UNIVERSITY OF CALIFORNIA,
IRVINE

Complex Signals: Reflexivity, Hierarchical Structure, and Modular Composition

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Logic & Philosophy of Science

by

Travis LaCroix

Dissertation Committee:
Chancellor's Professor Jeffrey A. Barrett, Chair
Distinguished Professor Brian Skyrms
Professor Simon Huttegger
Associate Professor Cailin O'Connor

2020

DEDICATION

To

Sarah LaCroix

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are two individuals in particular who deserve especial thanks, without whom I almost certainly would not have been able to complete this project or this programme. First, Jeff Barrett has been enormously gracious in aiding my pursuits with his time and understanding. Were it not for the continued support and independence that was afforded to me by Jeff during my PhD, who unquestioningly approved and encouraged my working abroad, it is entirely unclear whether I would have been able to continue on in the programme in the first place. It is due, in no small part, to Jeff's role as an advisor, mentor, and colleague that I have gotten to this point. Second, I would like to thank Yoshua Bengio for the trust and belief that he has placed in me and my work. The provision of generous space and funding in Montréal at Mila were unfathomably helpful to make my project feasible. So, to Jeff and Yoshua, I extend my utmost gratitude.

I must also thank the members of my dissertation committee, each of whose influence, help, and feedback during the writing of this dissertation in particular—but even more so the duration of my PhD—has been greatly appreciated. Many thanks to Jeff Barrett, Brian Skyrms, Simon Huttegger, and Cailin O'Connor, on this account.

I would like to thank Mila - l'Institut Québécois d'Intelligence Artificielle and l'Université de Montréal for generous financial support during the research and writing of this dissertation.

I would like to thank the following individuals, in no particular order, for helpful comments on earlier drafts of this dissertation, and general intellectual contributions throughout this process: Jeffrey A. Barrett, Brian Skyrms, Simon Huttegger, Josh Armstrong, Cailin O'Connor, Natalia Komorova, Nic Fillion, Adrian Currie, Shane Steinert-Threlkeld, Aydin Mohseni, Daniel Herrmann, Michael Noukhovitch, Aaron Courville, Calvin Cochran, Bert Baumgaertner, Rafael Ventura, Kino Zhao, Zoe Cocchiaro, C. Kenneth Waters, Tristan Taylor, Gabriel Larivière, and Melissa Berthet.

Finally, for more general support in the last several years, I would like to extend my thanks to Patty Jones, Lauren Ross, Jeremy Heis, and John Sommerhauser, as well as Allan LaCroix, Heather LaCroix, Walter Despot, and Barbara Despot.

And, of course, Sarah and Atlas, who have been beacons in the dark times that we call 'candidacy'. Sarah, especially, has been very patient over the years: From my first day of post-secondary studies at Camosun College in Victoria when I came home with an idea in my mind to get a PhD in philosophy, to the night I left Montréal to defend my dissertation. She has been there since before the beginning. Atlas, on the other hand, has no patience for anything; but she is a good dog.

Early versions of several of these chapters have been presented or accepted for presentation at a number of venues over the course of the previous two years. I am grateful to the organisers of these conferences, workshops, and talks for providing a platform for discussing this work,

and I am equally grateful to the respective audience members for stimulating questions and discussions. These are as follows:

Chapter 3

(a) Emergent Communication: Toward Natural Language. 3rd NeurIPS workshop on Emergent Communication. 2019 Conference on Neural Information Processing Systems. December 2019, Vancouver, Canada.

Chapter 4

(a) Interdisciplinary Workshop Series. McGill University, Department of Philosophy. October 2018, Montréal, Québec.

(b) Philosophy of Science Association, November 2018, Seattle, USA.

(c) Canadian Philosophical Association, June 2019, Vancouver, Canada.

Chapter 5

(a) Formal Epistemology Workshop, June 2019, Turin, Italy.

Chapter 6

(a) American Philosophical Association, Pacific Division. April 2019, Vancouver, British Columbia.

(b) Society for Exact Philosophy. May 2019. Toronto, Ontario.

(c) International Congress on Logic, Methodology, and Philosophy of Science and Technology. August 2019. Prague, Czechia.

Chapter Six is published in *British Journal for the Philosophy of Science*: 'Using Logic to Evolve More Logic: Composing Logical Operators via Self-Assembly' by Travis LaCroix, 2019, reproduced by permission of Oxford University Press https://doi.org/10.1093/bjps/axz049.

Finally, I would like to extend no thanks whatsoever to the undergraduate student from McGill University, who spilled her coffee on my laptop while I was working at Dispatch, St. Laurent in Montréal, that fateful October morning of 2019.

Montréal/Toronto/Irvine
01 March 2020

VITA

Travis LaCroix

Ph.D in Logic & Philosophy of Science **2020**
University of California, Irvine *Irvine, California*

M.A. in Social Science **2018**
University of California, Irvine *Irvine, California*

M.A. in Philosophy **2016**
Simon Fraser University *Burnaby, British Columbia*

B.A. in Philosophy (Hons.), English Literature **2014**
University of British Columbia *Vancouver, British Columbia*

A.A. in English **2011**
Camosun College *Victoria, British Columbia*

PUBLICATIONS

Epistemology and the Structure of Language (with J. A. Barrett) **2020**
Erkenntnis

Communicative Bottlenecks Lead to Maximal Information Transfer **2020**
Journal of Experimental and Theoretical Artificial Intelligence

Using Logic to Evolve More Logic **2019**
British Journal for the Philosophy of Science

Biology and Compositionality **2019**
Emergent Communication Workshop at NeurIPS 19

Evolutionary Explanations of Simple Communication **2019**
Journal for General Philosophy of Science

On Salience and Signalling in Sender-Receiver Games **2018**
Synthese

SELECTED PRESENTATIONS

Emerging Communication under Conflict of Interest *EvoLang XIII*	**Apr 2020**
Learning from Learning Machines *AI and Moral Learning (AIML) workshop at AISB-20*	**Apr 2020**
Epistemology and the Structure of Language *Concordia Research Symposium in Philosophy*	**Feb 2020**
Biology and Compositionality: Empirical Considerations for Emergent Communication Protocols *Conference on Neural Information Processing Systems (NeurIPS)*	**Dec 2019**
Learning from Learning Machines *Canadian Society for Epistemology*	**Nov 2019**
Using Logic to Evolve More Logic *Congress on Logic, Methodology, and Philosophy of Science and Technology*	**Aug 2019**
The Correction Game *Formal Epistemology Workshop*	**Jul 2019**
Accounting for Role-Asymmetries in the Evolution of Compositional Signals *Canadian Philosophical Association*	**Jun 2019**
Using Logic to Evolve More Logic *Society for Exact Philosophy*	**May 2019**
Using Logic to Evolve More Logic *American Philosophical Association, Pacific Division*	**Apr 2019**
Less is More: Degrees of Compositionality for Complex Signals *Philosophy of Science Association*	**Nov 2018**
Reference by Proxy and Truth-in-a-Model *Western Canadian Philosophical Association*	**Oct 2018**
On the Role of Power in the Evolution of Inequitable Norms *Canadian Philosophical Association*	**Jun 2018**
On the Role of Power in the Evolution of Inequitable Norms *Philogica V and ALFAn V*	**May 2018**
On Salience and Signalling in Sender-Receiver Games *Western Canadian Philosophical Association*	**Oct 2017**
Signalling Games and Their Models *Colombian Conference on Logic, Epistemology, and Philosophy of Science*	**Feb 2016**

ABSTRACT OF THE DISSERTATION

Complex Signals: Reflexivity, Hierarchical Structure, and Modular Composition

By

Travis LaCroix

Doctor of Philosophy in Logic & Philosophy of Science

University of California, Irvine, 2020

Chancellor's Professor Jeffrey A. Barrett, Chair

This dissertation argues that what drives the emergence of complex communication systems is a process of modular composition, whereby independent communicative dispositions combine to create more complex dispositions. This challenges the dominant view in language-origins research, which attempts to resolve the explanatory gap between (simple) communication and (natural) language by demonstrating how complex syntax evolved. I show that these accounts fail to maintain sensitivity to empirical data: genuinely compositional syntax is extremely rare or non-existent in nature. In contrast, I propose that the *reflexive* properties of natural language—the ability to use language to talk about language—provide a plausible alternative explanatory target.

