately equal in proportion to the correlation between intentionality and recursive syntax performance argues against the interpretation that recursive syntax performance can be explained solely as a by-product of short-term memory capacity.

4.4.3 Limitations

Finally, the measures of both higher-order intentionality and recursive syntax performance used in this study had certain clear limitations. The primary limitation in the intentionality measures was largely in the fact that each was itself contained within a sentence with some degree of syntactic embedding. Therefore, it is likely the intentionality questions imposed a double cognitive load; the intentionality task problem-solving, as well as the embedded syntax problem-solving. The recursive syntax questions, by comparison, had no aspects of mindreading or intentionality. Although efforts were made to reduce the amount of syntactic embedding within each intentionality question, realistically it is probable that there was still some degree of interference.

As a result, some have argued that adult higher-order intentionality ability may be greater than those reported in previous studies, due to the bias inherent in written narratives (O'Grady et al. 2015). However, given that written narratives can be sequentially read, reread and cross-checked in a way that cannot be done with spoken statements, it is certainly not intuitively obvious why this should actually be the case. If anything, performance on spoken narratives should be lower than on written narratives, yet the O'Grady et al. 2015 study suggests they are not on their particular tasks (and, indeed, with little meaningful variation in performance across higher levels of intentionality). Accordingly, given that all previous text-

based tasks have found that performance levels steadily decline with increasing intentionality levels (i.e. more individuals' mind states involved in the story), this might suggest that their tasks are in fact being solved without the need to process all of the mind states putatively included in their tasks. However, this remains to be investigated.

The recursive syntax measure used in this study also had clear limitations. Namely, the recursive syntax measure consisted of author-written sentences, based on the number of embedded clauses within each sentence. Unfortunately, no prior studies had previously investigated recursive syntax at equivalent levels of complexity; therefore, embedded syntax sentences were developed for the current study, which varied in their level of recursion. Accordingly, because this was not a validated scale, it may have been subject to various unidentified confounds. Future studies should perhaps attempt to construct a more accurate scale of both higher-order intentionality and recursive syntax with rigorous validity and reliability.

Chapter V

Discussion

The aim of the current thesis was to explore the biological function of human language. As outlined in the Chapter I introduction, investigation into potential explanations for the function of language is motivated by the fact that language is a ubiquitous and defining feature of human nature. In congruence with the apparent complexity and multi-purpose use of language in various diverse human contexts, this thesis employed a broad-in-scope approach to investigate possible explanations for why human language evolved. I did so by way of three sets of experiments, within data Chapters II, III and IV, each of which was designed to test a particular hypothesis or set of hypotheses about one aspect of the way language might function. Lastly, although not directly tested in the current thesis, this analysis was nevertheless embedded within the historical background of the biological processes responsible for the evolution of language in the human species.

Chapter II introduced a novel experimentally-controlled human social foraging paradigm to determine whether language facilitates cooperation, as predicted by the interdependence hypothesis. Previous studies of this hypothesis had never actually experimentally tested prosociality, communication and behavioural cooperation in the context of a human social foraging context, even though this hypothesis is explicitly conceived in terms of cooperation during foraging.

The results from this experiment did not support the prediction that gesture or language-based communication facilitates cooperation, as predicted by the interdependence hypothesis (Tomasello et al. 2012). Indeed, the findings revealed that cooperation in an experimental social foraging context is probably facilitated through observation and behavioural mimicry. Nonetheless, group size had a significant impact, demonstrating that

more information was likely transmitted through this imitative medium in large groups compared to small groups. However, there was suggestive evidence for the proposal that subjective ratings of prosociality and communication would be related to better group performance; indeed, several of the self-rated relationship variables were positively correlated with social foraging performance.

Given that the results of Chapter II did not seem to support the interdependence (cooperation) hypothesis of human language function, Chapter III tested a series of alternative proposals. Here, a series of experimentally controlled author-written vignettes was used to test four proposed hypotheses which claim to characterise language in terms of honest signalling theory. Previous studies of this hypothesis had not experimentally tested any of these diverse proposals in the literature, by any experimental method.

The results from Experiments I and II of Chapter III did not support the prediction that use of low-frequency or poetic vocabulary enhances the sexual attractiveness of either males or females in the way predicted by the sexual selection hypothesis (Miller, 2000). Moreover, the results of Experiment III did not support the prediction that charisma, at least as reflected in being able to speak in large public gatherings, is attractive to the opposite sex (Burling, 1986). However, the results of Experiment IV supported the hypothesis that the high prevalence of linguistically honest information in the everyday use of language (and potential reputation-based social costs of lying and deception) may have evolved as a socially or sexually selected handicap (Bliege Bird and Smith, 2005; Lachmann et al. 2001). In other words, truth-telling and self-deprecation may be an honest signal (as a receiver-dependent cost) of humility, prosociality and social status with respect to human mate choice.

Indeed, the results of Experiment IV supported the prediction that revealed courtship deception leads to significant modulations in attractiveness ratings, for both males and females. The overstated attorney and janitor vignettes were rated as less attractive following

revealed *self-enhancing* deception (in comparison to the initial description), indicating that lies which inflated the traits of a potential romantic interest were costly in terms of attractiveness ratings. On the other hand, the understated janitor and attorney vignettes were rated as more attractive, following revealed *self-deprecating* deception, indicating that lies which 'undersold' the traits of a potential romantic interest were beneficial in terms of attractiveness ratings.

Given that the results of Chapter III seemed to support an honest signalling theory approach to language function, and signalling of the kind tested in Experiment IV necessarily depends on understanding how signals would be interpreted by recipients, the next data chapter elected to evaluate a key mechanism likely to be important in this process, namely mentalising - the ability to read others' minds. Chapter IV employed a standardised measure of higher-order intentionality performance, a novel recursive syntax measure and a basic short-term memory measure to determine whether higher-order intentionality capacity is positively related to recursive syntax comprehension, as predicted by the cognitive bootstrapping hypothesis. Indeed, this hypothesis represents a unique proposal argued to explain language structure; that is, why many of the world's roughly 7,000 languages have recursive syntax. Previous studies had not experimentally tested higher-order intentionality, recursive syntax and short-term memory capacity, in the context of a social vignette method.

The results from this experiment indicate that higher-order intentionality competence correlates with recursive syntax performance independently of short term memory capacity. In addition, the findings also revealed a potentially complex causal relationship: while first through fifth-order mindreading may be necessary to assist the processing of simpler syntactic structures, once mentalising ability has been taxed to its normal limit, the cognitive scaffolding provided by recursive syntax may be recruited to enable higher-order mentalising beyond fifth-order intentionality, at least in some individuals.

5.1 Overview

Following decades of research into the communicative abilities of human children and other animals, most linguists, psychologists and evolutionary biologists now agree that language is unique to humans (Barrett, Dunbar and Lycett, 2002; Fitch, 2010; Fromkin, Rodman and Hyams, 2013). Yet, its functional significance, and therefore its evolutionary origins, remains unresolved (Számadó and Szathmáry, 2006). As argued in Chapter I, a comprehensive review of the contemporary literature suggests that there are probably three serious contenders for the adaptive significance of human language: the interdependence or cooperation hypothesis (Tomasello et al. 2012), the sexual selection or honest signalling theory (Bliege Bird and Smith, 2005; Lachmann et al. 2001) and the social bonding hypothesis (Dunbar, 2004; Freeberg, Dunbar and Ord, 2012).

However, as I noted in Chapter I, nearly every study to have considered the function(s) of language has simply presented evidence in favour of its preferred hypothesis. Few attempts have been made to test between competing hypotheses. Indeed, many proposals often fail to recognise that there may be competing hypotheses. Confirmatory evidence is always, at best, weak evidence for a hypothesis, not least because it is quite common for evidence, in many cases, to be compatible with several otherwise mutually incompatible hypotheses. Thus, both the cooperation and the social bonding hypothesis might predict that language use or language complexity increases with group size, and so evidence that this is so could not be used to prove the validity of either hypothesis. A direct comparison between competing hypotheses thus provides a heuristically more powerful way of testing a particular hypothesis and particularly so where different hypotheses yield incompatible predictions. Only once alternative hypotheses have been eliminated can one definitively conclude that the most prominent hypothesis is in fact the most likely explanation.

In addition, there are at least two further reasons for undertaking an experimental study of all three of the most prominent hypotheses. First, most alternative hypotheses for the function of language – albeit not the social bonding hypothesis – have relied mainly on theoretical arguments, and much less on data or experimental methods. In cases where quantitative data or experimental tests have been presented, they have mostly focussed on derivative considerations (e.g. development, mechanisms, context, etc.) and simply assumed a preferred function. It was thus desirable to break this pattern and put some of these untested hypotheses directly to the test. Second, testing both the social complexity hypothesis and the social bonding hypothesis is not especially straightforward or intuitively obvious, at least in human subjects. In testing the social complexity hypothesis, perhaps the biggest problem experienced in the present study was in recruiting the large numbers of subjects necessary for properly evaluating this proposal. This was true even with the quite modest sized groups, of no more than 20 individuals per group, tested in this study. Testing the social bonding hypothesis presents equally complex challenges. Although several studies have shown that the spread of reputational information through social gossip promotes social bonding (Feinberg, Willer and Shultz, 2014; Weaver and Bosson, 2011), fewer studies have shown how or in fact whether language might promote social bonding outside the context of social gossip (although see Ireland et al. 2011 for one possibility in romantic contexts).

In Chapter II, I used a novel experimentally-controlled human social foraging method to test between the three aforementioned alternative hypotheses. Participants, working in small and large groups, collected tokens from a number of distinct 'foraging patches' and cached them at a central point. The principal findings were that: (i) language was used more frequently in large group than in small group coordination, (ii) performance increased in larger groups as a result of more independent sources of information, however (iii) the multiple regression analyses indicated that neither verbal nor gestural communication were

critical for overall group performance, and that the human social foraging groups seemed to cooperate well in spite of their gestural and vocal communication, ostensibly relying much more on imitation and behavioural mimicry; finally, that (iv) there was some evidence for the potential social bonding and cooperation effects, at least as measured by the self-report prosociality index questions, in this human social foraging experiment. On balance, then the results showed only limited evidence for the interdependence hypothesis of human language function.

Indeed, the human social foraging groups seemed to cooperate well irrespective of how much gestural and vocal communication there was; it seemed that success in these experiments depended much more on imitation and behavioural mimicry. Nevertheless, the positive correlations between the prosociality ratings and the group performance measures indicated that groups that were subjectively more cooperative and familiar with one another performed better on the human social foraging task.

Although, at face value, human social foraging experiments would seem to be an ecologically valid way to test hypotheses regarding the effect of language on cooperation and/or social bonding, in retrospect it may not have been as ideal as initially conceived. For one thing, the social foraging paradigm is very task-specific and considers only a particular kind of task – a foraging context. Many of the social decisions that humans face are not at all related to foraging. Some of them are concerned with mate choice or mate advertising, others are related to the more general business of living in peaceful communities, and still others are concerned with coordinating communal defence against raiders or other threats (Johnson and Earle, 2001). None of these functions is indexed well by a foraging task. That said, given that the results from this experiment did not provide results consistent with the hypothesis that language (or gesture) facilitates cooperation within an ecologically valid context, I explored the use of language in a wider variety of contexts, namely courtship and mate choice.

Given this, Chapter III introduced an experimentally-controlled vignette-based design to examine some of the more explicitly social functions of language under the rubric of the sexual selection hypothesis and honest signalling theory. Here, the aim was to investigate three versions of the sexual selection hypothesis and one account of the deception hypothesis, all argued to be congruent with honest signalling theory. All four experiments used the same vignette-based methodology in which participants rated each vignette protagonist on a series of personality-based criteria.

In toto, these experiments provided very little evidence for the sexual selection hypothesis of human language function. Experiments I to III tested three different versions of the honest-signalling hypothesis. They focused, respectively, on the value of being able to use low-frequency vocabulary, being able to use more poetic vocabulary (as receiver-independent costs) and being seen to be a good public speaker (as a receiver-dependent cost) as mate choice criteria. The experiments yielded no evidence to support any of these predictions. In Experiment IV, I therefore tested an alternative possibility: that self-disclosure of revealing or culpable personal information (as a receiver-dependent cost), as a kind of social risk-taking, is a form of social advertisement simultaneously signalling imperviousness to danger and social repercussions. Experiment IV yielded some support for the prediction that prosocial courtship deception leads to increased attractiveness ratings, while antisocial courtship deception lead to decreased attractiveness ratings (as a direct test of both increased social status and social sanctions for deception within the context of mate attraction). In conclusion, supporting evidence was found only for the deception hypothesis of human language function, at least in so far as mate choice processes are concerned.

In view of the rather mixed results in Chapters II and III, in Chapter IV I shifted focus to explore some of the cognitive mechanisms underpinning language. Since mindreading is a core feature of human social behaviour, I wanted to see if individual differences in

mentalising abilities correlated with individual differences in grammatical competences. If there was such a correlation, it would constitute some prima facie evidence for a social function of language (e.g. favoured evidence for the social bonding hypothesis, over say, the cultural learning hypothesis). I did this using a within-subjects design that compared subjects' abilities handling syntactically embedded sentences with their abilities of equivalent mentalising sentences for the same story. The results demonstrated that, overall, mentalising questions are more cognitively demanding than embedded syntax questions, which are in turn more cognitively demanding than pure factual memory questions, as indexed by the number of levels of facts that subjects were able to cope with successfully. In addition, a path analysis indicated that the most plausible interpretation of the possible relationships between the three variables was that memory and higher-order intentionality performance had significant and independent main effects on recursive syntax performance.

5.2 Broader implications

Given the results from these varied series of experiments, several broader implications are worthy of note. Indeed, while the social bonding hypothesis for the evolution of language was not itself tested directly, the empirical results from the various experiments in this thesis lend at least indirect support to this hypothesis against the alternatives. The results from Chapter II, in particular, would seem to provide some indirect support for the social bonding hypothesis. Firstly, groups seemed to forage quite effectively in spite of the use of either language or gesture, ostensibly discounting the cooperation hypothesis for language function. Secondly, self-report measures of both group cooperation and the request for verbal information were positive predictors of group performance, at least in the case of the percentage of 'good' tokens collected. In addition, both group closeness and familiarity were

also positive predictors of convergence time. Importantly, this suggests that when groups are familiar with each other, their performance improves, perhaps because they are then more willing to seek and share information. Therefore, while language by itself doesn't cause a group to become more efficient, it might well allow a group that is already well integrated to exchange more information (i.e. groups of strangers, it would seem, do not gain from being able to talk to each other). In other words, these results might suggest that social bonding is perhaps more fundamental than cooperation; in the absence of a group common bond, cooperation may be much more challenging to achieve. Thirdly, as reviewed in Chapter I, among traditional societies, hunting and gathering itself is typically done in silence or utilizing only gestures and more often relying on observation and mindreading abilities to know what to do and when, without the need for language of any kind (Smith, 1991). Nevertheless, it is noteworthy that the results from Chapter II de-emphasized the importance of gesture; on both the criteria used in Chapter II, performance did not seem to dramatically improve when there was more than one individual gesturing. This suggests that there may be an upper limit on the number of individuals that can productively gesture at any one time.

In contrast, performance did seem to improve with more speakers, possibly reflecting the fact that it may be possible to maintain several independent conversations simultaneously when the group is large enough. If so, this may have important implications for the debate over the gestural versus vocal origin of language: speech would appear to have a clear advantage in group-based tasks. As such, the results from Chapter II suggest that gesturing is less efficient than speaking; although lots of participants gestured, they didn't actually coordinate each other's behaviour as well as they did with speech. Nevertheless, the limit to conversation group size at four suggests that large groups cannot be coordinated through language alone. As implied by the positive effect of familiarity on convergence time, it may require formal or informal rules of coordination based on prior experience and intentional

familiarity with fellow group members. This result would also be consistent with the research from traditional societies; in rare cases where hunter-gatherers do engage in extensive sharing of information, the conditions under which it occurs are highly variable and condition dependent, typically occurring only prior to the act of foraging (Smith, 1991).