Part I provides the philosophical foundation of this novel account using the theoretical framework of Lewis-Skyrms signalling games and drawing upon relevant work in evolutionary biology, linguistics, cognitive systems, and machine learning. Part II provides a concrete set of models, along with analytic and simulation results, that show precisely how (and under what circumstances) this process of modular composition is supposed to work.

Introduction

So long as we are ignorant of how a thing arose, we cannot be said to know it.

— Schleicher,
Darwinism Tested by the Science of Language

Communication is found everywhere in nature. However, language is often taken to be unique to humans. Thus, the question arises:

How did language evolve?

It has been suggested that this question represents the hardest problem in science.[1] Although this is a bold claim, there are several reasons to take this suggestion seriously. For one, it is difficult to define what precisely constitutes a language—language is always in flux, it is infinitely flexible and omnipresent, and it is (arguably) the most complex behaviour known (Christiansen and Kirby, 2003, 15). One of the main difficulties arising in the study of language origins is a lack of direct evidence: language does not fossilise, and we cannot go back in time to observe the actual precursors of human-level linguistic capacities.

Despite the many difficulties that arise in studying the evolution of language, this does not imply that concrete and plausible answers to this question are unattainable. Nonetheless,

[1] This is a strong suggestion in light of stiff competition—for example, the problem of reconciling quantum theory and gravity or solving the so-called 'hard problem' of consciousness (Chalmers, 1995).

any inquiry that relies on indirect evidence does require some degree of care: it is necessary to maintain sensitivity to an increasing pool of empirical data from a variety of disciplines, such as modern biology—including evolutionary theory, developmental and molecular genetics, and neuroscience—and the contemporary language sciences—including theoretical linguistics, psycholinguistics, and comparative linguistics (Fitch, 2010). Since language is a composite system, it is necessary to take a *multi-component* approach to the evolution of language rather than singling out a particular aspect of language and ignoring all else. Fitch (2010, 2017) additionally argues for a broadly comparative approach to the evolution of language, which uses variation and disparity to understand patterns in nature. In this particular case, a comparative approach involves studying communication in a wide variety of species to develop and test hypotheses about the evolution of specific abilities.

In addition to these empirical sensitivities, the use of computer modelling and simulation (though labelled 'controversial' in some circles) can help us to overcome problematic aspects of theories of language evolution—in particular, such theories tend to be stated in vague and general terms, which makes it challenging to generate and test specific predictions. Cangelosi and Parisi (2002) highlight that a simulation is the implementation of a theory in a computer. As such, computer simulations can be used as (1) virtual laboratories for conducting bona fide experiments, (2) tools for testing the internal validity of theories, and (3) means for studying language as a complex system. A primary benefit of computer simulations is that a computer program requires an explicit, detailed, consistent, and complete expression of one's theory to run and generate results.

In this dissertation, my analysis of the evolution of language is couched in the theory of *signalling games*. The signalling game was initially proposed by Lewis (1969) and subsequently reinterpreted in an evolutionary context by Skyrms (1996, 2010a). On this view, language is governed by rules, and these rules are conventional; as a result, meaning depends precisely on the *conventions of use* of particular signals in a given context. (I will set this

aside for now, but much more will be said about the relationship between conventional rules and meaning in later chapters.)[2] The signalling-game framework uses formal models from evolutionary game theory to explain how *meaning* emerges via repeated interactions. In its most basic form, the signalling-game framework gives a plausible explanation as to why simple communication systems appear in nature. Further, the simple signalling game has been extended to shed light on a variety of philosophically interesting phenomena that arise in communicative contexts. These include, e.g., the difference between indicatives and imperatives (Huttegger, 2007b; Zollman, 2011); signalling in social dilemmas (Godfrey-Smith and Martínez, 2013; Wagner, 2014; Martínez and Godfrey-Smith, 2016; Brusse and Bruner, 2017; LaCroix et al., 2020); network formation (Pemantle and Skyrms, 2004; Barrett et al., 2017b); deception (Zollman et al., 2012; Martínez, 2015; Skyrms and Barrett, 2018); meta-linguistic notions of truth and probability (Barrett, 2016, 2017); syntactic structure and compositionality (Franke, 2016; Steinert-Threlkeld, 2016; Barrett et al., 2018; LaCroix, 2019c); vagueness (O'Connor, 2014); and epistemic representations, such as how the structure of one's language evolves to maintain sensitivity to the structure of the world (Barrett and LaCroix, 2020).[3]

This framework serves to bridge the gap between the absence and presence of simple communication in nature. Further, computer simulations, insofar as they are philosophically well-grounded, serve as a proxy for observing the precursors of communication, to the extent that we can obtain a clearer understanding of the necessary and sufficient preconditions under which we should expect meaningful communication to arise. Mathematical and computational modelling should be taken as *complementary to*, rather than a replacement of, other methodologies. For example, in addition to comparative studies between language and animal communication—e.g., Hauser (1996); Arnold and Zuberbühler (2006a); Fitch (2010)—plausible explanations of the evolution of language may draw upon research in ani-

[2]This is in line with the views of a number of philosophers who study language in a more traditional way, including, e.g., Wittgenstein (1953); Strawson (1970, 1974); Lewis (1975); Dummett (1975, 1978, 1989, 1993b); Searle (1976, 1980); Wiggins (1997), among others.

[3]See LaCroix (2019b) for a recent overview.

mal cognition (Griffin, 1992), neuroscience (Rizzolatti et al., 1996), speech physiology (Fitch, 2000), genetics (Enard et al., 2002), experimental psychology (Kirby et al., 2008), gesturology (McNeill, 2005) or sign-language studies (Emmorey, 2002), and paleoanthropology (Wilkins and Wakefield, 1995) or archaeology (McBrearty and Brooks, 2000).

Though language is a rare phenomenon, existing in a single extant species, communication is ubiquitous in nature.[4] As such, in order to answer how language might have evolved—i.e., out of simpler communicative precursors—we must also investigate the following question:

What are the relevant differences between communication and language?

One of the crucial differences between communication and language that researchers often point to is the *productive capacity* or *openness* of natural languages: with a limited vocabulary and a finite set of grammatical rules, human languages allow for the production of an unlimited number of unique expressions. A principle of such productive features of natural language captures how arbitrary sounds can be combined in endless variations to form semantically meaningful and syntactically permissible units—e.g., phonemes form morphemes and words, and words form phrasal expressions and sentences. This is often referred to as the *Principle of Compositionality*—the meaning of a complex expression is a function of the meaning of its parts and the ways in which they are combined (Kamp and Partee, 1995). Simple communication systems that arise in nature lack this unbounded character.

Therefore, many researchers who tackle the question of how language evolved focus on how complex syntax might have evolved or emerged from simpler systems that lack syntax. In this

[4]Throughout this dissertation, I will refer to and compare two sorts of phenomena. In general, when a description is appended with a token of 'language', 'linguistic', etc., I mean the *capacities* in general that are available to all humans and unavailable to nonhuman animals (rather than, e.g., a specific language). In contrast, the same descriptions appended with tokens of, e.g., 'communicative', 'signalling', etc., will mean the dispositions that are available to nonhuman animals. For example, 'language', 'linguistic capacities', 'linguistic disposition', 'system of language', etc., all fall into the former concept, whereas 'communication', 'communicative capacities', 'signalling disposition', 'communication system', etc., all fall into the latter concept. We will have occasion to more clearly demarcate *simple* communicative capacities, such as atomic signalling, from *complex* communicative abilities, such as syntactic signalling.

case, we might say that the *target* of an evolutionary explanation—i.e., the explanandum—is complex syntax. Thus, we can reformulate the formidable question of how languages arise to the (arguably) more tractable problem of explaining how complex syntax arises.

If complex syntax is the primary distinguishing characteristic of human language—and thus, our explanatory target—then a satisfactory account must explain how syntax becomes complex over time, thus giving rise to language. Any evolutionary account that takes compositional syntax as its desideratum needs, minimally, to address how compositionality might arise from non-compositional communication. Further, if compositionality is itself an evolutionary adaptation, then such an account should explain why it might be selected for in the first place (or why it should be a byproduct of other selected-for properties). Finally, such an evolutionary account must explain why compositionality is rare in nature, though communication is ubiquitous. There are, roughly speaking, two major schools of thought about how this may have happened: the *saltationist* (or *emergent*) perspective and the *gradualist* perspective.