From an even broader perspective, when viewed through the lens of the evolution of human language, the results would nevertheless seem to give at least some tentative support for a gestural origin of human communication. Indeed, it has been argued that pointing and pantomiming is a natural human instinct for directing attention (Tomasello, 2008). Moreover, studies have shown that even prelinguistic infants use and understand the pointing gesture (Tomasello, 2008), while gesture is used ubiquitously among adults across a crowded room or among tourists finding their way around a foreign country, in which no one speaks their conventional language (Tomasello, 2008). Accordingly, although the multiple regression analyses showed that neither language nor gesture was critical for group cooperation, given the obvious prevalence of gesture in this study, it could well be that gesture is more important for merely directing and signalling attention, as opposed to facilitating cooperation.

Nevertheless, if gesture was indeed somehow important in the evolution of human language, it has yet to be persuasively demonstrated how gesture might fit into a social bonding hypothesis framework. Given the fact that at some point during primate evolution, a transition was ostensibly made from manual grooming to vocal communication (as a method of social bonding), one speculative proposal is that gesture somehow bridged that gap. Indeed, several studies have shown that gestural and behavioural mimicry promotes social bonding in both human and non-human primates (Lakin and Chartrand, 2003; Paukner et al. 2009), perhaps providing a plausible evolutionary transition from manual grooming, to gestural and behavioural mimicry, and eventually to a vocal channel of social bonding.

Tangentially to this, it is worth noting that there was an approximate ratio of one

speaker for every three listeners, as has been found by previous studies to be the limit on conversation group size (Dunbar et al. 1995; Waller et al. 2011; Dezecache and Dunbar, 2012). Consequently, there may be a critical limit on the number of conversation participants a social group can maintain at around four (Dunbar, 1993; Dunbar et al. 1995; Waller et al. 2011), although this did not necessarily seem to impede the effective transfer of information. Indeed, large groups did significantly better overall than smaller groups. This may, in part, reflect the fact that individuals resorted to gesture, which does not suffer the same limitations from noise interference as auditory communication. Nevertheless, overall, neither language nor gesture seemed to have a significant influence in groups below twenty individuals, although the possibility that either or both might be important in groups larger than twenty cannot be ruled out. Perhaps the safest conclusion is that both language and gesture may have evolved to socially bond and signal large-scale planning and strategic arrangements rather than manage small-scale labour groups with clear goals.

The results from Chapter III would also seem to provide some indirect support for the social bonding hypothesis. Firstly, participants did not appear to be attracted to low-frequency or poetic vocabulary, ostensibly undermining the sexual selection hypothesis for language function. Secondly, some preliminary support was found for the deception hypothesis of language function. However, as reviewed in Chapter III, since deception is commonly assumed among behavioural ecologists to be a derived signalling function (and in combination with the context-specific results found here), it seems reasonably safe to assume that deception is probably a derived function of the language and social cognitive mechanisms that underpin social bonding. Thirdly, self-deprecation deception may be an honest signal (as a receiver-dependent cost) of humility, prosociality and social status with respect to human mate choice, further illustrating the social bonding function of language, even in deceptive contexts.

From a broader interpretation, when viewed from the perspective of the evolution of human language, the results clearly suggest that honest signalling theory can be successfully applied to particular aspects of human language function. Indeed, many theorists have strenuously argued that honest signalling theory, though useful in application to many different types of animal signalling systems, has absolutely no relevance to explaining or describing any aspects of human language evolution (Fitch, 2010; Zahavi and Zahavi, 1997). More specifically, it has been argued that human language (unlike most animal signalling systems) has no component that guarantees reliability and prevents lying and deception (Zahavi and Zahavi, 1997).

Nevertheless, as humans and other primates have evolved an intensely social nature, such social systems often depend on an implicit social contract where reputation and the identification of cheats is a perennial concern (Dunbar, Barrett and Lycett, 2011). As such, the results found here indicate that deception may be subject to reputation-based punishment, based on lower attractiveness ratings for dishonest self-aggrandizement. This result is therefore congruent with computational modelling experiments of honest signalling theory which demonstrate this proposal is both mathematically consistent and theoretically plausible (Lachmann et al. 2001). Thus, the results give additional support to the claim for the potential reputation-based social costs of lying and deception, thereby providing a potential explanation for the high prevalence of linguistically honest information in the everyday use of language (Bliege Bird and Smith, 2005; Lachmann et al. 2001). Indeed, the apparent mystery of linguistically honest information in the daily use of language has been identified as one of the central issues in understanding the evolution of human language (Fitch, 2010).

Lastly, the results from Chapter IV would also seem to provide some indirect support for the social bonding hypothesis. Firstly, the child development of theory of mind is characterised by many aspects of both simple and complex social cognitions, including

shared attention, directed eye gaze, imitation, understanding desires, and pretence, eventually resulting in the full understanding of false beliefs (de Villiers, 2007; Gergely, Bekkering and Kiraly, 2002; Onishi et al. 2007; Tomasello, 2009). Indeed, theory of mind is thought to be critical for the development of the kinds of advanced social cognitions that characterise humans, forming the core part of what it means to be a fundamentally socially animal (Barrett, Dunbar and Lycett, 2002; Call and Tomasello, 2008). Secondly, developmental psychologists have discovered a strong interaction exists between theory of mind and language development (de Villiers, 2007; Wellman et al. 2001) indirectly indicating that language itself may be fundamentally social. Accordingly, given the strong positive correlations found here between higher-order mentalising and recursive syntax, this would seem to further substantiate the claim that complex language abilities both require and support complex social cognition.

In addition, the path analysis results in Chapter IV allowed several important questions to be investigated with important implications for the evolution of language. First, it allowed me to ask whether one of these was evolutionarily prior to the other: in other words, did mentalising abilities scaffold the syntactical structure of language, or did the prior existence of recursive syntax-based language scaffold our higher-order mentalising abilities? As mentioned previously, the path analysis indicated that memory and higher-order intentionality performance had significant and independent main effects on recursive syntax performance, tentatively suggesting that mentalising abilities might predate formal language and have been important in scaffolding the evolution of grammatically structured language.

More interestingly, however, the results suggested that while this seemed to be the case at low intentionality levels, at higher levels the relationship is reversed, and grammatical recursion appears to scaffold high levels of intentionality. The point of transition appears to be at fifth-order intentionality, the level of competence of young teenagers and normal adults

(Berman, 2004; Nippold et al. 2005; Nippold, Ward-Lonergan and Fanning, 2005; Nippold, 2009). In effect, to get beyond the normative levels of adult human mentalising at fifth-order intentionality recursive language may be necessary. In summary, to the extent that high-order mentalising abilities of the kind found in normal, rather than exceptional adults, are crucial for a complex social life, these results here would seem to provide some indirect support for the social bonding hypothesis for the origin of language.

Therefore, when viewed from the perspective of the evolution of human language, the results clearly argue in favour of a long and slow development of increasingly complex language-like communication. Contrary to some views, which maintain language must have been either present or absent (Bickerton, 1990, 1998; Gould, 1995), these results suggest that at the start, human communication did not involve the same complexity of language as we currently know. Given the rather tight coupling between complex language and higher-order intentionality, fully developed language was likely not required until group sizes had become sufficiently large and complex enough to provide the reason for understanding and negotiating several mental states at the same time. Indeed, it is striking, and perhaps not a coincidence, that the most intimate layer of human social networks is composed of an individual's five closest friends and relatives (Zhou et al 2005); that is, the same number of individual mental states equivalent to adult human mentalising.

Moreover, the gradual phylogenetic sequence, as suggested based on these results, would also clearly argue in favour of a late emergence for fully developed language. Indeed, recent research on Neanderthal brain-size analyses demonstrate that if they had language, it could not have been fully modern language; they could only aspire to fourth-order intentionality and that would likely have meant much reduced language complexity (Gamble, Gowlett and Dunbar, 2014).

In conclusion, between them, the experiments included in this thesis allow us to rule out some of the alternative explanations for the function of language and in doing so provide some indirect evidence for the social bonding hypothesis for the function of language. This thesis also provided evidence for a unique hypothesis addressing language structure, namely recursive syntax, there providing additional purchase on how language came to evolve its complex form and structure. Indeed, the results indicated an origin for recursive syntax in complex social knowledge and understanding (Cheney and Seyfarth, 2007; Dunbar, 2009). Nevertheless, although this thesis presents an important step forward, both theoretically and empirically, into investigating the function, form and evolution of language, it is not without qualifications or limitations, which will now be examined in the following section.

5.3 Methodological challenges and limitations

As with any study, the experiments described in this thesis are not without their limitations. In this section, I consider potential limitations of the three main sets of experiments in turn.

5.3.1 Chapter II: The role of language for cooperation during human social foraging

It is possible to identify at least five possible problematic issues with the human social foraging experiment. The first of these concerns the sizes of groups that were tested. It proved particularly difficult to get large numbers of subjects to turn up at the same time to take part in an experiment, even when using occasions like university open days when large numbers of individuals were potentially available. Not only did this limit the number of large groups that could be tested, it also placed a relatively modest limit of twenty subjects on the largest size of groups that could be tested. In several cases, the results strongly suggested that performance on this task is an asymptotic function of communication frequencies, and it

would have been desirable to have groups larger than twenty in size to be sure that this was in fact accurate in all cases.

Because of this, it may have been the case that the group sizes tested here were simply not large enough to detect significant differences in performance due to vocal communication. As suggested in Chapter I, language has been proposed as a mechanism by which humans can effectively bond with many individuals at a time, thereby facilitating the creation of larger social networks than those allowed by more conventional primate bonding strategies such as dyadic manual grooming (Dunbar, 2004). The time constraints of manual grooming seem to impose an upper limit on the size of groups that monkeys and apes can maintain as coherent social entities, and this limit seems to be groups of about fifty individuals. The group sizes tested in this experiment never exceeded twenty individuals. Accordingly, if it is true that humans needed language to break through a sort of 'glass ceiling' at fifty individuals, then the foraging experiment may not have been on a large enough scale. It may be, for example, that participants simply adopted the kinds of observation or imitation strategies that the more social monkeys and apes habitually use to coordinate groups of the same size. Language may only come into play when group size exceeds fifty individuals or greater. If this is the case, then future studies should try to investigate larger groups as well as different types of cooperation and coordination.

Despite this, there are some reasons to think that this concern may be less intrusive than one might fear. One of these is that for language-using participants group performance seemed to plateau at around four individuals. This is suspiciously close to the natural limit on the size of human conversation groups, suggesting perhaps that, at least as far as work groups are concerned, audibility and other aspects of language reception may simply impose a relatively low upper limit.

A second issue relates to the quantification of cooperation. Although cooperation, in

the human social foraging experiment, was ostensibly measured as the percentage of good tokens collected and mean group consensus time, it is possible that neither of these measures were an accurate index of cooperation. Although it seems likely, given the results, that participants worked mostly independently, and converged on a group consensus through observation and behavioural mimicry alone, it is also possible that a different measure of cooperation would have yielded rather different results, perhaps ones more consistent with the interdependence hypothesis. Given that I had video footage available for all trials, it might, for example, have been possible to directly quantify the level of cooperation by having multiple naïve coders code the video files and/or through motion energy analysis software that allowed for continuous objective quantification of body positions and nonverbal behaviour. As it happens, coding the video files proved to be too time-consuming to warrant the time investment. Even so, some kind of arbitrary criterion as a definition of cooperation would still be needed in order to code events in the files and I remain somewhat unconvinced that the results would have been any different. Nonetheless, future studies should perhaps consider alternative ways of defining coordination.

A third possible issue may have been related to the way participants were recruited and assigned to groups. For instance, all groups tested were females only. This was done partly to avoid problems associated with male subjects behaving in such a way as to impress the girls rather than concentrating on completing the task and other issues commonly associated with mixed sex groups. Future studies should perhaps replicate this experiment on all male and mixed sex groups. Nonetheless, previous research has shown, in accordance with expectation states theory, that men have an advantage over women in mixed sex groups, other things being equal (Wittenbaum, 1998). Accordingly, this could have quite drastic consequences on group performance, and may make such groups undesirable for testing functional hypotheses of the kind considered here.

Fourth, it is always possible that the assignment of participants to group size conditions may not have been truly random. Although every effort was made for random assignment at the beginning of each trial, the negative correlations (albeit nonsignificant ones) found between group performance and several measures of group cooperation and familiarity might indicate otherwise. Presumably, this may have been due to the fact that many experimental trials involved psychology undergraduate students, who may well have known each other (from prior classroom affiliation) before the experiment. Future research should attempt to control for this possible confounding variable more effectively, perhaps by ensuring no prior affiliation of any kind. Alternatively, subjects could also be assessed for previous acquaintanceship at the beginning of the experiment and subsequently allocated to familiar and unfamiliar conditions in order to ensure that social cohesion acted as an independent variable. However, given that recruiting large numbers of subjects proved to be a considerable challenge, finding a viable solution to this rather basic logistical problem may be necessary if greater numbers and larger groups are to be tested effectively.

Lastly, given that the majority of self-report measures of prosociality did not demonstrate positive correlations with performance, and therefore group size conditions, it is possible that Likert-scale questionnaires may not be the most suitable method for measuring prosociality. This is likely to be even more the case when investigating the between-subjects effects of language on social bonding and cooperation between strangers. I discuss this further below in section *5.4 Future directions for research*.

In summary, future studies should perhaps replicate this experiment on all male groups, address the challenges of potential variability in how well acquainted participants are prior to the experiment, and standardise indices for measuring cooperation and prosociality.

5.3.2 Chapter III: The role of honest signalling theory and language function

Perhaps the main concern to highlight with the experiments in Chapter III is the fact

that the experimental stimuli that tested for vocabulary attractiveness in Experiments I-II examined a very narrow range of low-frequency vocabulary, as well as a rather limited number of vignettes. It is possible that some, or perhaps even all, of the participants found the particular low-frequency words used in these series of vignettes unfamiliar, unattractive, pretentious, or in some other way unappealing. It therefore might be desirable to consider re-running these experiments using several different vignettes, identical in all respects, apart from having different low-frequency vocabulary words. In contrast, Experiment IV used vignettes for two socially very different target individuals and found virtually identical results – suggesting that this experiment did not suffer from this problem.

A second potential limitation may have been related to the somewhat artificial testing methodology used in these particular series of experiments. More specifically, the results found here were based on self-report measures of opposite sex attractiveness ratings in the context of vignette-based narratives. Consequently, the results may have limited generalizability to actual situations which involve real-world people. Moreover, as the attractiveness ratings were based on self-report measures, there may have been some element of subjective bias introduced into these findings. Indeed, the difference between what individuals say and what they actually do in a given context is often considerable.

In summary, future studies should perhaps attempt to test attractiveness ratings of low-frequency vocabulary in more naturalistic environments, such as at coffee shops or conferences using actors, or in realistic dating situations (such as speed-dating events), using several different types of low-frequency vocabulary words.

5.3.3 Chapter IV: Mentalising, recursive syntax and language form

The experiment in Chapter IV was less susceptible to this particular problem – that is, to say, the limited number of vignettes in Chapter II – in that it used three vignette stories, all

of which had been used successfully in a number of previous studies. However, the measures of both syntax and higher-order intentionality could potentially be improved upon through other methodologies. Namely, the recursive syntax measure consisted of author-written sentences, based on the number of recursive syntax elements. For instance, the following example contained five embedded clauses (subscripts counting the number of embedded clauses): *Betty is a roommate of Brian$_1$, who walked into the bedroom of Frank$_2$, who received £10 from Brian$_3$, who likes lettuce and tomatoes for a salad$_4$, that Betty and Brian helped to prepare$_5$.* Unfortunately, no prior studies had previously investigated recursive syntax at equivalent levels of complexity; therefore, embedded syntax sentences were developed for the current study, which varied in their level of recursive syntax. Accordingly, because this was not a validated scale, it may have been subject to various unidentified confounds. Future studies should perhaps attempt to construct a validated scale with rigorous validity and reliability.

Similarly, although the lower range of mentalising questions were the same as those used in previous studies, the higher order questions (i.e. those at 7-9th order) were mainly constructed for the purposes of this study. Therefore, these questions may have potentially suffered from the same kinds of validity issues as already raised for the recursive syntax questions.

An additional possible problem with the intentionality measures was that each was itself contained within a sentence with some degree of syntactic embedding. Therefore, it is likely the intentionality questions imposed a double cognitive load: the intentionality task problem-solving, as well as the embedded syntax problem-solving. The recursive syntax questions, by comparison, had no aspects of mindreading or intentionality. Although efforts were made to reduce the amount of syntactic embedding within each intentionality question, realistically it is probable that there was still some degree of interference. It is difficult to

know how to avoid this problem, although the use of cartoons to provide implicit mental state embedding might be one way forward. This, however, is not without its own problems, since it is probably rather challenging to construct language-free cartoons with more than three or four embedded mental states. That in itself might suggest that language is needed to construct mental states above fifth order.

In summary, future studies should attempt to use alternative measures of both higher-order intentionality and recursive syntax to determine whether or not the positive regression analyses found here remain.

5.4 Future directions for research

5.4.1 Further tests of the social complexity hypothesis

Based on the comprehensive literature review presented in Chapter I, and the lack of supporting evidence found for many of the alternative hypotheses tested in this thesis, the weight of current evidence would seem to suggest social bonding as the most likely *ultimate* explanation for *why* humans have language, and the social complexity hypothesis as the most likely *proximate* explanation for *how* language evolved in the human species. In line with this research paradigm, the social complexity hypothesis for animal communication offers a promising new direction for future research into the function and evolution of human language. Namely, the social complexity hypothesis predicts that communicative complexity - and thereby, interactions among group members - should be regulated by five influential factors: 1) social unit size 2) social density 3) member roles 4) egalitarian structure 5) and inter-unit density. Unfortunately, although an increasing amount of data currently shows that this hypothesis explains a great deal of complex animal communication, few studies have been done on humans, thereby limiting the generalizability of the social complexity

hypothesis for explaining the evolution of human language.

Nevertheless, several experiments are currently underway (or nearing completion) to test several predictions following from the first two factors of the social complexity hypothesis. For example, in a manuscript recently submitted for publication entitled, "Social density processes regulate the functioning and performance of foraging human teams", my collaborator Andy King and I demonstrate the importance of unit density for regulating the function and performance of human social interactions, using high-resolution GPS data. More specifically, using the same methodology as the human social foraging experiment in Chapter II, we show that coordination performance is greatest at intermediate social densities, and these groups collected the largest amounts of 'forage'. However, contrary to expectations, increased coordination at intermediate densities did not result in improved foraging precision, as accuracy was equivalent across all density conditions. As such, the findings suggest a very similar set of results to those found for the social foraging experiment in Chapter II; namely, that social foragers rely upon visual information to achieve coordination (i.e. observation and behavioural mimicry), but auditory channels to maximise foraging performance. In summary, this experiment provides support for a central prediction of the social complexity hypothesis, in that unit density appears to be critical for regulating interactions among group members.

Moreover, additional data – collected, but unanalysed from my thesis – is currently being analysed by myself to test the prediction that greater communicative complexity is found in larger unit sizes of human social interactions. Consistent with predictions from the social complexity hypothesis, preliminary analyses do suggest that large groups show greater communicative complexity than small groups. More specifically, large groups speak more often, thereby resulting in larger aggregates of words, as well as use a greater diversity of words, such as nouns, verbs, adjectives, and deictic pronouns. Nevertheless, despite the supportive findings from these two sets of experiments, future studies should test the

remaining predictions of the social complexity hypothesis which address the factors of member role, egalitarian structure and inter-unit density. In summary, this experiment also provides support for a central prediction of the social complexity hypothesis, in that groups with complex social systems require more complex communicative systems for regulating interactions and relations among group members.

5.4.2 Developing accurate measures of social bonding

The present thesis, predominantly in Chapter II, included a range of behavioural measures (i.e. cooperative social foraging) and self-report measures of social closeness, which could be broadly referred to as indices of prosociality. In the psychological literature, the methods used to assess prosociality generally include helpfulness tasks (Kirschner and Tomasello, 2010), economic games (Kurzban, 2001), subjective questions of closeness and intimacy (Aron et al. 1992), and synchrony and/or coordination (Wiltermuth and Heath, 2009). In many cases, the selection of just what sort of prosociality measures to include in any given study can be very unstandardized, and therefore ostensibly arbitrary. Accordingly, future studies on prosociality and social bonding would likely benefit enormously from a set of consistently valid and reliable indices to accurately represent these complex psychological constructs.

As discussed in Chapter II, the measures used in the present thesis were selected based on their successful use in many similar studies in psychology. Nevertheless, it is important to highlight the inconsistency between the behavioural and self-report prosociality measures in Chapter II, and that support for the validity and reliability of these measures in the psychological literature is mixed and somewhat inconclusive. For example, economic games, although widely used in the psychological, sociological and economic literature, have been recently criticised for their lack of ecological validity, as well as whether they constitute

a genuine measure of prosociality (Burton-Chellew and West, 2013). As such, it is even more difficult to see how economic games might have been applicable for determining prosociality in the context of the present study, where a between-subject experiment was used to investigate the effect of communication and group size in a cooperative social foraging task. Additional measures, including implicit self-report questions, have also been used in many recent studies. For example, the willingness to donate a kidney or give blood to another individual, have been used in several studies (Curry and Dunbar, 2013). Eye gaze contact has also been argued to be useful method of demonstrating affiliative intentions (Dunbar and Shultz, 2010).

Nevertheless, the Inclusion of Other in the Self (IOS) scale has recently been touted to have a high test-retest validity and reliability and to be one of the more accurate measures of subjective interpersonal closeness and intimacy (Aron et al. 1992; Dibble et al. 2012; Hornsey et al. 2012). As such, a before-and-after the experiment administration of the IOS scale would likely also have been desirable, to adjust for baseline feelings of interpersonal affiliation. Moreover, as previously mentioned, given that there may have been interference of random allocation, future studies should attempt to ensure larger sample sizes and true random allocation to experimental conditions, to reduce confounds due to prior affiliation. In conclusion, the best metric for measuring prosociality remains a controversial area, but future studies would likely benefit from a standardised subjective scale of prosociality.

5.4.3 Experimental tests of the social bonding hypothesis

Primate social bonding ostensibly operates through a dual process mechanism that involves a psychopharmacological component (mediated through oxytocin, endorphins, dopamine and various other neuropeptides: Dunbar, 2010) and a cognitive component (whereby relationships of trust, obligation and reciprocity are established: Dunbar and Shultz,

2010; Silk, Cheney and Seyfarth, 2013). Accordingly, while an increasing amount of evidence currently suggests an important cognitive component; namely, the connection between social gossip and social bonding (Feinberg, Willer and Shultz, 2014; Weaver and Bosson, 2011), there is currently less evidence to suggest that the conversational use of language might induce a simultaneous psychopharmacological component, mediated through any of several different neuropeptides. Indeed, previous research has found that endorphins are released in response to socially synchronised versus non-social non-synchronised sporting events (Cohen et al. 2010), and that behavioural synchrony itself promotes social bonding (Wiltermuth and Heath, 2009). However, similar studies of endorphin or other neuropeptide activation, during conversationally synchronised language use, have yet to be directly investigated.

Nevertheless, the need to coordinate and synchronise vocal behaviour has often been identified as an important aspect of bonded relationships in many animal species, including humans (Dunbar and Shultz, 2010). For example, compared to the sexually dimorphic polygynous red-winged blackbird, synchronised duetting is practiced almost exclusively among the monogamous red-shouldered blackbird, primarily facilitating a long-term pairbond (Whittingham, Kirkconnell, and Ratcliffe, 1997). In addition, among primate species, gibbons differ from chimpanzees, bonobos, gorillas and orangutans in that gibbons are obligately monogamous, primarily due to raising dependent offspring under the threat of infanticide (Opie et al. 2013). Unlike their polygamous ape cousins, both male and female gibbons sing long and complex song bouts, which in most cases are combined to produce highly coordinated and synchronised duets, primarily in service of long-term pairbonds (Geissmann and Orgeldinger, 2000). Among gibbon pairs that perform the most highly synchronised duets, they typically spend more time grooming each other, engaging in the same activities, and spend more time in close proximity (Geissmann and Orgeldinger, 2000).

Finally, similar to animal pair-bonding, the coordination and synchrony of human vocal behaviour has been shown to have important implications for social bonding within the context of romantic relationships. For instance, one study in particular found that among speed-dating dyads, both romantic initiation and follow-up relationship stability were predicted by unconscious conversational language style matching (LSM) scores (determined as an index of equivalent pronouns, articles, verbs, etc. used in conversation) (Ireland et al. 2011). In summary, given the current evidence, highlighting the importance of vocal synchrony for social bonding in many animal species (including humans and other primates) future studies should investigate whether synchronised vocal behaviours are simultaneously supported by a complex psychopharmacological component mediated through the release of any of several various neuropeptides.

Accordingly, several additional experiments could potentially be designed to test this psychopharmacological component of the social bonding hypothesis. For example, one potential within-subjects experiment might involve an experimental intervention, where a long-term romantic couple is encouraged to talk to each other for five minutes - while following conventional practice, using an increase in pain threshold as a proxy assay for endorphin activation (Cohen et al. 2010; Dunbar et al. 2012); and a control intervention, where the same romantic couple is again encouraged to talk to each other for five minutes - while using the administration of naltrexone, a β-endorphin-specific antagonist, to confirm an opiate effect (Depue and Morrone-Strupinsky, 2005; Madsen et al. 2007; Zubietta et al. 2001). The introduction of naltrexone, an endorphin blocker, would aim to directly manipulate the endorphin system, rather than relying entirely on pain threshold assays to infer a potential relationship between conversational language use, endorphin activation, and social bonding.

A second experiment could investigate the effects of dyadic talking together for 5

minutes, compared to the effects of a 5 minute light-stroking massage from a partner. The use of a blood pressure cuff could be employed to induce ischemic pain as a simple, low-cost method for indexing endorphin release (Cohen et al. 2010; Dunbar et al. 2012). A before-and-after administration of the IOS scale could be used to measure differences in subjective feelings of interpersonal closeness and affiliation. As such, if it is indeed the case that language evolved as a substitute for primate social grooming, then the prediction would be that touch would create the greatest difference both in pain tolerance as well as IOS scale measures. In conclusion, activation of the endorphin system is hypothesized to be a causal mediator between synchronised conversational language use or light stroking massage, and social bonding, as measured by a prosociality index, the IOS scale.

5.5 Final conclusions

This thesis aimed to comprehensively and critically evaluate the weight of current literature on human language function and evolution as well as experimentally test alternatives to the social bonding hypothesis. Previous studies, apart from the social bonding hypothesis, were limited due to the predominantly highly speculative and theoretical approach to this area of inquiry. The human social foraging experiment in Chapter II attempted to break this pattern for various versions of the cooperation hypothesis. The author-written vignettes in Chapter III attempted to do this for various versions of honest signalling theory. Lastly, the written narrative stories and questions in Chapter IV attempted to do this for the structure and form of language. The upshot of this research demonstrated that many of these questions are in fact amenable to empirical investigation.

That said, Chapter II demonstrated very little evidence for the interdependence (cooperation) hypothesis of human language function. Chapter III showed little support for

the function of language as a sexual advertisement strategy (as a receiver-independent cost), but tentative support for language having a deceptive function (as a receiver-dependent cost), in particular mate-specific contexts. Chapter IV showed preliminary support for the conclusion that mindreading capabilities may have influenced the structure of language; namely, recursive syntax. As mentalising is a key component of effective interpersonal empathy and understanding, this ostensibly provides indirect support for the social bonding hypothesis. Based on these results, and the comprehensive literature review surveyed in Chapter I, the social bonding hypothesis emerges as the most likely contender for explaining the biological function, and hence ultimate evolutionary explanation, for why humans have language.

References

Abeler, J., Becker, A. and Falk, A. (2012). Truth-telling: A representative assessment. *Institute for the Study of Labor (IZA), 6919*, 1-18.

Ackerman, P. L., Beier, M. E. and Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30-60.

Adachi, I. (2014). Spontaneous spatial mapping of learned sequence in chimpanzees: Evidence for a SNARC-like effect. *PLoS ONE, 9*(3), 1-4.

Akhtar, N. and Gernsbacher, M. A. (2007). Joint attention and vocabulary development: A critical look. *Language and Linguistic Compass, 1*(3), 195-207.

Allen, G. R. and Gardner, B. T. (1980). Comparative psychology and language acquisition. In T. A. Sebok and J. U. Sebok (Eds.), *Speaking of apes: A critical anthology of two-way communication with man* (pp. 287-329). New York: Plenum Press.

Apicella, C. L., Marlowe, F. W., Fowler, J. H. and Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature, 481*, 497-502.

Apperly, I. A., Samson, D. and Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology, 45*(1), 190-201.

Argyle, M., Lalljee, M. G. and Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human Relations, 21*, 3-17.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596-612.

Astington, J. W. and Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*, 1311-1320.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*, 346-349.

Baillargeon, R., Scott R. M. and He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110-118.

Baldwin, D. A. *(*1995*).* Understanding the link between joint attention and language*. In* C. Moore *and* P. J. Dunham *(Eds.), Joint attention: Its origins and role in development (pp.* 131-158*).* Hillsdale, NJ*:* Lawrence Erlbaum*.*

Bale, C., Morrison, R., and Caryl, P. G. (2006). Chat-up lines as male sexual displays. *Personality and Individual Differences, 40,* 655-664.

Bard, K. A. (1993). Commentary on Tomasello, Kruger and Ratner's "Cultural learning". [Peer commentary by K. A. Bard]. *Behavioral and Brain Sciences, 16*, 495-552.

Baron-Cohen, S. (2009). Autism: The empathizing-systemizing (E-S) theory. *Annals of the New York Academy of Sciences, 1156,* 68-80.

Barrett, L., Dunbar, R. I. M., and Lycett, J. (2002). *Human evolutionary psychology.* Princeton: Princeton University Press.

Baumeister, R. F., Zhang, L., Vohs, K. D. (2004). Gossip as cultural learning. *Review of General Psychology, 8*(2), 111-121.

Beal, D. J., Cohen, R. R., Burke, M. J. and McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, *88*(6), 989-1004.

Berman, R. A. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. A. Berman (Ed.), *Language development across childhood and adolescence: Volume 3. Trends in language acquistion research (TILAR)* (pp. 9-34). Amsterdam: John Benjamins.

Bickerton, D. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by D. Bickerton]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Bickerton, D. (1998). Catastrophic evolution: The case for a single step from protolanguage to full human language. In J. R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the evolution of language* (pp. 341-358). Cambridge: Cambridge University Press.

Bickerton, D. (1990). *Language and species.* Chicago: University of Chicago Press.

Bickerton, D. and Szathmáry, E. (2011). Confrontational scavenging as a possible source for language and cooperation. *BMC Evolutionary Biology, 11*(261), 1-7.

Bird, R. (1999). Cooperation and conflict: The behavioral ecology of the sexual division of labor. *Evolutionary Anthropology, 8*(2), 65-75.

Bischoping, K. (1993). Gender differences in conversation topics 1922-1990. *Sex Roles, 28,* 1-17.

Blackshaw, A. J., Kinderman, P., Hare, D. J. and Hatton, C. (2001). Theory of mind, causal attributions and paranoia in Asperger's syndrome. *Autism*, ***5***, 147-163.

Bliege Bird, R. and Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology 46*(2), 221-248.

Blumstein, D. T. and Armitage, K. B. (1997). Does sociality drive the evolution of communicative complexity? A comparative test with ground-dwelling sciurid alarm calls. *American Naturalist, 150*, 179-200.

Boesch, C. (2002). Cooperative hunting roles among Taï chimpanzees. *Human Nature, 13*(1), 27-46.

Boesch, C. (1994). Cooperative hunting in wild chimpanzees. *Animal Behavior, 48(3),* 653-667.