Saltationism—from the Latin *saltus*, meaning 'leap'—is the view that language sprang into existence suddenly and recently and that there is a *complete discontinuity* between human language capacities and the communication systems of nonhuman animals. 'Suddenly and recently' can be taken to mean approximately $50,000$ (Chomsky, 2005) to $200,000$ years ago (Berwick and Chomsky, 2016). 'Complete discontinuity' means language is present in *Homo sapiens* alone, and no other species.[5] This approach is endorsed, to some degree or another by, e.g., Berwick (1998); Bickerton (1990, 1998); Lightfoot (1991); Hauser et al. (2002); Chomsky (2002, 2005, 2010); Piattelli-Palmarini and Uriagereka (2004, 2011); Moro

[5]Though it is logically possible that language emerged suddenly in an ancestral species—such as *Homo heidelbergensis*—Progovac (2019) suggests that this possibility has not been entertained by saltationists (3). Note that the timeline in Chomsky (2005) implies that Neanderthals did not have language, but Berwick and Chomsky (2016) say that it is an open question. Even so, It is contested whether Neanderthals—a less distant ancestor to *H. sapiens* than *H. heidelbergensis*—had language.

(2008); Hornstein (2008); Piattelli-Palmarini (2010); Berwick and Chomsky (2011, 2016); Di Sciullo (2011, 2013); Bolhuis et al. (2014); Miyagawa et al. (2015); Miyagawa (2017), etc.

A saltationist account might suggest that the compositional nature of language is a mere byproduct of some other adaptive feature unique to humans, or that compositional language was the result of a 'catastrophic' change (Bickerton, 1998), whereby several precursors culminated in a 'leap' from non-language to language.

However, in this case, an explanation is still required: the saltationist who holds that compositional syntax is the key differentiating feature of human language must still explain how it (suddenly) arose, how the leap is made from non-compositional communication to compositional language, and why compositional syntax is so rare.

Some authors—e.g., Chomsky (1980b); Bickerton (1990); Wray (1998)—avoid this question to some extent by suggesting that the *primary* purpose of language is/was not communication, *per se*. For example:

> Suppose that in the quiet of my study I think about a problem, using language, and even write down what I think. Suppose that someone speaks honestly, merely out of a sense of integrity, fully aware that his audience will refuse to comprehend or even consider what he is saying. Consider informal conversation conducted for the sole purpose of maintaining casual friendly relations, with no particular concern as to its content. Are these examples of 'communication'? If so, what do we mean by 'communication' in the absence of an audience, or with an audience assumed to be completely unresponsive, or with no intention to convey information or modify belief or attitude?
>
> It seems that either we must deprive the notion 'communication' of all significance, or else we must reject the view that the purpose of language is communication. (Chomsky, 1980b, 230)[6]

[6]Note that one obvious response to this might be that the uses of language which Chomsky describes here are derivative of communicative uses. It is easy to imagine how, if an individual can communicate with others, she might be able to sit quietly in her study and write down her thoughts using the tools that she uses to communicate socially. This evolutionary description strikes me as more parsimonious than suggesting that language is a fundamentally different type of thing from communication. Nonetheless, Fitch (2010) points out that Chomsky's views are often uncharitably caricatured:

In such a case, the function of a fully-syntactic language is understood to be the expression of *thought*.

This conception of language derives from a version of Plato's problem of 'explaining how we can know so much given that we have such limited evidence' (Chomsky, 1986, xxv). This is generally known as the *poverty of the stimulus* (Chomsky, 1980a). Chomsky's solution to this problem led him to posit the notion of a *Universal Grammar*, which is a 'part of our biological endowment, genetically determined, on a par with elements of our common nature that cause us to grow arms and legs rather than wings' (Chomsky, 1988, 4). Thus, according to Chomsky, humans possess (cognitively) a distinct language faculty (called the *faculty of language*); linguistic knowledge consists in abstract and unconscious principles; and language is acquired rather than learned.[7]

Chomsky later distinguishes two conceptual components of the faculty of language. The faculty of language in the broad sense (FLB) is the group of organism-internal systems used for the production and comprehension of language, whereas the faculty of language in the narrow sense (FLN) is the subset of mechanisms in FLB which are uniquely human and thus unique to language (Hauser et al., 2002).

In general, saltationist theories are couched in terms of the *minimalist program* (Chomsky, 1995), which is supposed to be a theory-neutral reduction of the number of mechanisms needed to account for language (Chomsky, 2010). The minimalist program posits a single

> Chomsky's perspective was clear: that only some aspects of language could be understood as adaptations for communication. Other aspects—including many of the minutiae of syntax or semantics—seemed more likely to result from cognitive, historical, or developmental constraints. This perspective seemed both congenial and diametrically opposed to the 'Chomsky' caricature painted by the literature opposing him. Deep confusion was caused, it seemed, by some scholars using 'language' to denote language *in toto*, while others like Chomsky used it to denote a far more circumscribed set of mechanisms: essentially the computational mechanisms central to syntax, which allow an unbounded set of structures to be formed by a finite set of rules operating on a finite vocabulary. It became clear ... that much of the heated debate in the field was based on terminological misunderstandings, intertwined with a fundamental issue of ongoing biological debate (constraints versus adaptation). (21)

See also Hurford's response to this passage from Chomsky (Hurford, 2007, 174-7).

[7] See the discussion in Lin (1999).

syntactic operation called *Merge*. This is a recursive binary operation which derives hierarchical binary branching structures and is supposed to account for all syntactic manipulations (Chomsky, 1999, 2010; Bickerton, 2009; Collins and Stabler, 2016; Berwick and Chomsky, 2016).

An alternative to the saltationist view of language emergence is a *gradualist* view. *Gradualism*—from the Latin *gradus*, meaning 'step'—is the view that language (i.e., complex syntax) evolved slowly over long periods. Some argue that a form of language was available to Neanderthals; Dediu and Levinson (2013) propose that language might date as far back as *H. heidelbergensis*, up to 500,000 years ago—however, this is controversial, and there is not yet a consensus as to whether Neanderthals had language. A gradualist perspective does not necessarily require that such an evolutionary process was continuous, nor constant rate: a Darwinian notion of adaptation involves incremental processes moving forward in steps as opposed to leaps, but this is consistent with varying rates of change over time and small jump-discontinuities in evolutionary progress.[8]

The gradualist view—though not necessarily natural selection—is endorsed by, e.g., Givón (1979, 2002a,b, 2009); Pinker and Bloom (1990); Bickerton (1990, 1998); Newmeyer (1991, 1998, 2005); Jackendoff (1999, 2002); Fitch (2008, 2010); Culicover and Jackendoff (2005); Progovac (2006, 2009a,b, 2013, 2015, 2019); Tallerman (2007, 2013a,b, 2014a,b); Heine and Kuteva (2007); Hurford (2007, 2012), among many others.[9] In this context, the explanation of how linguistic capacities might arise is a *diachronic* story about how languages get to be complex over (potentially significant expanses of) historical time by a combination of genetic and cultural evolution. Hurford (2012) points out that, in a long-term evolutionary account, the full story must be one of *complexification*: complex languages evolved from simpler

[8]See the discussion in Dawkins (1996) and Fitch (2010).

[9]Note that Derek Bickerton appears on both the saltationist and gradualist lists. This is not a mistake: his view posits a protolanguage, which is generally denied by the saltationist, but that the emergence of syntax between protolanguage and language was 'catastrophic', which is usually rejected by the gradualist.

communication systems (372). How does this 'complexification' occur? What allows, or causes, complex language to evolve from simpler systems of communication?

Some gradualists hold that there is no selective pressure for complex syntax; instead, it developed primarily via *cultural* evolution.[10] For example, Tomasello (1999) argues that the syntactic characteristics of human language arise from the *purely* cultural process of *grammaticalisation*—i.e., the process by which words that represent concreta, like objects and actions, become grammatical markers, thus creating new function words. Grammaticalisation is a cultural evolutionary process to the extent that it is contingent upon the constraints of communicative interactions (Hopper and Traugott, 2003). Others suggest that complex syntax is a *spandrel* that develops out of some additional evolutionary product which itself is subject to selection.[11] For example, syntax might be a byproduct of speech-motor control, which in turn develops from changes in basal ganglia due to speech evolution (Lieberman, 2000); these *physical* changes are subject to selective pressure, though syntax itself is not.