Boesch, C., and Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Taï National Park. *American Journal of Physical Anthropology, 78,* 547-573.

Bolund, E., Schielzeth, H. and Forstmeier, W. (2012). Singing activity stimulates partner reproductive investment rather than increasing paternity success in zebra finches. *Behavioral Ecology and Sociobiology, 66*(6), 975-984.

Bornstein, R. F., Riggs, J. M., Hill, E. L. and Calabrese, C. (1996). Activity, passivity, self-denigration, and self-promotion: Toward an interactionist model of interpersonal dependency. *Journal of Personality, 64*(3), 637-673.

Boughman, J. W. and Wilkinson, G. S. (1998). Greater spear-nosed bats discriminate group mates by vocalizations. *Animal Behaviour, 55*, 1717-1732.

Bradbury, J. W. and Vehrencamp, S. L. (2011). *Principles of animal communication.* Sunderland, MA: Sinauer Associates.

Bressler, C. and Balshine, P. G. (2006). Chat-up lines as male sexual displays. *Evolution and Human Behavior, 40,* 29-664.

Briscoe, T. (2008). Language learning, power laws, and sexual selection. *Mind and Society, 7*, 65-76.

Bro-Jørgensen, J. and Pangle, W. M. (2010). Male topi antelopes alarm snort deceptively to retain females for mating. *American Naturalist, 176*(1), 33-39.

Bruner, J. S. (1972). The uses of immaturity. *American Psychologist, 27,* 1-22.

Buchanan, T. and Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125-144.

Bugnyarf, T. and Kotrschal, K., (2002). Observational learning and the raiding of food caches in ravens, *Corvus corax*: is it 'tactical' deception?" *Animal Behaviour, 64*, 185–195.

Burling, R. (1993). Primate calls, human language and nonverbal communication [and comments and reply]. *Current Anthropology, 34*(1), 25-53.

Burling, R. (1986). The selective advantage of complex language. *Ethology and Sociobiology, 7*(1), 1-16.

Burton-Chellew, M. N. and West, S. A. (2013). Pro-social preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences, USA, 110*, 216-221.

Buss, D. M. (1989). Sex differences in human mate selection: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*, 1-49.

Buss, D. M., Abbott, M., Angleitner, A., Biaggio, A., Blanco-Villasenor, A., Bruchon-Schweitzer, M [and 45 additional authors]. (1990). International preferences in selecting mates: A study of 37 cultures. *Journal of Cross-Cultural Psychology, 21*, 5-47.

Byrne, R. and Whiten. A. (1991). Computation and mindreading in primate tactical deception. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 127-141). Cambridge: Basil Blackwell.

Byrne, R. W. and Whiten, A. (1990). Tactical deception in primates: The 1990 database. *Primate Report, 27*, 1-101.

Byrne, R. and Whiten, A. (1985). Tactical deception of familiar individuals in baboons (*Papio ursinus*). *Animal Behaviour, 33*(2), 669-673.

Calhim, S., Shi, J., and Dunbar, R. I. M. (2006). Sexual segregation among feral goats: Testing between alternative hypotheses. *Animal Behaviour, 72*, 31-41.

Call, J. and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science, 12*, 187-192.

Carey, S. 2004. Bootstrapping and the origin of concepts. *Daedalus, 133*(1), 59-68.

Carey, S. 2009. *The origin of concepts*. New York: Oxford University Press.

Carney, J., Wlodarski, R. and Dunbar, R. (2014). Inference or enaction? The impact of genre on the narrative processing of other minds. *PLoS ONE, 9*(12), 1-14.

Chabris, C. F. (2007). Cognitive and neurobiological mechanisms for the law of general intelligence. In Roberts, M. J. (Ed.) *Integrating the mind: Domain general versus domain specific processes in higher cognition.* Hove, UK: Psychology Press.

Cheney, D. L. and Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind.* Chicago: University of Chicago Press.

Cheney, D. L. and Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species.* Chicago: University of Chicago Press.

Christiansen, M. H. and Kirby, S. (Eds). (2003). *Language evolution: States of the art.* Oxford: Oxford University Press.

Clark, B. (2011). Scavenging, the stag hunt, and the evolution of language. *Journal of Linguistics, 47*(2), 447-480.

Clutton-Brock, T. H., and Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209-216.

Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology, 53*(5), 588-616.

Cohen, E. E., Ejsmond-Frey, R., Knight, N., Dunbar, R. I. M. (2010). Rowers' high: Behavioural synchrony is correlated with elevated pain thresholds. *Biology Letters, 6*(1), 106-108.

Cohen, J. E. (1971). *Casual groups of monkeys and men: Stochastic models of elemental social systems.* Cambridge, MA: Harvard University Press.

Collier, J. (1993). Commentary on Tomasello, Kruger and Ratner's "Cultural learning". [Peer commentary by J. Collier]. *Behavioral and Brain Sciences, 16*, 495-552.

Collins, E. C., Percy, E. J., Smith, E. R. and Kruschke, J. K. (2011). Integrating advice and experience: Learning and decision making with social and nonsocial cues. *Journal of Personality and Social Psychology, 100*, 967-982.

Colombo, D., Hauser, A., Kalisch, M. and Maechler, M. (2014). pcalg: Methods for graphical models and causal inference. *R Package Version, 2.0-3.* http://pcalg.r-forge.r-project.org.

Conway, A. R., Kane, M. J. and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552.

Craik, K. H. (2009). *Reputation: A network interpretation.* New York: Oxford University Press.

Crockford, C., Herbinger, I., Vigilant, L. and Boesch, C. (2004). Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology, 110*, 221-243.

Curry, O. and Dunbar, R. I. M. (2013). Do birds of a feather flock together? The relationship between similarity and altruism in social networks. *Human Nature, 24*(3), 336-47.

Dávid-Barrett, T. and Dunbar, R. I. M. (2013). Processing power limits social group size: Computational evidence for the cognitive costs of sociality. *Proceedings of the Royal Society B, 280*, 1-8.

Dawkins, R. and Krebs, J. R. (1978). Animal signals: information or manipulation? In J. R. Krebs and N. B. Davies (Eds.)., *Behavioural ecology* (pp. 282-309). Oxford: Blackwell Scientific Publications.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain.* New York: W. W. Norton.

De Backer, C. J. S and Gurven, M. (2006). Whispering down the lane: The economics of vicarious information transfer. *Adaptive Behaviour, 14(3), 249-264.*

DePaulo, B. M., and Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*(1), 63-79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M. and Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979-995.

Depue, R. A. and Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualising a human trait of affiliation. *Behavioral and Brain Sciences, 28*(3), 313-350.

Deutscher, G. (2005). *The unfolding of language: An evolutionary tour of mankind's greatest invention*. New York: Henry Holt.

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*(11), 1858-1878.

de Villiers, J. and de Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell and K. J. Riggs (Eds.) *Children's reasoning and the mind* (pp. 191-228). Hove, England: Psychology Press.

de Villiers, J. and Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex recursive syntax and false-belief-understanding. *Cognitive Development, 17*, 1037-1060.

deWaal, F. (2007). *Chimpanzee politics: Power and sex among apes.* Baltimore: Johns Hopkins University Press.

deWaal, F. (1986). Deception in the natural communication of chimpanzees. In R. W. Mitchell and N. S. Thompson (Eds.), *Deception: Perspectives on human and non-human deceit* (pp. 221-224). Albany: State University of New York.

Dezecache, G. and Dunbar R. I. M. (2012). Sharing the joke: The size of natural laughter groups. *Evolution & Human Behavior*, *33*(6), 775-779.

Dibble, J. L., Levine, T. R. and Park, H. S. (2012). The unidimensional relationship closeness scale (URCS): Reliability and validity evidence for a new measure of relationship closeness. *Psychological Assessment, 24*(3), 565-72.

Donath, J. (in press). *Signals, truth, and design.* Cambridge, MA: MIT Press.

Donath, J. (2008). Signals in social supernets. *Journal of Computer-Mediated Communication, 13*, 231-251.

Donnai, D. and Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics: Seminars in Medical Genetics, 97*(2), 164-171.

Dugatkin, L. A. (1992). Sexual selection and imitation: Females copy the mate choice of others. *American Naturalist, 139*(6), 1384-1389.

Dumontheil, I., Apperly, I. A. and Blakemore, S-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13*, 331-338.

Dunbar, R. I. M. (2010). The social role of touch in humans and primates: Behavioural function and neurobiological mechanisms. *Neuroscience and Biobehavioral Reviews, 34*, 260-268.

Dunbar, R. I. M. (2009). Why only humans have language. In R. Botha and C. Knight (Eds.), *The prehistory of language* (pp. 12-35). Oxford: Oxford University Press.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8*, 100-110.

Dunbar, R. I. M. (2003). The origin and subsequent evolution of language. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution: States of the art* (pp. 219-234). Oxford: Oxford University Press.

Dunbar, R. I. M. (1998). Theory of mind and the evolution of language. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the evolution of language* (pp. 92-110). Cambridge: Cambridge University Press.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences, 16*(4), 681-735.

Dunbar, R. I. M, Baron, R., Frangou, A., Pearce, E., van Leeuwen, E. J. C., Stow, J., Partridge, G., MacDonald, I., Barra, V., and van Vugt, M. (2012). Social laughter is correlated with an elevated pain threshold. *Proceedings of the Royal Society B: Biological Sciences, 22*, 279, 1161-1167.

Dunbar, R., Barrett, L., and Lycett, J. (2011). *Evolutionary psychology: A beginner's guide.* Oxford: Oneworld.

Dunbar, R. I. M., Duncan, N. D. C. and Marriott, A. (1997). Human conversational behaviour. *Human Nature, 8*, 231-246.

Dunbar, R. I. M., Duncan, N. D. C. and Nettle, D. (1995). Size and structure of freely forming conversational groups. *Human Nature, 6*, 67-78.

Dunbar, R. I. M. and Shultz, S. (2010). Bondedness and sociality. *Behaviour, 147*, 775-803.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C. and Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*, 1352-1366.

Easley, D. and Kleinberg, J. (2010). *Networks, crowds and markets: Reasoning about a highly connected world.* New York: Cambridge University Press.

Eisenberg, J. (1981). *The mammalian radiations*. London: Athlone Press.

Ekman, P. (2009). Lie catching and microexpressions. In C. Martin (Ed.), *The philosophy of deception* (pp. 118-135). Oxford: Oxford University Press.

Ellison, N., Heino, R. and Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication, 11*(2), 415-441.

Epstein, J., Klinkenberg, W. D., Wiley, D. and McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior, 17*(3), 339-346.

Farrar, M. J. and Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language, 22*, 197-213.

Feinberg, M., Willer, R. and Shultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*(3), 656-664.

Fitch, W. T. (2010). *The evolution of language.* New York: Cambridge University Press.

Fitch, W. T. (2005). The evolution of language: A comparative review. *Biology and Philosophy, 20*, 193-230.

Fitch, W. T. and Hauser, M. D. (2002). Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals. In *Acoustic communication* (Eds.) A. Simmons, R. R. Fay and A. N. Popper. New York: Springer.

Francis, W. N and Kučera, H. (1967). *Computational analysis of present-day American English.* Providence, Rhode Island: Brown University Press.

Freeberg, T. M. (2006). Social complexity can drive vocal complexity: Group size influences vocal information in Carolina chickadees. *Psychological Science, 17*, 557-561.

Freeberg, T. M., Dunbar, R. I. M. and Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B, 367*, 1785-1801.

Fromkin, V., Rodman, R. and Hyams, N. (2013). *An introduction to language.* 10th ed. New York: Cengage Learning.

Fuertes, J. N., Potere, J. C. and Ramirez, K. Y. (2002). Effects of speech accents on interpersonal evaluations: Implications for counselling practice and research. *Cultural Diversity and Ethnic Minority Psychology, 8*, 346-356.

Galantucci, B. and Roberts, G. (2014). Do we notice when communication goes awry? An investigation of people's sensitivity to coherence in spontaneous conversation. *PLoS ONE, 9*(7), 1-5.

Gale, E., de Villiers, P., de Villiers, J., and Pyers, J. (1996). *Language and theory of mind in oral deaf children.* Paper presented at the 20th Boston University Conference on Language Development, November, Somerville, MA.

Gamble, C., Gowlett, J. and Dunbar, R. (2014). *Thinking big: How the evolution of social life shaped the human mind.* London: Thames and Hudson.

Geissmann T. and Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Animal Behaviour, 60*, 805-809.

Gentner, D. and Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, *2*(2), 261-283.

Gergely, G. Bekkering, H. and Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature, 415,* 755.

Giles, H. (1973). Communicative effectiveness as a function of accented speech. *Speech Monographs, 40*, 330-331.

Gluszek, A. and Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review, 14*, 214-237.

Goodall, J. (1986). The chimpanzees of Gombe. Cambridge, MA: Belknap Press.

Gosling, S. D., Vazire, S., Srivastava, S. and John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*(2), 93-104.

Gould, S. J. (1995). *The human language series*. Video. Program Three. By Gene Searchinger. New York: Equinox Films/Ways of Knowing Inc.

Graf Estes, K., Evans, J. L., Alibali, M. W. and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254-260.

Green, M. (2005). *Book of lies*. Kansas City, MO: Andrews McMeel Publishing.

Greenfield, P. and Lave, J. (1982). Cognitive aspects of informal education. In D. Wagner and H. Stevenson (Eds.), *Cultural perspectives in child development* (pp. 181-207). New York: Freeman.

Greengross, G., and Miller, G. (2008). Dissing oneself versus dissing rivals: Effects of status, personality, and sex on the short-term and long-term attractiveness of self-deprecating and other-deprecating humor. *Evolutionary Psychology, 6*(3), 393-408.

Gustison, M. L., Roux, A. and Bergman, T. J. (2012). Derived vocalisations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society B, 367*(1597), 1847-1859.

Hall, J., Park, N., Song, H., and Cody, M. (2010). Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships, 27*(1), 117-135.

Hamao, S. (2012). Acoustic structure of songs in island populations of the Japanese bush warbler, *Cettia diphone*, in relation to sexual selection. *Journal of Ethology, 31*, 9-15.

Hamilton, A. F. (2009). Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881-892.

Hancock, J. T., Thom-Santelli, J., and Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behaviour. *CHI Letters, 6*(1), 129-134.

Haselton, M. G. & Miller, G. F. (2006). Women's fertility across the cycle increases the short-term attractiveness of creative intelligence compared to wealth. *Human Nature, 17*, *50-73*.

Hausberger, M., Bigot, E. and Clergeau, P. (2008). Dialect use in large assemblies: A study in European starling *Sturnus vulgaris* roosts. *Journal of Avian Biology, 39*, 672-682.

Hauser, M. D. (1999). Sex and status effects on primate volubility: Clues to the origin of vocal languages? *Evolution and Human Behavior, 18*, 151-158.

Hauser, M. D. (1997*). The evolution of communication*. Cambridge: MIT Press.

Hauser, M. D. (1993). Commentary on Tomasello, Kruger and Ratner's "Cultural learning". [Peer commentary by M. D. Hauser]. *Behavioral and Brain Sciences, 16*, *495-552.*

Hauser, M. D. (1992). Costs of deception: Cheaters are punished in rhesus monkeys. *Proceedings National Academy of Sciences, 89*, 12137-12139.

Haviland, J. B. (1991). *Gossip, reputation and knowledge in Zinacantan.* Chicago: University of Chicago Press.

Henzi, S., Hawker-Bond, D., Stiller, J. and de Sousa Pereira, L. (2007). Look who's talking: developmental trends in the size of conversational cliques. *Evolution and Human Behavior, 28(1)*, 66-74.

Herbinger, I., Pepworth, S., Boesch, C., and Zuberbüler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: A playback study. *Animal Behaviour, 78*, 1389-1396.

Herwegen, J. V. and Karmiloff-Smith, A. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by J. V. Herwegen and A. Karmiloff-Smith]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Hess, N. H. and Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature, 17*, 337-354.

Hopp, S. L., Jablonski, P. and Brown, J. L. (2001). Recognition of group membership by voice in Mexican jays, *Aphelocoma ultramarina*. *Animal Behaviour, 62*, 297-303.