Other gradualists suggest that syntax genuinely *is* dependent upon selective pressure, and complex syntax evolved due to a specific set of rich and complex adaptations, as a result of natural selection, for increased communicative efficacy or efficiency (Jackendoff, 1999, 2002; Pinker and Bloom, 1990; Pinker and Jackendoff, 2005), or that the move from simple to complex syntax was determined by sexual selection (Progovac, 2015). Similarly, Bickerton (1998, 2000); Seyfarth and Cheney (2005) argue that social knowledge is a necessary factor in the origins of complex syntax; however, they also suggest that this 'pre-adaptation' (or possibly *exaptation*) was indeed subject to independent selection pressures, thus positing a

[10] Note that this does not avoid the second requirement above, that an evolutionary account should explain why compositionality is selected for, per se. Rather, this shifts the explanation as to why compositionality is should arise, e.g., by giving a cultural story or explaining why compositionality is a byproduct of some other feature that was selected for.

[11] A *spandrel* in evolutionary biology is a characteristic that is a byproduct of the evolution of some other feature, rather than a direct product of adaptive selection. This terminology was introduced by Gould and Lewontin (1979) and was meant to be analogous to the architectural spandrel—a triangular gap at the corner of an arch—which is a 'secondary epiphenomenon . . . not a cause of the entire system' (584).

real adaptive advantage for syntax. Givón (1995), for example, suggests that syntax arises initially as a byproduct of visual processing; but, once this has happened, syntax itself becomes subject to further selective pressures.

The main difference between the saltationist and the gradualist views might be summarised thus: the saltationists hold that it is inconceivable that there could have existed a *language* that is not complete with all the complexities of modern syntax, whereas the gradualists allow for the possibility of an intermediate stage (or stages) between simple communication and language—i.e., *protolanguage*. In both cases, whether in steps or leaps, what is supposed to have evolved (or emerged) is *complex syntax*. Since this is unique to language, complex syntax is the explanatory target of both camps.

Even though arguments concerning language origins can be partitioned into two exhaustive and mutually exclusive camps, very few researchers actually argue *for* a gradualist stance over a saltationist stance, or vice-versa.[12] Part of the reason for this, as is aptly pointed out by Jackendoff (2010), is that one's theory of language evolution depends upon one's theory of language. For example, if one thinks that the primary purpose of language is to communicate, then one would hold that language is continuous with communication, and therefore, will likely argue a gradualist position to language origins.

Given that the signalling-game framework offers, by design, an *evolutionary* account of the emergence of meaning, my view in this dissertation can be taken to be firmly in line with the gradualist perspective; I will not argue for the virtues of gradualism over saltationism here, but I will assume that this is the correct view.[13] In light of the gradualist requirement

[12]For two exceptions, see Muszynski (2015); LaCroix (2020b).

[13]See LaCroix (2020b). Progovac (2019) offers a critical discussion of both these positions, but she highlights that

> the purpose of the scientific method is to narrow down the range of possibilities by attempting to exclude (falsify) as many of the available hypotheses as possible[.] ... The methods for testing evolutionary hypotheses include computer simulations; fMRI experiments; and statistical correlations between linguistic variation and genetic variation[.] ... the best theories from a variety of diverse fields will need to come together, the fields as disparate as: evolutionary biology; typological, theoretical, and historical linguistics; anthropology; genetics; neuroscience.

of relative continuity between simple communication systems and language—i.e., that the latter evolved *out of* the former—I will also assume that the *primary* purpose of language is (or at least initially was) to communicate.

This bias toward syntax arises from Chomsky (1957, 1965), who showed that language requires a generative system that makes an unlimited variety of sentences possible; however, he assumed without explicit argument that the generativity of language arises entirely because of the syntactic component of grammar. On this account, the combinatorial properties of phonology and semantics are strictly derivative of the combinatorial properties of syntax. Jackendoff (2007) refers to this tendency as *syntactocentrism* and argues that it was a 'scientific mistake'.

This syntactocentric tendency has also pervaded evolutionary models in the signalling-game framework. In recent years, several formal models of signalling conventions have been proposed to explain how and under what circumstances *compositional* signalling might arise. To that end, researchers who employ formal models from evolutionary game theory to test hypotheses of how simple communicative contexts might evolve have striven to explain how compositionality might itself evolve out of these more straightforward contexts.

I believe that the emphasis on syntax in language origins is mistaken. That is to say, it is a mistake to assume that *since* complex syntax provides a crucial difference between language and simple communication, research on language origins must centre on the evolution of complex syntax itself. Thus, the purpose of this dissertation is to explain why the focus on compositional syntax is wrong—primarily in light of a lack of empirical evidence. Further,

---

> As daunting as the task may seem, ... evolution (via selection) is a force which can bring all these fields together, and most probably the only force that can accomplish it. (83)

She suggests that, though saltationism has held some sway for a significant amount of time, at least in the sense that language is a relatively recent development, 'today the pendulum seems to be swinging in the direction of the belief that Neanderthals commanded some kind of language' (3), which would make this capacity much older than saltationists are often willing to allow.

in lieu of complex syntax, I offer a new, alternative account of how complex communication might naturally arise from simpler precursors via a process of *reflexivity*.

Using the formal framework of the signalling game (Chapter 4), I demonstrate that evolutionary accounts of compositional signalling fail to explain how complex syntax evolved insofar as they do not distinguish between a full-blooded notion of linguistic compositionality and the simple sort of communicative (proto-)compositionality which they purport to model. This conceptual obscurity gives rise to several problems, as we shall see. Further, such explanations fail to account for the inherent role-asymmetries of the sender and receiver in the signalling game and, in doing so, fail to explain why compositional signalling is beneficial to *both* agents.

In the context of gradualist perspectives concerning the evolution of compositional syntax more generally (Chapter 2), I argue that such accounts fail to maintain sensitivity to empirical data regarding evolutionary precursors. That is, genuinely compositional syntax is rare or non-existent in nature. On the other hand, I argue that a system of communication is either compositional or it is not. As such, there is no room for so-called *proto-compositionality*. Therefore, theories that purport to explain the evolution of complex syntax run afoul of the gradualist assumptions in which these explanations are couched.

Though compositional syntax is indeed unique to language, it is not the only such property. Rather than compositionality, I focus on *reflexivity* as a crucial feature of language. Reflexivity refers to the fact that language can be used to talk about language.

I demonstrate that communication is a unique evolved system in the following sense. Once a group of individuals has learned some simple communication convention, those learned conventions may be used to influence future communication behaviour, thus giving rise to a feedback loop. When faced with a novel context, an individual can learn a brand-new disposition from scratch; however, the individual may also learn to take advantage of previ-

ously evolved dispositions. Indeed, individuals may learn to take advantage of pre-evolved *communicative* dispositions to thereby influence the evolution of future communication; this is the conception of reflexivity, as an evolutionary mechanism, which I examine.

I argue (Chapter 3) that what drives the emergence of complex communication systems is a process of *modular composition*, whereby independently evolved communicative dispositions combine to create more complex dispositions, and that a notion of modularity is necessary for an evolutionary account of linguistic capacities. I further argue that this process of modular composition depends inherently upon reflexivity. Once some complexity is exhibited, at a small scale, it may lead to a 'feedback loop' between communication and cognition that gives rise to the complexity we see in natural language. This serves to connect parallel research in the evolution of language and cognitive systems.

Finally, I argue that the reflexivity of language, as opposed to compositionality, is the correct explanatory target of an evolutionary account of language to the extent that it has salient precursors in simple communication systems, so it can account for empirical data; it offers a genuinely gradualist perspective; and it is able to give rise to hierarchical compositional structures. Thus, complex syntax is a *byproduct*, rather than a target, on my view.

Outline of the Dissertation

This dissertation is divided into two parts. The first part (Chapters 1-3) provides the philosophical foundations for my account of the evolution of language via modular composition using the theoretical framework of the signalling game. This further draws upon relevant work in evolutionary biology, linguistics, cognitive systems, and machine learning. This background is meant to situate my contributions within this literature while simultaneously justifying my novel contributions in light of empirical research done in several different fields. The second part of this dissertation (Chapters 4-6) provides a novel set of models, along with

analytic and simulation results, that show precisely how and under what circumstances this process of modular composition is supposed to work.

I present a summary of the main content of each chapter below.

Chapter 1. In this first chapter, I introduce the signalling game and relate it to questions in traditional philosophy of language. This chapter will serve as a technical and philosophical introduction to the methods used in the signalling-game literature and, as such, will be referred back to throughout the dissertation.

I begin by examining the insight from Lewis (1969) that language *qua* communication can be understood in game-theoretic terms and that it requires coordination. Thus, the first step in analysing language involves understanding how individuals coordinate in the first place. I highlight that this insight laid the foundation for what would become the theory of signalling games, where signalling conventions are evolved or learned in a population over time—namely, the insight of Skyrms (1996) to analyse social dynamical phenomena using the tools of evolutionary, as opposed to classical, game theory. I then survey how the simple evolutionary signalling model has been extended and generalised to try to explain a variety of more complex communicative phenomena, or to make more empirically realistic modelling assumptions.

I further draw connections between the signalling-game framework and traditional philosophy of language, to highlight how the target concepts of the latter can be well expressed in terms of the former. Finally, I address some criticisms that have been raised against the signalling-game framework specifically, and the use of computer simulations more generally.

Chapter 2. In this second chapter, I address the salient differences between communication and language. I further present empirical data from biology and linguistics and argue that

complex syntax is not the most apt explanatory target for how language might have evolved out of simple communication.

Thus, the main question I address in this chapter concerns the distinction between animal communication and language. Though it is perhaps impossible (or otherwise unfruitful) to try to give a clear distinction between these two—i.e., in terms of a set of concrete necessary and sufficient conditions—it will be helpful nonetheless for highlighting some of the possible candidates for distinguishing characteristics.

While many analyses focus on what capacities are given by human language—i.e., abilities that are not present in animal communication—we can also ask what is common between these two phenomena, to get a clearer picture of what possible precursors to language might exist in nature.

I highlight several recent models that use signalling games to try to explain the emergence of compositionality. However, I argue that these models are not sensitive to empirical data from animal communication. This leads to the negative conclusion that compositional syntax is the wrong target for an evolutionary explanation.

Chapter 3. In this chapter, I argue that the reflexivity of language is a more fruitful property to consider with respect to uniqueness of human language.

First, I introduce the notion of *modular composition*, whereby a distinct game takes the output of a previous game as its input. This arises from the discussion of *self-assembly* in Barrett and Skyrms (2017). However, I make explicit several things that were left implicit in their work.

Such compositional processes, I show, can arise in a variety of forms—each increasing in complexity. These include transfer (of) learning (or template transfer), analogical reasoning, and modular composition more generally.

I show how each of these has some simpler precursors that are evident in nature, and that they are related to increasing complexity. I further show how this process of modular composition connects communication to cognition and social structure. Finally, I suggest that what underlies this process of modular composition is reflexivity.

Chapter 4. Though I suggested arguments against compositionality and for reflexivity as the main explanatory target of an evolutionary account of language, in this chapter I make these arguments more precise by presenting compositionality with respect to a model of information transfer.

On the one hand, I argue that evolutionary accounts of compositionality are conceptually vague and under-determined with respect to what they mean by 'compositionality'. I highlight that pre-theoretic biases concerning the meaning of compositionality seep into these accounts. On the other hand, such accounts ignore an inherent role-asymmetry between the sender and receiver in the signalling game; therefore, they fail to explain how compositional syntax can be beneficial to both the sender and the receiver in the signalling game.

I offer a formal definition of compositional signals that does not fall prey to these criticisms, and then I show precisely why, under this formal framework, extant models that purport to show how compositional syntax might evolve fail to be compositional.

Further, I show that those models which do give rise to a genuinely compositional signalling system do so because of the emergent reflexivity of the system. Thus, compositional syntax is a mere byproduct of evolutionary processes that are driven primarily by modular composition and reflexivity.

Chapter 5. This chapter presents a novel model of signalling which shows explicitly how and under what circumstances pre-evolved communicative dispositions might affect how individuals learn to communicate in a novel context. I present a set of models that vary the reward for coordination in the signalling-game framework under simple reinforcement learning as a function of the agents' actions. These take advantage of a type of modular compositional communicative bootstrapping by which the sender and receiver use pre-evolved communicative dispositions—a 'yes/no' command—to evolve new dispositions. The dynamic examined is simple reinforcement learning in a variety of contexts, including a cue-reading game, a signalling game, and a signalling game with invention. I discuss the effects of pre-evolved capacities, in addition to the possibility for co-evolved capacities. Finally, I ground the empirical adequacy of this model with respect to theories in theoretical linguistics which posit 'fossils' of language.

Chapter 6. This chapter presents a set of models which show how simple *unary* logical operations can be composed to form more complex *binary* operations via modular composition under simple reinforcement learning. These models extend the work of Barrett and Skyrms (2017), and I show how modular composition in this sense can account for phenomena which cannot be accommodated by simple template transfer.

I consider how complex logical operations might self-assemble in a signalling-game context via composition of simpler underlying dispositions. On the one hand, agents may take advantage of pre-evolved dispositions; on the other hand, they may co-evolve dispositions as they simultaneously learn to combine them to display more complex behaviour. In either case, the evolution of complex logical operations can be more efficient than evolving such capacities from scratch. Showing how complex phenomena like these might evolve provides an additional path to the possibility of evolving more or less rich notions of compositionality. However, as I highlight, it is because of the *reflexive* abilities of these systems that such

capacities arise. This helps provide another facet of the evolutionary story of how sufficiently rich, human-level cognitive or linguistic capacities may arise from simpler precursors.

Summary. In order to explain how language evolved out of simpler precursors—e.g., communication systems—it is necessary to understand the salient differences between language and simple communication. Two properties of language are taken as viable candidates: *compositionality* and *reflexivity*. Most researchers working on this problem take compositionality as their primary target, with the idea being that demonstrating how compositionality may have evolved could be sufficient for explaining how language evolved. I believe this focus is mistaken. On the one hand, there is no empirical evidence for any precursor to compositional syntax in nature. On the other hand, there is no gradualist explanation of compositionality, insofar as its presence or non-presence in a communication system is antipodal.

In opposition to this paradigm, I suggest that *reflexivity* should be our primary target of inquiry. Reflexivity allows for the evolution of complexity via self-reference. This can be modelled using the signalling game, to the extent that the *functional referent* of a signal can be the elements of the game, and eventually signals themselves. On my view, compositionality is an emergent property of complex signalling, which arises once hierarchical communication structures are evolved via modular composition, which (in turn) is fundamentally supported by reflexivity.

This view is both parsimonious and plausible. There are empirical precursors to this sort of phenomenon that are observable in nature, and this process has demonstrable effects on efficiency with respect to signalling.

My position is couched in terms of a gradualist approach to the evolution of language, and thus it is firmly against the saltationist stance that language emerged suddenly. I believe this is the correct view; for example, Givón (2002b) argues that 'like other biological phenomena,

language cannot be fully understood without reference to its evolution, whether proven or hypothesized' (39). However, I assume this position rather than arguing for it. In this context, I might simply defer to Dobzhansky (1973):

Nothing in biology makes sense except in the light of evolution.

And, with that, we begin.

Part I

Philosophical Background

To imagine a language means to imagine a form of life.

– Wittgenstein, *Philosophical Investigations*

Chapter 1

Communication and Conventions

> *Sie selbst (die Sprache) ist kein Werk (Ergon), sondern eine Thätigkeit (Energeia).*
>
> — Humboldt, *Über die Verschiedenheit des Menschlichen Sprachbaues*

This chapter characterises the signalling game and explicitly relates it to questions in traditional philosophy of language. I additionally address some criticisms that have been raised against the use of computer simulations in general and the signalling-game framework in particular.

In Section 1.1, I introduce the signalling-game framework. This will serve as the technical background for subsequent chapters. I begin (Section 1.1.1) by highlighting Lewis' (1969) insight that conventional meaning is a problem of coordination which requires joint action. Since communication is analysed in terms of conventions, and conventions are analysed in terms of coordination problems, we can use game-theoretic tools, including equilibrium concepts, to analyse communication. As we shall see, Lewis (1969) suggests several ways of solving such coordination problems—including prior communication, salience, and precedent.

However, as a philosophical response to the sceptical position that conventional communication requires antecedent communicative conventions, these solutions will not suffice.