Hornsey, M. J., Olsen, S., Barlow, F. K., Oei, T. P. S. (2012). Testing a single-item visual analogue scale as a proxy for cohesiveness in group psychotherapy. *Group Dynamics: Theory, Research, and Practice, 16*(1), 80-90.

Hughes, S. M., Dispenza, F. and Gallup, G. G. J. (2004). *Evolution and Human Behavior, 25*, 295-304.

Hurford, J. R. (2011). *The origins of grammar: Language in the light of evolution II.* Oxford: Oxford University Press.

Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution.* Oxford: Oxford University Press.

Huttenlocher, J., Vasilyeva, M., Cymerman, E. and Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45,* 337-374.

Hyde, J S. and Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Iacoboni, M. and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7*(12), 942-951.

Iñiguez, G., Govezensky, T., Dunbar, R., Kaski, K. and Barrio, R. A. (2014). Effects of deception in social networks. *Proceedings of the Royal Society B, 281*, 1-9.

Ireland, M. E., Slatcher, R. B, Eastwick, P. W., Scissors, L. E., Finkel, E. J. and Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science, 22*(1), 39-44.

Jackson, M. O. (2010). *Social and economic networks.* Princeton: Princeton University Press.

James, S. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations.* New York: Doubleday.

Janik, V. M., Sayigh, L. S., and Wells, R. S. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *Proceedings of the National Academy of Sciences USA, 103*(21), 8293-8297.

Jehn, K. A. and Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediating processes in friendship and acquaintance groups. *Journal of Personality and Social Psychology, 72*(4), 775-790.

Jensen, A. (1998). *The g factor: The science of mental ability.* London: Praeger.

Johnson, A. W. and Earle, T. K. (2001). *The evolution of human societies: From foraging group to agrarian state.* 2nd ed. Palo Alto, CA: Stanford University Press.

Kaplan, H., Hill, K., Lancaster, J. and Hurtado, A. (2000). A theory of human life history evolution: Diet, intelligence and longevity. *Evolutionary Anthropology, 9*, 156-185.

Kappeler, P. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by P. Kappeler]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Karmiloff-Smith, A. (2006). Modules, genes and evolution: What have we learned from atypical development. In Y. Munakata and M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance series,* Vol. 21 (pp. 563-583). Oxford: Oxford University Press.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge: MIT Press.

Keating, C. F. and Heltman, K. R. (1994). Dominance and deception in children and adults: Are leaders the best misleaders? *Personality and Social Psychology Bulletin, 20*(3), 312-321.

Kerr, N., Dunbar, R. I. M. and Bentall, R. P. (2003). Theory of mind deficits in bipolar affective disorder. *Journal of Affective Disorders, 73*(3), 253-259.

Keulers, E. H. H., Evers, E. A. T., Stiers, P. and Jolles, J. (2010). Age, sex and pubertal phase influence mentalizing about emotions and actions in adolescents. *Developmental Neuropsychology, 35*(5), 555-569.

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

Kinderman, P., Dunbar, R. and Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology, 89*, 191-204.

King, A. J., Narraway, C., Hodgson, L., Weatherill, A., Sommer, V. and Sumner, S. (2011). Performance of human groups in social foraging: The role of communication in consensus decision making. *Biology Letters, 7*, 237-240.

King, A. J. and Sueur, C. (2011). Where next? Group coordination and collective decision-making by primates. *International Journal of Primatology, 32*, 1245-1267.

Kinzler, K. D., Corriveau, K. H. and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*, 106-111.

Kinzler, K. D., Shutts, K., DeJesus, J. and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*(4), 623-634.

Kinzler, K. D., Dupoux, E. and Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*, 12577-12580.

Kirkpatrick, C. (2007). Tactical deception and the great apes: Insight into the question of theory of mind. *Totem: The University of Western Ontario Journal of Anthropology, 15*(4), 1-7.

Kirschner, S. and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior, 31*(5), 354-364.

Klapwijk, E. T., Goddings, A-L., Heyes, S. B., Bird, G., Viner, R. M. and Blakemore, S-J. (2013). Increased functional connectivity with puberty in the mentalising network involved in social emotion processing. *Hormones and Behavior, 64*, 314-322.

Kniffin, K. M. and Wilson, D. S. (2005). Utilities of gossip across organizational levels: Multilevel selection, free-riders, and teams. *Human Nature, 16*, 278-292.

Knight, C. (1998). Ritual/speech coevolution: A solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the evolution of language* (pp. 68-91). New York: Cambridge University Press.

Kolling, N., Behrens, T. E. J., Mars, R. B., and Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science, 336*, 95-98.

Krebs, J. R. and Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs and N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach.* 2nd ed. (pp. 380-402). Oxford: Blackwell Scientific Publications.

Kroodsma, D. E. (1977). Correlates of song organization among North American wrens. *American Naturalist, 111*, 995-1008.

Kurzban, R. (2001). The social psychophysics of cooperation: Nonverbal communication in a public goods game. *Journal of Nonverbal Behavior, 25*(4), 241-259.

Labov, W. (1980). *Locating language in time and space.* New York: Academic Press.

Labov, W. (2001). *Social factors: Principles of linguistic change*, vol. II. Malden, MA: Blackwell.

Lachmann, M., Szamado, S. and Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *PNAS*, *98*(23), 13189-13194.

Laidre, M. E., Lamb, A., Shultz, S. and Olsen, M. (2012). Making sense of information in noisy networks: Human communication, gossip and distortion. *Journal of Theoretical Biology, 317*, 152-160.

Lakin, J. L. and Chartrand, T. L. (2003). Using nonconscious behavioural mimicry to create affiliation and rapport. *Psychological Science, 14*(4), 334-339.

Larson, R. K., Déprez, V. and Yamakido, H. (Eds). (2010). *The evolution of human language.* Cambridge: Cambridge University Press.

Leavitt, M. J. and Mueller, R. A. H. (1955). Some effects of feedback on communication. In A. P. Hare, E. F. Borgatta and R. F. Bales (Eds.), *Small Groups* (pp. 414-433). New York: Knopf.

Lee, K. (2013). Little liars: Development of verbal deception in children. *Child Development Perspectives, 7*(2), 91-96.

Lee, R. B. (2005). *Cambridge encyclopedia of hunters and gatherers.* Cambridge: Cambridge University Press.

Lev-Ari, S. and Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093-96.

Lewis, P., Rezaie, R., Browne, R., Roberts, N. and Dunbar, R. I. M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuro-image, 57*, 1624-1629.

Li, A. S., Kelley, E. A., Evans, A. D. and Lee, K. (2011). Exploring the ability to deceive in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41,* 185-195.

Locke, J. L. and Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences, 29*(3), 259-280.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition, 121*, 289-298.

Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE, 5*(1), 1-10.

Madsen, E. A., Tunney, R. J., Fieldman, G., Plotkin, H. C., Dunbar, R. I. M., Richardson, J. M., McFarland, D. (2007). Kinship and altruism: A cross-cultural experimental study. *British Journal of Psychology, 98*, 339-359.

Maestripieri, D. (1999). Primate social organization, gestural repertoire size, and communication dynamics. A comparative study of macaques. In B. J. King (Ed.), *The origins of language: What nonhuman primates can tell us* (pp. 55-77). Santa Fe, NM: School of American Research Press.

Marlowe, F. (2010). *The Hadza: The hunter-gatherers of Tanzania.* University of California Press.

Marlowe, F. (2007). Hunting and gathering: The human sexual division of foraging labor. *Cross-Cultural Research, 41*(2), 170-195.

Mascie-Taylor, C. G. N. (1988). Assortative mating from psychometric characters. In C. G. N. Mascie-Taylor and A. J. Boyce (Eds.), *Human mating patterns* (pp. 62-82). Cambridge: Cambridge University Press.

Maynard Smith, J. and Harper, D. (2003). *Animal signals.* Oxford: Oxford University Press.

McComb, K., and Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters, 1*, 381-385.

McCracken, G. (1998). *Plenitude: Culture by commotion* (Vol. 1). Toronto, ON: Periph.:Fluide.

Mech, D. L. and Boitani, L. (2003). *Wolves: Behaviour, ecology and conservation*. Chicago: University of Chicago Press.

Mesmer-Magnus, J. R. and DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535-546.

Mesoudi, A. Whiten, A. and Dunbar, R. I. M. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology, 97*(3), 405-423.

Metzger, M. M., Kristof, V. L. and Yoest, D. J. (2003). The World Wide Web and the laboratory: A comparison using face recognition. *Cyberpsychology and Behavior, 6*(6), 613-621.

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information, 1956. *Psychological Review, 101*(2), 343-352.

Miller, G. F. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology, 82*, 97-125.

Miller, G. F. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. London: Heineman.

Miller, G. F. (1999). Sexual selection for cultural displays. In R. I. M. Dunbar, C. Knight, and C. Power (Eds.), *The evolution of culture* (pp. 71-91). Edinburgh: Edinburgh University Press.

Miller, J. (1907). *Light: A narrative poem.* Boston: Herbert B. Turner Press.

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin, 135*(5), 749-773.

Milligan, K., Astington, J. W. and Dack, L. A. (2007). Language and theory of mind: A meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Millikan, R. (1984). *Language, thought and other biological categories.* Cambridge, MA: MIT Press.

Mills, M. G. L. (2003). *Kalahari hyenas: Comparative behavioral ecology of two species.* London: Blackburn Press.

Mitani, J. C. and Watts, D. P. (2001). Why do chimpanzees hunt and share meat? *Animal Behaviour, 61*, 915-924.

Mitani, J. C. and Watts, D. P. (1999). Demographic influences on the hunting behaviour of chimpanzees. *American Journal of Physical Anthropology, 109,* 439-454.

Møller, A. P. (1990). Deceptive use of alarm calls by male swallows, *Hirundo rustica*: A new paternity guard. *Behavioral Ecology, 1*(1), 1-6.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others.* New York: Guilford Press.

Nettle, D. (1999). *Linguistic diversity.* Oxford: Oxford University Press.

Nettle, D., and Dunbar, R. I. M. (1997). Social markers and the evolution of cooperative exchange. *Current Anthropology, 38*(1), 93-99.

Nippold, M. A., Hesketh, L. J., Duthie, J. K. and Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language and Hearing Research, 48*, 1048-1064.

Nippold, M. A., Ward-Lonergan, J. M. and Fanning, J. L. (2005). Persuasive writing in children, adolescents and adults: A study of syntactic, semantic and pragmatic development. *Language, Speech and Hearing Services in Schools, 36*, 125-138.

Nippold, M. A. (2009). School-age children talk about chess: Does knowledge drive syntactic complexity? *Journal of Speech, Language, and Hearing Research, 52*, 856-871.

Nyberg, D. (1993). *The varnished truth: Truth telling and deceiving in ordinary life.* Chicago: University of Chicago Press.

O'Grady, C., Kilesch, C., Smith, K. and Scott-Phillips, T. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior, 36*(4), 313-322.

Oller, D. K. and Griebel, U. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by D. K. Oller and U. Griebel]. *Behavioral and Brain Sciences, 29*(3), 259-325.

O'Neill D. and Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology, 27,* 390-397.

Onishi, K. H., Baillargeon, R. and Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica, 124*(1), 106-128.

Onishi, K. H. and Baillargeon, R. (2005). Do 15-month old infants understand false beliefs? *Science, 308*(5719), 255-258.

Opie, C., Atkinson, Q. D., Dunbar, R. I. M. and Shultz, S. (2013). Male infanticide leads to social monogamy in primates. *Proceedings of the National Academy of Sciences, 110*(33), 13328-13332.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20*, 139-156.

Origgi, G. and Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers and A. Chamberlain (Eds.), *Evolution and the human mind: Language, modularity and social cognition* (pp. 140-169). Cambridge: Cambridge University Press.

Paal, T. and Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences, 43*(4), 541-551.

Packer, C., Scheel, D. and Pusey, A. E. (1990). Why lions form groups: Food is not enough. *American Naturalist, 136,* 1-19.

Palmer, C. T. (1991). Kin selection, reciprocal altruism and information sharing among Maine lobstermen. *Ethology and Sociobiology, 12,* 221-235.

Parker, S. T. (1993). Commentary on Burling's "Primate calls, human language and nonverbal communication". [Peer commentary by S. T. Parker]. *Current Anthropology, 34*(1), 25-53.

Paukner, A., Suomi, S. J., Visalberghi, E., Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science, 325*(5942)*,* 880-883.

Pawlowski, B., and Dunbar, R. I. M. (2001). Human mate choice decisions. In R. Noe, P. Hammerstein, and J. A. R. A. M. van Hooff (Eds.), *Economic models of human and animal behaviour* (pp. 187-202). Cambridge: Cambridge University Press.

Pietraszewski, D. and Schwartz, A. (2014a). Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing. *Evolution and Human Behavior, 35*, 43-50.

Pietraszewski, D. and Schwartz, A. (2014b). Evidence that accent is a dimension of social categorization, not a byproduct of coalitional categorization. *Evolution and Human Behavior, 35*, 51-57.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen and S. Kirby (Eds.)., *Language evolution: States of the art* (pp. 16-37). Oxford: Oxford University Press.

Pinker, S. (1999). *Words and rules*. New York: Harper Perennial.

Pinker, S. (1994). *The language instinct.* New York: William Morrow.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*(4), 707-784.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*, 201-236.

Plomin, R., DeFries, J., McClearn, G., and Rutter, M. (1997). *Behavioral genetics.* (3rd ed.). San Francisco: W. H. Freeman.

Powell, J. L., Lewis, P. A., Dunbar, R. I. M., García-Fiñana, M. and Roberts, N. (2010). Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia, 48*(12), 3554-3562.

Preckel, F. and Thiemann, H. (2003). Online versus paper-pencil version of a high potential intelligence test. *Swiss Journal of Psychology, 62*(2), 131-138.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory-of-mind? *Behavioral and Brain Sciences, 1*, 515-526.

Radick, G. (2007). *The simian tongue: The long debate about animal language.* Chicago: University of Chicago Press.

Ragir, S. and Brooks, P. J. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by S. Ragir and P. J. Brooks]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Rakić, T., Steffens, M. and Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology, 100*(1), 16-29.

Randall, F., Corcoran, R., Day, J. and Bentall, R. (2003). Attention, theory of mind, and causal attributions in people with persecutory delusions: A preliminary investigation. *Cognitive Neuropsychiatry, 8*(4), 287-294.

Rauwolf, P., Mitchell, D. and Bryson, J. (2014). Cooperation benefits when homophily motivates dishonesty in gossip. *European Human Behaviour and Evolution Association Annual Conference,* University of Bristol, Bristol, April 6-9, European Human Behaviour and Evolution Association. Available at: http://www.bristol.ac.uk/expsych/media/ehbea/EHBEA2014FullProgramme.pdf

Redhead, G. and Dunbar, R. I. M. (2013). The functions of language: An experimental study. *Evolutionary Psychology, 11*(4), 845-854.

Rendell, L. and Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences, 24*(2), 309-324.

Richerson, P. J. and Boyd, R. (2010). Why possibly language evolved. *Biolinguistics, 4*(2-3), 289-306.

Roberts, G. (2013). Perspectives on language as a source of social markers. *Language and Linguistics Compass, 7*(12), 619-632.

Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies, 11*(10), 138-159.

Robbins, S. P. and Judge, T. A. (2013). *Essentials of organizational behavior.* (12th ed.). London: Prentice Hall.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* Oxford: Oxford University Press.

Rosenberg, J. and Tunney, R. J. (2008). Human vocabulary use as display. *Evolutionary Psychology, 6*(3), 538-549.

Rushworth, M. F. S, Mars, R. B. and Sallet, J. (2013). Are there specialised circuits for social cognition and are they unique to humans? *Current Opinion in Neurobiology*, *23*, 436-442.

Ryan, M. J. (2013). The importance of integrative biology to sexual selection and communication. In U. Stegmann (Ed.), *Animal communication theory: Information and influence* (pp. 233-255). Cambridge: Cambridge University Press.

Ryan, M. J. and Rand, A. S. (1993). Species recognition and sexual selection as a unitary problem in animal communication. *Evolution, 47*(2), 647-657.

Sadalla, E. K., Kenrick, D. T. and Vershure, B. (1987). Dominance and heterosexual attraction. *Journal of Personality and Social Psychology, 52*(4), 730-738.

Salgado, J. F. and Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment, 11*(2-3), 194-205.

Sakura, O. (1993). Commentary on Locke and Bogin's "Primate calls, human language and nonverbal communication". [Peer commentary by O. Sakura]. *Current Anthropology, 34*(1), 25-53.

Schleidt, W. M. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by W. M. Schleidt]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Schlenker, B. R. and Leary, M. R. (1982). Audiences' reactions to self-enhancing, self-denigrating and accurate self-presentations. *Journal of Experimental Social Psychology, 18,* 89-104.

Scofield, J. and Behrend, D. A. (2011). Clarifying the role of joint attention in early word learning. *First Language, 31*(3), 326-341.

Scott-Phillips, T. C. (2006). Why talk? Speaking as selfish behaviour. In A. Cangelosi, A. D. M. Smith, and K. Smith (Eds.), *The evolution of language: Proceedings of the 6th international conference* (pp. 299-306). London: World Scientific.

Searcy, W. A. and Nowicki, S. (2009). Sexual selection and the evolution of animal signals. In L. R. Squires (Ed.), *Encyclopedia of neuroscience, Vol. 8,* (pp. 769-776). Oxford: Academic Press.

Searcy, W. A. and Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signalling systems*. Princeton: Princeton University Press.

Seijts, G. H. and Latham, G. P. (2000). The effects of goal setting and group size on performance in a social dilemma. *Canadian Journal of Behavioural Science*, *32*(2), 104-116.

Serota, K. B., Levine, T. R. and Boster, F. J. (2010). The prevalence of lying in America: Studies of self-reported lies. *Human Communication Research, 36*(1), 2-25.

Silk, J. B., Cheney, D. L. and Seyfarth, R. M. (2013). A practical guide to the study of social relationships. *Evolutionary Anthropology, 22,* 213-225.

Smith, E. A. (1991). *Inujjuamiut* foraging *strategies: Evolutionary ecology of an arctic hunting economy.* Hawthorn, NY: Aldine de Gruyter.

Smith, M. A. and Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments and Computers, 29*(4), 496-505.

Sommerfeld, R. D., Krambeck, H.-J., Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. *Proceedings of the Royal Society B, 275* (1650), 2529-2536.

Stammbach, E. (1988). Group responses to specially skilled individuals in a *Macaca fascicularis* group. *Behaviour, 107*(3), 241-266.

Stander, P. E. (1992). Cooperative hunting in lions: The role of the individual. *Behavioral Ecology and Sociobiology, 29*, 445-454.

Stanford, C. B. (1998). The social behavior of chimpanzees and bonobos: Empirical evidence and shifting assumptions. *Current Anthropology, 39*(4), 399-420.

Starck, J. (1998). *Avian growth and development*. Oxford Oxfordshire: Oxford University Press.

Stiller, J. and Dunbar, R. I. M. (2007). Perspective-taking and social network size in humans. *Social Networks, 29,* 93-104.

Stromswold, K. (2001). The heritability of language: A review and metanalysis of twin, adoption and linkage studies. *Language, 77*, 647-723.

Surian, L., Caldi, S. and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580-586.

Számadó, S. (2010). Pre-hunt communication provides context for the evolution of early human language. *Biological Theory, 5*(4), 366-382.

Számadó, S. and Szathmáry, E. (2006). Selective scenarios for the emergence of natural language. *Trends in Ecology and Evolution, 21*(10), 555-561.

Talwar, V. and Lee, K. (2008). Little liars: Origins of verbal deception in children. In S. Itakura and K. Fujita (Eds.), *Origins of the social mind: Evolutionary and developmental views* (pp.157-178). New York: Springer.

Taylor, J. D. and Kinderman, P. (2002). An analogue study of attributional complexity, theory of mind deficits and paranoia. *British Journal of Psychology, 93*, 137-140.

Terkel, J. (1996). Cultural transmission of feeding behavior in the black rat (Rattus rattus). In Cecilia M. Heyes and Bennett G. Galef (Eds.), *Social learning in animals: The roots of culture* (pp.17-48). San Diego: Academic Press.

Thornton, S. (1996). *Club cultures: Music, media and subcultural capital.* Middletown, CT: Wesleyan University Press.

Tibbetts, E. A. and Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature, 432*, 218-222.

Toma, C., Hancock, J. T., and Ellison, N. B. (2008). Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin, 34*, 1023-1036.

Tomasello, M. (2009). The usage-based theory of language acquisition. In Edith L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 69-87). Cambridge: Cambridge University Press.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M. and Farrar, J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454-1463.

Tomasello, M., Kruger, A. C. and Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16, 495-552.*

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E. and Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*(6), 673-692.

Tomasello, M. and Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*, 197-212.

Trivers, R. (2014). *Deceit and self-deception: Fooling yourself the better to fool others.* New York: Penguin.

van Schaik, C. P. and Dunbar, R. I. M. (1990). The evolution of monogamy in large primates: A new hypothesis and some crucial tests. *Behaviour, 115*, 30-62.

Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review, 10*(4), 354-371.

Van Vugt, M. and Hart, C. M. (2004). Social identity as social glue: The origins of group loyalty. *Journal of Personality and Social Psychology, 86*(4), 585-598.

Varley, R. and Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology, 10*(12), 723-726.

Voracek, M., Hofhansl A., and Fisher, M. L. (2005). Clark and Hatfield's evidence of women's low receptivity to male strangers′ sexual offers revisited. *Psychological Reports, 97*, 11-20.

Vul, E. and Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*. *19*(7), 645-647.

Waller, B. M., Hope, L., Burrowes, N. and Morrison, E.R. (2011). Twelve (not so) angry men: Managing conversational group size increases perceived contribution by decision makers. *Group Processes and Intergroup Relations, 14*, 835-843.

Warneken, F. Chen, F. and Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*(3), 640-663.

Watson, A., Painter, K. and Bornstein, M. (2002). Longitudinal relations between 2-year olds' language and 4-year-olds' theory of mind. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Journal of Cognition and Development, 2*, 449-457.

Watts, D. P. and Mitani, J. C. (2002). Hunting behaviour of chimpanzees at Ngogo, Kibale National Park, Uganda. *International Journal of Primatology, 23*(1), 1-28.

Weaver, J. R. and Bosson, J. K. (2011). I feel like I know you: Sharing negative attitudes of others promotes feelings of familiarity. *Personality and Social Psychology Bulletin, 37*(4), 481-491.

Weisfeld, G. E. (2006). Commentary on Locke and Bogin's "Language and life history: A new perspective on the development and evolution of human language". [Peer commentary by G. E. Weisfeld]. *Behavioral and Brain Sciences, 29*(3), 259-325.

Wellman, H. M., Cross, D. and Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655-684.

Wheeler, B. C. (2009). Monkeys crying wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics. *Proceedings of the Royal Society B, 276,* 3013-3018.

Whiten, A. and Byrne, R. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Oxford University Press.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y. Tutin, C. E. G., Wrangham, R. W. and Boesch, C. (1999). Cultures in chimpanzees. *Nature, 399*, 682-685.

Whittingham, L. A., Kirkconnell, A. and Ratcliffe, L. M. (1997). The context and function of duet and solo songs in the red-shouldered blackbird. *Wilson Bulletin, 109*(2), 279-289.

Wilkinson, G. S. (2003). Social and vocal complexity in bats. In F. B. M. DeWaal and P. L. Tyack (Eds.), *Animal social complexity: Intelligence, culture and individualized societies* (pp. 322-341). Cambridge, MA: Harvard University Press.

Willems, R. M., Benn, Y., Hagoort, P., Toni, I. and Varley, R. (2011). Communicating without a functioning language system: Implications for the role of language in mentalizing. *Neuropsychologia, 49*, 3130-3135.

Williams, D. L. Goldstein, G. and Minshew, N. J. (2006). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology, 12*(4-5), 279-298.

Wilson, D. S., Wilczynski, C., Wells, A., and Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In C. Heyes and L. Huber (Eds.), *The evolution of cognition* (pp. 347-366). Cambridge, MA: MIT Press.

Wiltermuth, S. S. and Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5.

Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Wittenbaum, G. M. (1998). Information sampling in decision-making groups: The impact of members' task-relevant status. *Small Group Research, 29*(1), 57-84.

Wright, T. F. and Wilkinson, G. S. (2001). Population genetic structure and vocal dialects in an Amazon parrot. *Proceedings of the Royal Society B*, *268*, 609-616.

Yurk, H., Barrett-Lennard, L., Ford, J. K. B., and Matkin, C. O. (2002). Cultural transmission within maternal lineages: Vocal clans in resident killer whales in southern Alaska. *Animal Behaviour*, *63*, 1103-1119.

Zahavi, A. (1975). Mate selection – selection for a handicap. *Journal of Theoretical Biology, 53,* 205-214.

Zahavi, A. and Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle.* New York: Oxford University Press.

Zihlman, A. and Tanner, N. M. (2001). Gathering and the hominid adaptation. *Anthropology Origins, 10*(99), 163-194.

Zhou, W.-X., Sornette, D., Hill, R. A., and Dunbar R. I. M. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B, 272,* 439-444.

Zubieta, J. K., Smith, Y. R., Bueller, J. A., Xu, Y., Kilbourn, M. R., Jewett, D. M., Meyer, C. R., Koeppe, R. A., Stohler, C. S. (2001). Regional mu opioid receptor regulation of sensory and affective dimensions of pain. *Science, 293*(5528), 311-315.

Appendix A.2: Chapter 2

Appendix A.2.1 Social foraging questionnaire

Age _____ Sex _____ Height _____ Ethnicity _____ Occupation _____

Highest level of education: *Secondary/A-Level* ☐ *BA/BS* ☐ *MA/MS/MPhil* ☐ *PhD/DPhil* ☐ *MD* ☐ *JD* ☐

			please circle the number that applies				
	VERY LITTLE			**SOMEWHAT**			**A LOT**
How cooperative and helpful were you to your group members?	1	2	3	4	5	6	7
How much cooperative and helpful behaviour did the group provide?	1	2	3	4	5	6	7
How much social pressure did you feel from the group to do well?	1	2	3	4	5	6	7
How much encouragement did you provide?	1	2	3	4	5	6	7
How much encouragement did the group provide?	1	2	3	4	5	6	7
How much did you talk about the task with the group?	1	2	3	4	5	6	7
How much suggestion or opinion did you provide to the group?	1	2	3	4	5	6	7
How much did you ask for opinions or suggestions from the group?	1	2	3	4	5	6	7
How much do you think you gestured (hand-waving, pointing) to the group?	1	2	3	4	5	6	7
How much did you see others gesturing (hand-waving, pointing) to you?	1	2	3	4	5	6	7
How much did you help plan and organise task duties for the group?	1	2	3	4	5	6	7
How much did someone else help plan and organise task duties for the group?	1	2	3	4	5	6	7
How confused were you about how best to complete the task?	1	2	3	4	5	6	7
How much do you believe the group was confused about how best to complete the task?	1	2	3	4	5	6	7
How much did you disagree with the group about how to complete the task?	1	2	3	4	5	6	7
How much did your group disagree about how to complete the task?	1	2	3	4	5	6	7
How well did you know each person in the group?	1	2	3	4	5	6	7
How much did you trust other people in the group?	1	2	3	4	5	6	7
How friendly did you feel toward other people in the group?	1	2	3	4	5	6	7
How close did you feel toward other people in the group?	1	2	3	4	5	6	7
Do you consider the people in the group to be charitable and kind?	1	2	3	4	5	6	7
Did you know the name of other people in your group?	1	2	3	4	5	6	7
How motivated were you to work with the group?	1	2	3	4	5	6	7
How motivated were you to do the task?	1	2	3	4	5	6	7
How proud would you be to tell others you were "happy to be a part" of the group?	1	2	3	4	5	6	7

Appendix A.3: Chapter 3

Appendix A.3.1 Male protagonist small and large group public speaking vignettes. Respective word counts were matched to be approximately equal and are presented at the end of each description. Female protagonist vignettes were identical apart from names and relevant pronouns.

A.3.1.1. Version 1: Small group public speaker

Steve socially awkward:

> Steve is 5'11" tall, 170 lbs. He is an articulate speaker and frequently captures the attention of friends, relatives and strangers. Although his arguments are generally well-reasoned and cogent, he tends to shy away from the spotlight and sometimes buckles under social pressure, forgetting the stories he intended to present. He sometimes finds himself in compromising situations. He recently led a less than engaging discussion with a small group of people who remarked he was a bit unrehearsed and awkward. (80)

Tim socially shy (filler):

> *Tim is 5'9" tall, 160 lbs. He is a friendly individual and enjoys interacting with other people. He is often sociable and a well-liked person and often attends social gatherings. Although he is a nice person, he is normally quite shy and rarely speaks up unless he has something important to say. He is generally very polite and courteous. He recently led a discussion with a small group of people and agreed with many of the points made by the other people. (82)*

John socially clever (filler):

> *John is 5'10" tall, 165 lbs. He is a gifted conversationalist and has many acquaintances and several close friends. He is a generally sociable and well-liked person and often attends social gatherings with friends. He is not the most talkative person, but when he does speak up, he often voices intelligent and profound observations on politics, science, sociology and philosophy. He is generally thought to be brilliant among close friends. He recently led a successful and engaging discussion with a small group of people. (84)*

Steve socially relevant:

> Steve is 5'11" tall, 170 lbs. He is an articulate speaker and frequently captures the attention of friends, relatives and strangers. Although his arguments are generally well-reasoned and cogent, he tends to shy away from the spotlight and often disfavors the attention brought by being the center of attention. He generally chooses relevance and content over delivery and showmanship. He recently led an engaging discussion with a small group of people who remarked he was a fascinating and charismatic speaker. (80)

A.3.1.2. Version 2: Large group public speaker

Steve socially awkward:

> Steve is 5'11" tall, 170 lbs. He is a talented public speaker and has appeared at many live speaking events. Although his arguments are generally well-reasoned and cogent, he tends to shy away from the spotlight and sometimes buckles under social pressure, forgetting the material he intended to present. He sometimes finds himself in compromising situations. He recently presented a less than engaging lecture to an audience of 500 people who remarked he was a bit unrehearsed and awkward. (79)

Tim socially shy (filler):

> *Tim is 5'9" tall, 160 lbs. He is a highly skilled orator and frequent guest lecturer at many different live speaking events. Although he is a nice person, he is normally quite shy and rarely speaks up unless he has something important to say. He is generally very polite and courteous. He recently presented an engaging lecture to an audience of several hundred people who remarked he was a reserved but articulate speaker. (73)*

John socially clever (filler):

> *John is 5'10" tall, 165 lbs. He is a very articulate speaker and regularly appears on live debates and public lectures. He is not the most talkative person among his friends, but when he does speak up, he often voices intelligent and profound observations on politics, science, sociology and philosophy. He is generally thought to be brilliant among close friends. He recently presented a brilliant and engaging lecture to a large auditorium of more than 300 people. (77)*

Steve socially relevant:

> Steve is 5'11" tall, 170 lbs. He is a talented public speaker and has appeared at many live speaking events. Although his arguments are generally well-reasoned and cogent, he tends to shy away from the spotlight and often disfavors the attention brought by public speaking. He generally chooses relevance and content over delivery and showmanship. He recently presented an engaging lecture to an audience of 500 people who remarked he was a fascinating and charismatic speaker. (76)

Appendix A.3.2. Male protagonist vignettes with before and after descriptions. Respective word counts were matched to be approximately equal and are presented at the end of each description. Female protagonist vignettes were identical apart from relevant pronouns.