Though salience is perhaps sufficient to get a communication system off the ground, it is not necessary. In particular, we can couch the signalling game in the framework of evolutionary—as opposed to classical—game theory (Section 1.1.2). This is the insight of Skyrms (1996, 2010a). The problem of coordination then becomes a problem of breaking symmetry. I present a number of evolutionary and learning dynamics (Sections 1.1.3 and 1.1.4, respectively), in addition to known results that show how and under what circumstances simple signalling might arise. Finally, I show several ways in which the simple signalling game can be extended and modified to account for a variety of interesting phenomena (Section 1.1.5).

In Section 1.1.6, I mention some insights that the signalling-game framework provides for studying language and the emergence of language, highlighting some benefits over analyses in the philosophy of language more traditionally construed. I conclude (Section 1.2) by addressing some theoretical concerns that have been raised against the signalling-game framework specifically, and the use of computer simulations more generally. Ultimately, I argue that these concerns are not well-founded.

1.1 Signalling Games

1.1.1 Communication Conventions

The signalling game, due to Lewis (1969), arises as an explanation of language as a convention. Lewis' central insight is that successful communication constitutes a coordination problem—an individual ought to (or wants to) use a particular signal to mean one thing if most everyone else uses (understands) that signal in the same way. Another way of putting

this insight is that for a signal to have its meaning—or at least to have so successfully in a communication context—it must do so by convention. That is, a sufficient number of individuals in the community must also use that signal in the same way. Thus, the question of conventional meaning becomes one of coordination. Realising that language is conventional means that to understand what comprises successful communication will require a thorough analysis of conventions; further, to understand conventions, one must analyse coordination problems.

Lewis (1969) outlines the generality of the coordination-problem framework by pointing out several distinct and varied situations that can be understood as coordination problems. This includes meeting someone at some time and place, phoning back or waiting after a call is disconnected, rowing a boat (an example taken from Hume (1739)), driving on one or the other side of the road, dressing appropriately for an event, setting prices in an oligopoly, hunting stag, dividing common resources, trading commodities, and, finally, communicating. Cooperation in this sense requires a notion of joint action—the idea that individuals have shared intentions and, perhaps, awareness of their roles.[1]

Once we couch communication as success or failure in coordinating (e.g., meanings to state-act pairs), it is necessary to provide some analysis of coordination problems themselves to determine how they might be solved. These types of problems can be analysed from a game-theoretic perspective, giving rise to equilibria concepts—namely, situations wherein each actor has done the best that she can, given what others are doing. The foremost equilibrium concept in classical game theory is the *Nash equilibrium* (Neumann and Morgenstern, 1944; Nash, 1951).[2] This consists of a strategy for each player whereby no player can unilaterally deviate from her strategy to guarantee an increase in her payoff. A *strict* Nash equilibrium is one in which a player may be strictly worse-off after unilateral deviation from her strategy.

[1]See Gilbert (1989); Cohen and Levesque (1991); Searle (1995); Clark (1996); Bratman (1999).

[2]See also Bellhouse and Fillion (2015); Fillion (2015) for a discussion of the historical origin of the equilibrium concepts prior to Nash.

When in equilibrium, no single player has an incentive to deviate from her strategy. This is true even if *multilateral* deviation—several players deviating from their own strategy at the same time—would result in a better outcome for the group as a whole.

For example, consider the normal form for a generic 2×2 symmetric game, given in Table 1.1.

		Player 2	
		A	B
Player 1	A	(a, a)	(b, c)
	B	(c, b)	(d, d)

Table 1.1: Generic normal form for 2×2 symmetric game

When $a > d$ and $b \leq d$—as in a prisoner's dilemma or a stag hunt, for example—if the players play the equilibrium constituted by $\langle B, B \rangle$, then no player has individual incentive to deviate, even though both players deviating—i.e., to $\langle A, A \rangle$—would result in a higher payoff for each player. Therefore, any particular equilibrium need not be globally optimal; rather, it depends solely upon the actions of others—indeed, in some cases, an equilibrium may be globally pessimal.[3]

Thus, Lewis (1969) offers a formal account of what a convention is in the first place, and then he uses the tools of classical game theory to explain communication in terms of conventional signalling games. This was intended as an answer to, e.g., Russell (1922), Alston (1964), and Quine (1967): namely, the sceptical position that theoretical deference to convention as an explanation for communication results in either circularity or infinite regress. Quine (1967) argues this point against the logical positivists, who held that the validity of logical arguments and the truth of logically true propositions were valid and true, respectively, by

[3]I mean 'global' in the sense of taking account of the expectation of both players. In the example in Table 1.1, suppose $a = 5$, $b = 0$, $c = 10$, and $d = 2$. Then $\langle B, B \rangle$ is a strict Nash equilibrium and is a strictly dominant strategy. However, the average payoff for the 'group' is $\mathbb{E}(\langle A, A \rangle) = 5$, $\mathbb{E}(\langle A, B \rangle) = 5$, $\mathbb{E}(\langle B, A \rangle) = 5$, and $\mathbb{E}(\langle B, B \rangle) = 2$. In this case, the payoff for the group is worst at the strictly dominant strict Nash equilibrium. This comes to bear on repeated games where the players may switch strategies between rounds: since the players roles are symmetric, in this case, the expectation of the payoff of the group is equivalent to the expectation of, e.g., a mixed strategy for a single individual alternating between A and B with equal probability.

virtue of the meanings of the terms involved. Quine contended that this account required the pre-existence of logic; thus, 'it is not clear wherein an adoption of the conventions, antecedent to their formulation consists' (123).

Concerning the conventionality of language, Russell (1922) puts the point thus:

> A new word can be added to an existing language by a mere convention, as is done, for instance, with new scientific terms. But the basis of a language is not conventional, either from the point of view of the individual or from that of the community. ... We can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf. The association of words with their meanings must have grown up by some natural process, though at present the nature of the process is unknown. (189-190)

However, Lewis (1969) shows how we can make sense of the conventionality of language by concretely *defining* conventions—namely, in terms of the equilibria of a coordination problem. Given that communication, on Lewis' view, can be understood as being constituted by a coordination problem, we can thus explain the conventionality of language.[4]

Lewis (1969) builds off the work of Schelling (1960), who notes that games involving interdependent decisions should be understood to range on a spectrum—with games of *pure coordination* on the one end and games of *pure-conflict* (zero-sum games) on the other. The difference between coordination and conflict is characterised by payoff type. At the (pure) coordination end of Schelling's spectrum, payoffs are identical for all players, and the maximum payoff is achieved only if the players can coordinate their actions.[5]

Coordination problems are thus defined in some way by common interest amongst the individuals faced with a situation that involves coordination or cooperation.[6] If the players do

[4]Lewis himself did not show how such equilibria might come to be instantiated; however, we will see how this can happen as we go on.

[5]Note that the converse of this does not hold, since, as we have seen, players may coordinate but still fail to obtain the maximum payoff.

[6]Note, that 'cooperation' understood in this way is not merely two individuals acting such that, by chance, their action happen to benefit both of them—this is the sense of cooperation that arises in, e.g.,

not coordinate, then they both receive some reduced payoff (possibly nothing). A standard generic normal form for this sort of coordination game is given in Table 1.2.

		Player 2	
		A	B
Player 1	A	(a, a)	(b, b)
	B	(c, c)	(d, d)

Table 1.2: Generic normal form for 2×2 pure coordination game

At the opposite end of Schelling's spectrum, in games of pure conflict, payoffs are zero-sum. That is, a game is of the pure-conflict type if whatever the one player wins, the other loses. The generic normal form of a game of pure conflict is given in Table 1.3.

		Player 2	
		A	B
Player 1	A	$(a, -a)$	$(b, -b)$
	B	$(c, -c)$	$(d, -d)$

Table 1.3: Generic normal form for 2×2 pure conflict game

Thus, in cooperation games, the players depend upon one another to (individually) achieve maximum payoff, whereas, in games of pure conflict, the players work against one another to (individually) achieve maximum payoff. Lewis (1969) notes that the particular games with which he is concerned lay arbitrarily near to the coordination end of the spectrum (14).

One way to solve such coordination problems depends inherently upon a notion that Lewis (1969) calls 'suitably concordant mutual expectation' (25). This, in part, requires high-order expectations for reasoning about what others will do, and further higher-order expectations for reasoning about others' expectations, etc. Lewis points out that when there are m agents involved in a coordination problem, any given individual may have as many as $(m-1)^n$ different nth-order expectations about anything (32). So, in his words, we are 'windowless

a prisoner's dilemma, where the individuals do not know what the other is going to do. We are interested instead in the sense of cooperation that has some stability over time—cooperation in a prisoner's dilemma does not have this property to the extent that mutant defectors—i.e., individuals who adopts the novel strategy to defect—can invade a population of cooperators.

monads doing our best to mirror each other, mirror each other mirroring each other, and so on' (32). Although an analysis of this sort of recursive structure gets unwieldy very quickly, Lewis suggests that, in practice, we do not often go any further than 4th-order expectations.[7] Note that these expectations are not *necessary* for the justification of an agent's decision; however, they *strengthen* such justifications by giving independent reasons for action.

Perhaps most obviously, individuals can solve a coordination problem by prior agreement. Agreement is a way of producing expectations about actions, in addition to the higher-order expectations thus posited: if you and I agree to do x, then I expect you to do x, and I expect you to expect me to do x, and so on. Prior agreement is a particularly strong method for coordinating precisely because it produces strong concordant mutual expectations.[8] However, for the particular coordination problem involving communicative coordination, this is not going to help quell sceptical worries of how communication might get a start without already having a language in place with which to make such a prior agreement.

Still, prior agreement is not the only way of achieving such expectations about others' actions. For example, if one or another coordination equilibrium is 'preeminently conspicuous' (Lewis, 1969, 38), then we would say it is a *salient* equilibrium, in the sense of a *Schelling focal point*.[9] Schelling (1960) showed that individuals who need to cooperate could often concert their behaviour without communicating by exploiting certain salient features of the world. Note that this still requires higher-order expectations, in the sense that the player who notices a salient choice must also expect the other player to notice, etc.[10] Again, such conspicuous

[7]In the context of games, the idea that players employ finite, rather than infinite, depths of reasoning has been well studied (Binmore, 1987, 1988; Selten, 1991; Aumann, 1992; Bacharach, 1992; Bicchieri, 1993; Stahl, 1993, 1996). Stahl and Wilson (1994) examine human behaviour in symmetric 3×3 games, and they conclude that subjects use depths of reasoning of orders 1 or 2. For a different game, Nagel (1995) finds that at least 0.80 of individuals use depths of reasoning less than 4, with the modal frequency at depth-2.

[8]Note that if the individuals already have an agreed-upon language, then this constitutes 'cheap talk', which can solve coordination problems (Farrell and Rabin, 1996).

[9]See, for example, the discussions in Mehta et al. (1984a,b); Holm (2000); Crawford et al. (2008); Isoni et al. (2013); Parravano and Poulsen (2015).

[10]Schelling (1960) gives several examples of salient choice points in a coordination problem. For example, when asked to choose a time and a place to meet in New York City, an absolute majority of respondents chose *Grand Central Station* for the location, and almost all respondents chose *noon* for the time. (Though,

equilibria need not be *good*, per se—coordinating on a (globally) suboptimal equilibrium is better than not coordinating at all, in this arrangement.

Salience might occur naturally. For example, a lone rock formation might be a naturally salient meeting place for individuals. In the context of natural salience for communication, *iconic* (as opposed to arbitrary) signals might be naturally salient—e.g., imitating buzzing to symbolise bees. This sort of salience was hypothesised by Darwin (1871) to give rise to simple communication:

> As monkeys certainly understand much that is said to them by man, and as in a state of nature they utter signal-cries of danger to their fellows, it does not appear altogether incredible, that some unusually wise ape-like animal should have thought of imitating the growl of a beast of prey, so as to indicate to his fellow monkeys the nature of the expected danger. And this would have been a first step in the formation of a language. (57)

In this way, for example, a dog baring its teeth to signal being on the verge of biting is naturally salient insofar as a dog bares its teeth to bite.[11] (Also, baring one's teeth to signal biting is less costly than actually biting, as it may avoid physical confrontation altogether.)

However, salience need not be naturally present. It might also arise from familiarity with, or by analogy to, a particular coordination problem. This gives rise to the notion of *precedent*. When the same *type* of coordination problem is repeated, the repetition might make some equilibria salient—namely, those equilibria that worked for coordination when previously faced with such a coordination problem. That is, one or another coordination equilibrium can become salient due to precedent precisely because it was previously chosen. But, since

see the discussion in Farrell and Rabin (1996) about how focal points can lead to sub-optimal equilibria.) Schelling concludes that '[p]eople can often concert their intentions or expectations with others if each knows the other is trying to do the same. Most situations ... provide some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do' (57). These informal observations have been replicated in a controlled environment Mehta et al. (1984a,b).

[11] Skyrms (2010a, 21) notes that humans bare their teeth to smile, so there is some sense in which such a signal is still arbitrary. However, this depends on one's frame of reference. If the population in question consists of just dogs, then the signal can be taken as iconic.

not every instantiation of a coordination problem will be identical to those before it, salience in this sense requires at least some kind of *analogical reasoning*. Namely, salience due to precedent is salient as a preeminently conspicuous analogy to what was done before. This is a version of Hume's (1739, 1748) *problem of induction*: individuals extrapolate information from past experience and project these experiences into the future to form expectations about what others might do in situations identical or similar to those that have arisen previously.

Note that this process of salience via precedent need not be computationally demanding. It could be instantiated by something as sophisticated as Bayesian updating, or as simple as doing whatever worked last. Furthermore, once precedents are set, repetition itself gives rise to a 'metastable, self-perpetuating system of preferences, expectations, and actions'—the key ingredients of the solution to a coordination problem—which can persist indefinitely (Lewis, 1969, 42).

While agreement, (natural) salience, and precedent are all possible means of solving a coordination problem (to the extent that they produce some concordant expectations in the actions), *conventions* require the additional notion of *common knowledge*, on Lewis' account. Common knowledge makes it possible to generate higher-order expectations from existing higher-order expectations—e.g., of rationality. All this gives rise to the (quasi-)formal notion of a convention, in Lewis' sense:

> **Definition 1.1.** *Lewisian Conventions*
>
> A regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P,
>
> (1) almost everyone conforms to R;
>
> (2) almost everyone expects almost everyone else to conform to R;

(3) almost everyone has approximately the same preferences regarding all possible combinations of actions;

(4) almost everyone prefers that any one more conform to R, on condition that almost everyone conform to R;

(5) almost everyone would prefer that any one more conform to R', on condition that almost everyone conforms to R',

where R' is some possible regularity in the behaviour of members of P in S, such that almost no one in almost any instance of S among members of P could conform both to R' and to R.

Note, in particular, that there is a sense in which conventions *must* be arbitrary on this definition.[12] A non-arbitrary unique coordination equilibrium need not be conventional—one could conform to it merely because it is the best thing to do, which would violate the social structure of the convention that is built into conditions (4) and (5). Even in the case of explicit agreement, the causal influence of the initial agreement might fade over time; however, the regularities that the agreement caused will be upheld. Since such an action bears no trace to the original agreement, it can be called conventional in Lewis' sense.[13]

Processes of self-organisation thus constitute Lewisian conventions: no single individual ordains the final outcome of the convention, but repeated interactions may give rise to social

[12] A nice example of the arbitrariness of conventional signals in natural language: the modern Greek words for 'yes' and 'no' are ναι and όχι, which are pronounced [ne̝] and ['o̝çi], respectively.

[13] In his analysis, Lewis (1969) compares and contrasts conventions, thus defined, to social contracts, norms, rules, conformative behaviour, and imitation—each of which is taken to be (at least) a slightly distinct concept. He claims, for example, that conventions are a *species* of norms. This is in spite of the fact that the definition of convention does not include normative language. Though norms are usually understood as prescriptive or evaluative, Millikan (2005) points out that there are other kinds of norms. In particular, she argues that the essential norms that apply to language are non-evaluative, in the same way that norms of function and behaviour, which account for the continued proliferation of a biological species, are non-evaluative. Instead, '[s]pecific linguistic forms survive and are reproduced together with cooperative hearer responses because often enough these patterns of production and response benefit both speakers and hearers' (vi).

patterns which are thus reinforced. As time goes on, straying from the convention becomes more difficult, more detrimental, or generally less appealing to any particular individual in the population. This model of conventions applies to the conventional use of arbitrary signs to encode information for communication—the main focus of this dissertation. However, it applies equally well to numerous other phenomena that we see in culture and society, and which we may describe as conventional.