A.3.2.1. Version 1: Self-enhancing attorney before/after description

I am a happy and successful trial attorney. I have a lovely palatial estate in Beverly Hills. I would be delighted to take you out on a romantic evening dinner on my 40-foot yacht. I very much enjoy children and look forward to having kids. I would describe myself as quite tall and lean and I am often told to be quite handsome. I enjoy hiking, basketball, water sports, hiking, and in general most outdoor activities. (76)

After reading over his profile, you are intrigued and agree to a date. He invites you out to his house for a nice evening dinner. Surprisingly, upon arriving at his address, you discover that he does not in fact live on a "palatial" estate. Instead, it seems to be a fairly typical middle-class suburban neighbourhood. Once inside, you notice the house is warm and inviting, yet not particularly large. It is nicely decorated, with a small kitchen and pleasant living area. You don't notice a body of water around and wonder about his yacht. Glancing him over, you judge him to be not especially tall and fairly average-looking in appearance. He takes you on a brief stroll through his study where you find an assortment of books and periodicals; the Wall Street Journal sits on his desk and his shelves are stacked with books on physics, biology, philosophy, art, music and travel. Intrigued, you ask him a bit more about his hobbies. He remarks that he enjoys reading, the outdoors, basketball and hiking. You eventually sit down to dinner and you ask him about whether there is a lake near-by where he keeps his yacht. He apologises and confesses to you that he does not in fact own a yacht. Furthermore, you also learn that he is not in fact an attorney, but a successful paralegal with aspirations to soon start law school. When you ask him about the misleading profile, he apologizes again and confesses that he was only trying to attract attention to his profile. Despite the otherwise awkward introduction, you otherwise have a very relaxing and exciting evening over a delicious dinner as you trade stories about various hiking and travel excursions. (285)

A.3.2.2. Version 1: Self-enhancing janitor before/after description

I am a happy CEO of a successful cleaning business. I have a lovely palatial estate in Beverly Hills. I would be delighted to take you out on a romantic evening dinner on my 40-foot yacht. I very much enjoy children and look forward to having kids. I would describe myself as quite tall and lean and I am often told to be quite handsome. I enjoy hiking, basketball, water sports, hiking, and in general most outdoor activities. (78)

After reading over his profile, you are intrigued and agree to a date. He invites you out to his house for a nice evening dinner. Surprisingly, upon arriving at his address, you discover that he does not in fact live on a "palatial" estate. Instead, it seems to be a fairly typical middle-class suburban neighbourhood. Once inside, you notice the house is warm and inviting, yet not particularly large. It is nicely decorated, with a small kitchen and pleasant living area. You don't notice a body of water around and wonder about his yacht. Glancing him over, you judge him to be not especially tall and fairly average-looking in appearance. He takes you on a brief stroll through his study where you find an assortment of books and periodicals; the Wall Street Journal sits on his desk and his shelves are stacked with books on physics, biology, philosophy, art, music and travel. Intrigued, you ask him a bit more about his hobbies. He remarks that he enjoys reading, the outdoors, basketball and hiking. You eventually sit down to dinner and you ask him about whether there is a lake near-by where he keeps his yacht. He apologises and confesses to you that he does not in fact own a yacht. Furthermore, you also learn that he is not in fact a CEO, but a successful janitor with aspirations to start his own cleaning business. When you ask him about the misleading profile, he apologizes again and confesses that he was only trying to attract attention to his profile. Despite the otherwise awkward introduction, you otherwise have a very relaxing and exciting evening over a delicious dinner as you trade stories about various hiking and travel excursions. (286)

A.3.2.3. Version 2: Self-deprecating attorney before/after description

I am an unemployed attorney. I live in a small cottage on the outside of town. I am a fairly decent cook and enjoy cooking for dinner guests. I also like animals and have a very friendly dog named Oliver to keep me company. I am not terribly active or athletic and enjoy watching television at home. I am slightly below average height, and slightly overweight, although I wouldn't describe myself as horrendously unattractive. (74)

After reading over his profile, you are intrigued and agree to a date. He invites you out to his house for a nice evening dinner. Upon arriving at his address, he invites you into his house. Once inside, you notice that far from being a "small cottage" it's actually quite large, spacious and open. It is decorated very nicely, with a large kitchen and comfortable living room. He introduces you to his dog Oliver who is a very friendly bloodhound. Glancing him over, you find him to be much taller than he described in his profile. He takes you on a brief tour of his personal library which you find to be an impressive collection; you see books on philosophy, history, science, religion and various works of poetry and nonfiction. You eventually sit down to dinner and you ask him about his hobbies, likes and dislikes. He remarks that he enjoys the outdoors, skiing and hiking. A bit surprised by the apparent modesty of his profile, you ask him about his job. You learn that he in fact recently quit his job at a very successful law firm to continue his own private law practice currently earning three times his law firm salary. You are prompted to ask him about his somewhat misleading profile; he explains that he deliberately down-played his work situation so as to avoid dating people only interested in his success, and attract people more interested in him as a person. As it happens, it turns out he is quite a skilled cook and apparently once worked as a staff chef in his father's catering business. The two of you have a very relaxing and enjoyable evening over a delicious dinner as you trade stories about various hiking and travel excursions. (294)

A.3.2.4. Vignette Version 2: Self-deprecating janitor before/after description

I am an unemployed janitor. I live in a small cottage on the outside of town. I am a fairly decent cook and enjoy cooking for dinner guests. I also like animals and have a very friendly dog named Oliver to keep me company. I am not terribly active or athletic and enjoy watching television at home. I am slightly below average height, and slightly overweight, although I wouldn't describe myself as horrendously unattractive. (74)

After reading over his profile, you are intrigued and agree to a date. He invites you out to his house for a nice evening dinner. Upon arriving at his address, he invites you into his house. Once inside, you notice that far from being a "small cottage" cottage it's actually quite large, spacious and open. It is decorated very nicely, with a large kitchen and comfortable living room. He introduces you to his dog Oliver who is a very friendly bloodhound. Glancing him over, you find him to be much taller than he described in his profile. He takes you on a brief tour of his personal library which you find to be an impressive collection; you see books on philosophy, history, science, religion and various works of poetry and nonfiction. You eventually sit down to dinner and you ask him about his hobbies, likes and dislikes. He remarks that he enjoys the outdoors, skiing and hiking. A bit surprised by the apparent modesty of his profile, you ask him about his job. You learn that he is in fact the CEO of a successful cleaning business that is currently downsizing to be restarted at the end of the year. You are prompted to ask him about his somewhat misleading profile; he explains that he deliberately downplayed his work situation so as to avoid dating people only interested in his success, and attract people more interested in him as a person. As it happens, it turns out he is quite a skilled cook and apparently once worked as a staff chef in his father's catering business. The two of you have a very relaxing and enjoyable evening over a delicious dinner as you trade stories about various hiking and travel excursions. (291)

Appendix A.4: Chapter 4

Appendix A.4.1. Socio-cognitive mentalising task (revised from Stiller and Dunbar, 2007).

Provided here are details of all three stories, as presented to participants and the list of mentalising, memory, and recursive syntax questions participants were then asked to complete. Subscripts count the number of metarepresentations for the mentalising questions, simple clauses for the memory questions and embedded clauses for the recursive syntax questions. Although included here for purposes of illustration, subscripts were excluded from the actual questions as presented to participants.

Participants were asked to indicate which answer they thought was the correct one for each question: true (T) or false (F). Memory questions are indicated by the letter *m*, recursive syntax questions are indicated by an *s* and intentionality questions are indicated by an *i*; each letter designation is preceded by the number of agents for intentionality questions, simple clauses for memory questions, and embedded clauses for recursive syntax questions.

A.4.1.1. Martin, Simon, Charlotte and Jane's story

Charlotte and Simon decide to arrange a blind date for their two friends. Simon's friend Martin is a florist and Charlotte's friend Jane is a doctor. Charlotte and Simon decide to go to the pub on the edge of town, as it is within walking distance of their house and a convenient place to meet Jane and Martin, after their blind date over dinner at a nearby restaurant. When Charlotte and Simon arrive at the pub, they greet Martin and Jane, and Martin buys a round of drinks. Charlotte and Jane both have a glass of red wine while Martin has a pint of beer and Simon has a whisky. Martin and Simon leave the women at the bar to play a game of pool. Jane tells Charlotte that she did not like Martin as she thinks he is a chauvinist. Meanwhile Martin tells Simon that the date went well, Jane fancied him and that he was the perfect gentleman. Simon wins the game and they return to the girls. Jane thanks Martin for a lovely date and decides to go home early.

A.4.1.1.1. Questions

1 (a) Jane thinks$_1$ the date went well. (1i) **F**
(b) Charlotte and Simon arrange$_1$ a blind date. (1s) **T**
(c) Martin is$_1$ a qualified and competent doctor. (1m) **F**

2 (a) Martin thinks$_1$ that Jane believes$_2$ the date went well for both of them. (2i) **T**
(b) Martin meets Charlotte and Simon at the pub,$_1$ at which Martin buys drinks.$_2$ (2s) **T**
(c) Charlotte and Simon go to a pub;$_1$ the pub is on the edge of town.$_2$ (2m) **T**

3 (a) Martin knows$_1$ that Charlotte believes$_2$ that Jane thinks$_3$ the restaurant and bar date went badly for all concerned. (3i) **F**
(b) The two went on a blind date,$_1$ that was arranged by Martin and Jane,$_2$ who met Charlotte and Simon at the pub.$_3$ (3s) **F**
(c) Martin went on a blind date,$_1$ the florist's name was Jane,$_2$ and they met at the pub on the edge of town.$_3$ (3m) **F**

4 (a) Simon knows$_1$ that Martin believes$_2$ that Charlotte supposes$_3$ that Jane thinks$_4$ the restaurant and pub date went well for all concerned. (4i) **T**
(b) Charlotte and Simon meet Martin,$_1$ who buys a round of drinks,$_2$ who plays pool with Simon,$_3$ who wins the game.$_4$ (4s) **T**
(c) Martin buys a round of drinks;$_1$ Jane has a glass of red wine,$_2$ Martin has a beer,$_3$ and Simon has a whisky.$_4$ (4m) **T**

5 (a) Simon believes$_1$ that Martin thinks$_2$ that Charlotte supposes$_3$ that Jane knows$_4$ that Simon thinks$_5$ they had a lovely date at the restaurant and a relaxing evening at the bar. (5i) **F**
(b) Simon is a friend of Charlotte,$_1$ who is a friend of Jane,$_2$ who is a doctor,$_3$ who goes on a date with Martin,$_4$ who gets told he is a chauvinist.$_5$ (5s) **F**
(c) Martin is a florist,$_1$ Jane is a doctor,$_2$ Simon goes out to the pub with Charlotte,$_3$ Martin plays pool with Simon,$_4$ and Simon wins the game of pool.$_5$ (5m) **T**

6 (a) Martin thinks$_1$ Charlotte thought$_2$ that Simon knew$_3$ Martin thought$_4$ Jane knew$_5$ Martin believed$_6$ the pub was a good place to meet Charlotte and Simon for a relaxing and enjoyable evening and an engaging game of pool. (6i) **F**
(b) Martin goes on a date with Jane,$_1$ who at the pub meets Charlotte,$_2$ who is a friend of Simon,$_3$ who plays pool with Martin,$_4$ despite being on a date with Jane,$_5$ who nonetheless thanks Martin for taking her out on the date.$_6$ (6s) **T**
(c) Charlotte and Simon go to the pub,$_1$ Martin buys a round of drinks,$_2$ Charlotte has a glass of wine,$_3$ Simon has a beer,$_4$ Martin has a whisky,$_5$ and Simon wins the game of pool.$_6$ (6m) **F**

7 (a) After the evening they had all spent together, Simon thought$_1$ that Charlotte believed$_2$ that Jane knew$_3$ that Martin was confident$_4$ that Jane believed$_5$ that Charlotte fully understood$_6$ Simon's belief$_7$ that they had all had a lovely meal at the restaurant followed by a rather relaxing evening at the bar where everyone had shared in a round of drinks. (7i) **F**
(b) Charlotte and Simon arrange a blind date for their friends;$_1$ Martin is a florist and Jane is a doctor,$_2$ whom they meet for drinks at the pub,$_3$ where Simon plays a game of pool with Martin,$_4$ whom Jane thanks for the lovely date,$_5$ that Jane and Martin had before going to the pub,$_6$ where Simon won the pool game.$_7$ (7s) **T**
(c) Martin is a florist,$_1$ Charlotte's friend$_2$ Jane is a doctor,$_3$ and Charlotte and Simon are roommates;$_4$ Martin buys a round of drinks at the pub, including two glasses of red wine, a pint of beer, and a glass of whisky,$_5$ while Martin and Simon play a game of pool at the pub,$_6$ and Simon wins the game of pool over Martin.$_7$ (7m) **T**

8 (a) Charlotte certainly thought$_1$ for several good reasons that Martin surely knew,$_2$ with regard to Jane and Simon's date, about Simon's clear belief$_3$ that Jane was of the opinion$_4$ that Charlotte most likely believed$_5$ in more ways than one that Martin probably suspected$_6$ that Simon definitely thought,$_7$ for presumably good reasons in his own mind from the date, that Jane undoubtedly fancied$_8$ him romantically and not Martin. (8i) **T**
(b) Charlotte and Simon decide to go to the pub on the edge of town,$_1$ that is within walking distance of their house,$_2$ which is a convenient place to meet Jane and Martin,$_3$ who arrive after their blind date at the pub,$_4$ where Martin and Simon played a game of pool,$_5$ that Simon won while the girls sat at the bar,$_6$ where they gossiped about Martin,$_7$ who went home early.$_8$ (8s) **F**
(c) Martin is a florist,$_1$ Charlotte's friend$_2$ Jane is a doctor,$_3$ Charlotte and Simon are neighbours and close friends,$_4$ Martin buys a round of drinks at the pub, including two glasses of red wine, a pint of beer, and a glass of whisky,$_5$ Martin and Simon play a game of pool at the pub,$_6$ Simon wins the game of pool over Martin,$_7$ after which they rejoined the girls.$_8$ (8m) **F**

9 (a) Simon was certainly convinced$_1$ that Charlotte probably knew,$_2$ for several good reasons, that Jane suspected$_3$ with respect to her date with Simon that Charlotte most likely realised$_4$ that Simon didn't believe$_5$ that Martin understood$_6$ at least in some respect that Simon supposed $_7$ that, based on his experience during the date, that Charlotte was convinced$_8$ that, Jane thought $_9$ that Martin was the most awful chauvinist. (9i) **T**
(b) Charlotte and Jane both have a glass of red wine,$_1$ that Martin purchased as part of a round of drinks,$_2$ which included a beer for Martin and a whisky for Simon,$_3$ that they drank while Martin and Simon played pool,$_4$ which Martin won,$_5$ that they played in the bar,$_6$ that was located on the edge of town,$_7$ that was walking distance from their house,$_8$ where Charlotte and Simon live together.$_9$ (9s) **F**
(c) Martin is a florist,$_1$ Charlotte's friend$_2$ Jane is a doctor,$_3$ Charlotte and Simon are neighbours and close friends,$_4$ and when the boys met Jane and Charlotte at the pub,$_5$ Martin bought a round of drinks, including two glasses of red wine, a beer, and a whisky,$_6$ and Martin and Simon played a game of pool,$_7$ which Simon won,$_8$ after which they returned to the girls.$_9$ (9m) **F**

A.4.1.2 Frank, Betty and Brian's story

Frank, Betty and Brian all share a three-bedroom house situated close to the high street. Frank is very lazy and never helps with the housework. One morning Brian and Betty decide that they should make him go and do the weekly shop. Brian walks into Frank's bedroom and wakes him up and tells Frank what they need. Betty would like some eggs and bacon for a cooked breakfast and Brian would like lettuce and tomatoes to make a salad for lunch. Frank says he can remember the list so Brian gives Frank £10 and leaves the house and goes to work. Betty reminds Frank that she only wants him to buy free range or organic produce. Frank becomes confused so Betty writes down the shopping list for him and tells him that Brian will not mind what type of lettuce Frank buys. Frank gets dressed and walks to the shops. When Frank gets to the supermarket he finds he has left the shopping list in his dressing gown and becomes worried that Brian might shout at him.

A.4.1.2.1 Questions

1 (a) Frank knew$_1$ Betty might shout at him. (1i) **F**
(b) Frank walks$_1$ to the shops with the list. (1s) **F**
(c) Betty, Brian and Margaret live$_1$ in a house. (1m) **F**

2 (a) Brian believed$_1$ that Frank was convinced$_2$ the type of lettuce was unimportant. (2i) **F**
(b) Frank is asked$_1$ to shop$_2$ for eggs, bacon, lettuce, tomatoes and carrots. (2s) **F**
(c) Betty asked for eggs and bacon$_1$ and Brian asked for lettuce and tomatoes$_2$. (2m) **T**

3 (a) Betty knew$_1$ that Brian was worried$_2$ that Frank might think$_3$ he was stupid and shout at him for incompetence. (3i) **F**
(b) Betty asked for bacon and eggs for breakfast,$_1$ to be purchased by Frank,$_2$ who did not remember the list$_3$. (3s) **T**
(c) Brian asked$_1$ for breakfast foods, Betty asked$_2$ for lunch foods, and Frank was assigned$_3$ to shopping for the entire house. (3m) **F**

4 (a) Brian knew$_1$ Frank thought$_2$ that Brian believed$_3$ that Betty wanted$_4$ some ingredients for breakfast. (4i) **T**
(b) Frank becomes$_1$ confused by Betty,$_2$ who tells him not to worry about Brian,$_3$ who has no preference for organic produce.$_4$ (4s) **T**
(c) Betty walks$_1$ into Frank's bedroom, and reminds him$_2$ about shopping, so Frank gets$_3$ dressed.$_4$ (4m) **F**

5 (a) Frank believed$_1$ that Betty understood$_2$ that Brian was of the opinion$_3$ that Betty thought$_4$ that Frank believed$_5$ all along that Brian was lazy about house duties. (5i) **F**
(b) Betty is a roommate of Brian,$_1$ who walked into the bedroom of Frank,$_2$ who received £10 from Brian,$_3$ who likes lettuce$_4$ and tomatoes for salad.$_5$ (5s) **F**
(c) The friends share a house,$_1$ but Frank is lazy,$_2$ so Brian gives him £10,$_3$ but this makes Frank confused,$_4$ so Brian writes a list.$_5$ (5m) **F**

6 (a) Frank understands$_1$ that Brian wanted$_2$ Betty to believe$_3$ that Brian maintains$_4$ that Frank knows$_5$ that Betty believes$_6$ that Frank is very lazy about house duties. (6i) **T**
(b) Frank takes the list$_1$ that Betty wrote down for him,$_2$ who said$_3$ he could remember the list,$_4$ that got left at home,$_5$ that is situated on High Street.$_6$ (6s) **T**
(c) Betty wanted breakfast,$_1$ and Brian wanted lunch,$_2$ so Frank was assigned$_3$ to do the shopping; thus, Brian gave Frank £10,$_4$ but Betty had to remind$_5$ Frank before he got dressed for the shops.$_6$ (6m) **T**

7 (a) Brian understood$_1$ Frank believed$_2$ that Betty knew$_3$ Frank worried$_4$ that Brian realised$_5$ that Frank was confused$_6$ as to whether Betty wanted$_7$ him to buy free range or organic produce from the shop. (7i) **F**
(b) Brian wakes Frank,$_1$ who shares a house with Betty,$_2$ who reminds Frank$_3$ that she only wants free range or organic produce,$_4$ which confuses Frank,$_5$ which causes Betty$_6$ to write down the shopping list.$_7$ (7s) **T**

(c) Frank is asleep,$_1$ Brain walks into Frank's bedroom,$_2$ Brian gives £10 to Frank,$_3$ Brian leaves for work soon after,$_4$ Betty reminds Frank,$_5$ Frank becomes confused,$_6$ and Frank gets dressed.$_7$ (7m) **F**
8 (a) Betty knew$_1$ Frank wanted$_2$ Brian to understand$_3$ that Frank thought$_4$ that Betty believed$_5$ that Frank wanted$_6$ Brian to think$_7$ that Frank would remember$_8$ the shopping list which contained the ingredients for a cooked breakfast of bacon and eggs for Betty and a lunch salad with lettuce and tomatoes for Brian. (8i) **T**
(b) Betty shares a house with Brian,$_1$ who gives £10 to Frank,$_2$ who becomes confused by Betty,$_3$ who reminds Frank about the shopping,$_4$ that will provide lunch for Brian,$_5$ who Frank is worried might shout at him,$_6$ because he doesn't help with the housework,$_7$ of which Betty and Brian want Frank to help contribute.$_8$ (8s) **T**
(c) Betty and Brian decide upon a joint solution,$_1$ Brian wakes Frank,$_2$ Frank claims he will remember the list,$_3$ Brian gives Frank £10,$_4$ and Brian leaves the house for work;$_5$ later, Betty reminds Frank about the shopping,$_6$ but Frank becomes confused and worries that Brian might shout at him,$_7$ and so Betty writes down the list.$_8$ (8m) **T**

9 (a) Frank realised$_1$ that Betty wanted$_2$ Brian to think$_3$ that Frank believed$_4$ that Betty knew$_5$ Frank was worried$_6$ that Brian knew$_7$ Frank was confused$_8$ with respect to Betty's desire$_9$ that Frank buy free range or organic produce for a cooked breakfast with bacon and eggs for Betty and a lunch salad with lettuce and tomatoes for Brian. (9i) **F**
(b) Frank takes the list from Betty,$_1$ who had decided the shopping should be done by Frank,$_2$ who never before helped Betty,$_3$ who wanted some eggs and bacon for a cooked breakfast,$_4$ that needed specific produce,$_5$ that he was to purchase with £10 from Brian,$_6$ who woke up Frank,$_7$ who didn't remember the list,$_8$ that was written down by Brian.$_9$ (9s) **F**
(c) Betty and Brian agree on a joint solution:$_1$ Brian wakes Frank,$_2$ Brian gives Frank £10 for the shopping,$_3$ Frank claims he will remember the list,$_4$ and Brian leaves the house for work;$_5$ later, Betty reminds Frank of her request,$_6$ and when Frank worries Brian might shout at him,$_7$ Betty writes down the list,$_8$ but Frank forgets the list.$_9$ (9m) **T**

A.4.1.3. Gavin, Peter, Fiona and Sophie's story

Gavin, Peter, Fiona and Sophie are all waiting for a job interview. They are applying to become an assistant editor of a high profile cookery magazine based in London. Gavin is a trained chef, Fiona has a degree in food technology, and Sophie has a degree in English literature. Peter has no experience in editing or cooking, but does have a diploma in computing. Gavin engages them all in conversation to assess the competition. He tells them he is a chef in a high profile restaurant and worked with food for the past 10 years. Fiona is next to respond and describes her degree course in food technology. Sophie mentions that she enjoys cooking and has a degree in English. Gavin laughs at Sophie and wishes her luck. Sophie, Peter and Fiona all think Gavin is rather rude. Gavin asks Peter what qualifications he has and Peter refuses to answer as he is embarrassed by his lack of relevant skills. Gavin suggests that Peter is refusing to answer, as Gavin believes he is obviously the most qualified. Peter again refuses to reply so as not to give any information away and ignores Gavin. The two women tell Gavin to be quiet.

A.4.1.3.1. Questions

1 (a) Peter thinks$_1$ Gavin is actually rather rude. (1i) **T**
(b) Gavin has applied$_1$ to become an editor in Tokyo. (1s) **F**
(c) Sophie has a degree$_1$ in English literature. (1m) **T**

2 (a) Gavin thinks$_1$ that Sophie believes$_2$ that Peter should keep his opinions to himself. (2i) **F**
(b) Gavin is a trained chef$_1$ that has at least ten years of cooking experience.$_2$ (2s) **T**
(c) Sophie mentions she enjoys cooking$_1$ and has a degree in English literature.$_2$ (2m) **T**

3 (a) Sophie believes$_1$ Fiona knows$_2$ that Peter thinks$_3$ Gavin is rather rude in the way he communicates his opinions. (3i) **T**
(b) Peter is a computer programmer,$_1$ who refuses to answer Sophie,$_2$ who interrogates Peter on his alleged lack of qualifications.$_3$ (3s) **F**
(c) Gavin is a trained chef,$_1$ Sophie has a degree in food technology,$_2$ and Fiona has a degree in English literature.$_3$ (3m) **F**

4 (a) Peter thinks$_1$ that Gavin knows$_2$ that Sophie believes$_3$ that Fiona thinks$_4$ that Peter is the least qualified of the four applicants at the job interview for assistant editor. (4i) **F**
(b) Fiona is a food technology specialist,$_1$ who shows up to the interview with Peter, Sophie and Gavin,$_2$ who immediately laughs at Sophie,$_3$ who tells Gavin he is a reasonable fellow.$_4$ (4s) **F**
(c) Gavin is a trained chef,$_1$ Fiona has a degree in food technology,$_2$ Sophie has a degree in English literature,$_3$ and Peter has no experience in editing or cooking.$_4$ (4m) **T**

5 (a) Gavin knows$_1$ that Peter understands$_2$ that Fiona supposes$_3$ that Gavin knows$_4$ that Sophie thinks$_5$ it would be helpful if Gavin adopted a less condescending and superior attitude toward the other applicants. (5i) **T**
(b) Sophie is an interviewee,$_1$ who is an English literature graduate,$_2$ who is wished good luck by Gavin,$_3$ who laughs at Sophie,$_4$ who is waiting with the three others for the job interview.$_5$ (5s) **T**
(c) Fiona has studied food technology,$_1$ Sophie has studied English literature,$_2$ Peter has studied computing,$_3$ Gavin has trained as a chef,$_4$ and they have all come for the same interview as assistant editor.$_5$ (5m) **T**

6 (a) Fiona believes$_1$ that Gavin thinks$_2$ that Sophie likely knows$_3$ that Peter is undeniably rather rude to believe$_4$ that Gavin supposes$_5$ that he knows less than Peter does$_6$ about cooking and has fewer qualifications for a major high profile job at a cookery magazine based in London. (6i) **F**
(b) Fiona is a food technology specialist,$_1$ who shows up to the interview with Peter, Sophie and also Gavin,$_2$ who immediately laughs at Sophie,$_3$ who is nonetheless wished good luck by Gavin,$_4$ who is told to be quiet by Peter,$_5$ who has a diploma in computing.$_6$ (6s) **F**
(c) Gavin engages everyone in conversation,$_1$ Fiona has studied food technology,$_2$ Sophie mentions that she enjoys cooking,$_3$ Gavin has studied computing,$_4$ Peter has trained as a chef,$_5$ and they have all come for the same interview as assistant editor of a high profile cookery magazine.$_6$ (6m) **F**

7 (a) Peter certainly seems to believe$_1$ that Fiona probably thinks$_2$ that Sophie likely knows$_3$ that Peter actually thinks$_4$ that he is unfortunately embarrassed by Gavin's resolute feeling$_5$ that Peter probably believes$_6$ that he is not quite as qualified as Gavin is inclined to think$_7$ he is with respect to the job. (7i) **T**
(b) Peter is a computer programmer,$_1$ who pays no attention to Gavin,$_2$ who laughs at Sophie,$_3$ who mentions that she enjoys cooking to Gavin,$_4$ who claims to be more qualified than Sophie Peter or Fiona,$_5$ who are all applying to a high profile cookery magazine,$_6$ that is based on London.$_7$ (7s) **T**
(c) Fiona has studied food technology,$_1$ Gavin has trained as a chef,$_2$ Sophie has studied English literature,$_3$ Gavin laughs at Sophie,$_4$ Peter has studied computing,$_5$ Gavin suggests that Peter refuses to answer,$_6$ and they have all come for the same interview as assistant editor of a high profile cookery magazine.$_7$ (7m) **T**

8 (a) Gavin genuinely imagines$_1$ that Sophie believes$_2$ that Fiona undoubtedly suspects$_3$ that Peter certainly thinks$_4$ that Gavin believes$_5$ that Peter knows$_6$ that he is embarrassed by Gavin's arrogant feeling$_7$ that Peter thinks$_8$ he is the best qualified of the four applicants for the position of assistant editor for a high profile cookery magazine based in London. (8i) **F**
(b) Fiona is a food technology specialist,$_1$ who shows up to the interview with Peter, Sophie and Gavin,$_2$ who immediately laughs at Sophie,$_3$ who is an English literature graduate,$_4$ who mentions that she enjoys cooking to Gavin,$_5$ who inquires about the relevant qualifications of Peter,$_6$ who reluctantly answers Gavin,$_7$ who claims to be the most qualified.$_8$ (8s) **F**
(c) Peter has studied computer programming,$_1$ Gavin has trained as a chef,$_2$ Fiona has studied food technology,$_3$ Sophie has studied English literature,$_4$ Gavin engages them in conversation,$_5$ Fiona describes her qualifications,$_6$ Peter becomes embarrassed,$_7$ and they have all four come for the same interview as assistant editor at a high profile cookery magazine based in London.$_8$ (8m) **T**

9 (a) Sophie seems to think$_1$ that Fiona knows$_2$ that Peter believes$_3$ that Gavin supposes$_4$ that he is the most qualified of the applicants, and at the same time, Sophie apparently thinks$_5$ that Fiona knows$_6$ that Peter supposes$_7$ that Sophie imagines$_8$ that Peter believes$_9$ that he is by far the least qualified for the job as assistant editor. (9i) **T**
(b) Sophie is an interviewee,$_1$ who is an English literature graduate,$_2$ who is wished good luck by Gavin,$_3$ who laughs at Sophie,$_4$ who mentions that she enjoys cooking to Gavin,$_5$ who gets ignored by Peter,$_6$ who is a computer programmer,$_7$ who refuses to reply to the questions from Gavin,$_8$ who gets told by everyone to be quiet.$_9$ (9s) **F**
(c) It was clear that Fiona had studied food technology,$_1$ Sophie had studied English literature,$_2$ Peter had studied computing,$_3$ and Gavin had trained as a chef;$_4$ later, Sophie mentions that she enjoys cooking,$_5$ Gavin laughs at Sophie,$_6$ Peter admits he has no experience in editing or cooking$_7$ and refuses to reply to Gavin's questions,$_8$ so Sophie and Fiona tell Gavin to be quiet.$_9$ (9m) **T**

Appendix A.4.2 Proportion of true-false answers to questions for each socio-cognitive story

FRANK, BETTY AND BRIAN STORY

INTENTIONALITY QUESTIONS			RECURSIVE SYNTAX QUESTIONS			MEMORY QUESTIONS		
	TRUE	FALSE		TRUE	FALSE		TRUE	FALSE
1		X	1		X	1		X
2		X	2		X	2	X	
3		X	3	X		3		X
4	X		4	X		4		X
5		X	5		X	5		X
6	X		6	X		6	X	
7		X	7	X		7		X
8	X		8	X		8	X	
9	X		9		X	9		X

CHARLOTTE, SIMON, JANE AND MARTIN STORY

INTENTIONALITY QUESTIONS			RECURSIVE SYNTAX QUESTIONS			MEMORY QUESTIONS		
	TRUE	FALSE		TRUE	FALSE		TRUE	FALSE
1		X	1	X		1		X
2	X		2	X		2	X	
3		X	3		X	3		X
4	X		4	X		4	X	
5		X	5		X	5	X	
6		X	6	X		6		X
7		X	7	X		7	X	
8	X		8		X	8		X
9	X		9		X	9		X

GAVIN, PETER, FIONA AND SOPHIE STORY

INTENTIONALITY QUESTIONS			RECURSIVE SYNTAX QUESTIONS			MEMORY QUESTIONS		
	TRUE	FALSE		TRUE	FALSE		TRUE	FALSE
1	X		1		X	1	X	
2		X	2	X		2	X	
3	X		3		X	3		X
4		X	4		X	4	X	
5	X		5	X		5	X	
6		X	6		X	6		X
7	X		7	X		7	X	
8		X	8		X	8	X	
9	X		9		X	9		X

	TRUE	FALSE		TRUE	FALSE		TRUE	FALSE
TOTAL	13	14		13	14		13	14