Young (1998) points out that markets historically come into existence at convenient meeting places, such as crossroads or shady areas. This is a type of *salience* for an individual determining where to start a market. Over time, it happens that customers will become accustomed to, and so learn to expect, certain types of goods to be offered at specific types of establishments, and in turn, sellers will come to meet those expectations. This is equally true for other aspects of the market—for example, particular days and hours of operation, specific procedures governing trade, and so on. These various features are often determined by the accumulation of historical precedents, as was suggested by Lewis. But these precedents depend upon 'the decisions of many individuals who were concerned only with making the best trade at the moment, not with the impact of their decisions on the long-run development of the market' (Young, 1998, 3). This applies equally to economic contracts, dress codes, forms of money and credit, courtship and marriage rituals, rules of the road, etc., and 'they are what they are due to the accumulation of precedent; they emerged from experimentation and historical accident' (Young, 1998, 4).

We can see a nice picture emerging as to how such conventions get started in the first place. In the case of communication, a signal might be salient or naturally salient in some way or another, or a signal might be happened upon by complete accident—an individual makes a noise, accidentally causing another individual to react appropriately (by chance) for some mutually beneficial outcome. Success in coordination makes that particular signal more salient because there is now a precedent for using it in a specific situation. Over time, the

convention becomes fixed, and thus it becomes strongly stable. However, we are already getting ahead of ourselves, theoretically speaking.

Since Lewis (1969) does explain conventions in terms of salience or precedent, one *might* argue (at the risk of hindsight bias) that he has in mind something like a dynamic process unfolding over time. Nonetheless, if this is what he has in mind, it is not explicitly stated as such. Further, he analyses the signalling game in the context of a *classical* game-theoretic framework.[14] Classical game-theoretic analyses of this sort are concerned primarily with notions of stability, and they often presuppose unrealistic rationality requirements. The agents in Lewis' model are assumed to be computationally unbounded insofar as his model makes strong cognitive assumptions about common knowledge and rationality which cannot be appealed to in order to quell sceptical worries about how communication might arise before having a language in place with which to agree upon the necessary conventions. From the classical perspective, it is obvious how conventions are maintained once they are started, but how can they get started in the first place?[15] This requires a shift to a dynamic conception of conventions, which is well-captured by evolutionary, as opposed to classical, game theory.

[14]Note that the structural analogy between certain phenomena in biology and the theory of games was not suggested until Hamilton (1967). This was the same year that Lewis (1967) filed his dissertation, upon which the publication, Lewis (1969), is based. The concepts of *evolutionary* game theory were not made precise until Maynard Smith and Price (1973). Even so, though not explicit, the notion of precedent clearly involves some repetition, and this is explicit in some of the cases that Lewis considers. It is true that Nash (1950b) considered a 'mass action' interpretation of his stability concepts in his (unpublished) Ph.D dissertation, whereby players are repeatedly and randomly drawn from an arbitrarily large population. However, this idea did not make it into his foundational (published) articles on cooperative and non-cooperative game theory (Nash, 1950a,c). Further, Weibull (1995) points out that this mass action interpretation was 'until recently' (circa 1995) unknown. See the discussion in Björnerstedt and Weibull (1993); Leonard (1994); Weibull (1994, 1995); Young (2011). Note however, that the idea of a repeated game—then called a 'supergame'—was already present as early as, e.g., Luce and Raiffa (1957), in addition to several subsequent works. Thus, we can suggest that Lewis (1967, 1969) did *not* have a biological picture in mind; however, it is entirely possible that he was aware of the formalism of repeated games. Furthermore, the examples he discusses often involve *groups* of individuals.

[15]Note that the idea from Lewis (1969) that conventions are equilibria in coordination problems is one such way of explaining why social norms exist; see also Ullmann-Margalit (1977). However, there are other such accounts: norms might be efficient means for achieving social welfare (Arrow, 1971; Akerlof, 1976), they might serve to prevent market failures (Coleman, 1989) or reduce social costs (Thibaut and Kelley, 1959; Homans, 1961), etc. However, many of these explanations are *functional*, rather than causal, and so though they may explain *why* social norms exist, they do not necessarily explain *how* they arise. See the discussion in Bicchieri and Muldoon (2014).

1.1.2 Evolutionary Game Theory

We have already seen that the main idea underlying Lewis' notion of a signalling game is that language can act as a facilitator for coordinating behaviour amongst individuals, and signalling games are based upon coordination problems concerning states and acts. The question, then, is how these conventions come to exist in the first place. Lewis highlighted three possible ways of solving this communicative coordination problem: prior agreement, shared salience, and precedent—i.e., the type of precedent that arises through repeated interactions. However, it has already been noted that the first way of arriving at concordant mutual expectations in this type of coordination problem will not satisfy the desired explanatory function: namely, we want to know how conventional meaning might arise without already having a language in place. Thus, the answer should not beg the question.

Further, though we may well appeal to natural salience as a catalyst for initial coordination, if we end our analysis here, this begins to sound like a *Just So Story*, rather than an explanation.[16] We might additionally like to find the conditions (necessary or sufficient) under which we can *expect* such coordination to arise naturally. An initial success due to natural salience might give rise to precedent, which in turn gives rise to stronger precedent, and so on. Thus, natural salience may be *sufficient* for arriving at signalling equilibria, but is it necessary?

Part of the difficulty that arises in determining *how* such signalling conventions might get off the ground in the first place comes from the following fact. Unless we explicitly build salience into our model, the state-space of the signalling game is entirely symmetric with no special saliences to differentiate one equilibrium from another. Further, if we do build salience into our model, then we necessarily take away from the explanatory capacities of

[16]Note, however, that Hurford (2012, Sec. 2.2) suggests synaesthesia as a possible form of salience that may have contributed to the origins of the first arbitrary associations between words and objects. It is suggested that babies are particularly synaesthetic in the first few months after birth (Maurer, 1993; Maurer and Mondlach, 2005).

the model. However, if it is possible to *evolve* a simple signalling convention in a situation where there are no special saliences, then it is that much more plausible that a successful language might evolve in contexts where there are special saliences, or where the agents utilise more sophisticated learning strategies. As Skyrms (2010a) puts it, 'there may be many signaling systems in nature which got an initial boost from some sort of natural salience. But it is worth considering ... the worst case scenario in which natural salience is absent and signaling systems are purely conventional' (8).[17] So, the question of how we get to any particular signalling system becomes a question of how we might naturally break symmetries. Evolutionary game theory allows us to sidestep the symmetry-breaking problem via some evolutionary dynamic.

The study of communication (understood in this way), now involves a significantly rich framework which draws from a variety of modern conceptual and mathematical tools—including the theory of signalling as a game structure, the mathematical theory of information, the Darwinian theory of evolution as reproductive fitness, and socio-behavioural theories of trial-and-error learning. Taking these tools into account, Skyrms (2010a) points out that communication, in general, should not be understood as 'some evolutionary miracle, but rather as the natural product of some gradual processes' (2). As was mentioned in the introduction, 'gradual' need not mean continuous nor constant-rate but can include varying rates of change over time and small jump discontinuities while maintaining consistency with a Darwinian notion of adaptation.

Game theory was initially developed to understand strategies and conflict among *rational* actors (Nash, 1996).[18] A central question of classical game theory involves optimisation—namely, finding the set of stable equilibrium solutions to a game. These two points give rise

[17] See also the discussion of salience in LaCroix (2018).

[18] Though, as was already mentioned, classical game theory has precursors circa 1700 for gambling; see Footnote 2.

to the following questions: *What if the players are not (ideally) rational? What happens outside of equilibria?*

Evolutionary game theory suggests answers to these questions to the extent that an evolutionary model drops the rationality requirement, and the dynamical nature of the model gives us tools to discuss properties of states that are outside of equilibria. An evolutionary game-theoretic model consists of an underlying game, which tells us the payoffs for interactions between the players, and a dynamic, which tells us how the players' strategies change over time. Thus, a signalling game is characterised by some given structure, which specifies the possible situations that might obtain, the relevant stimuli to which the agents in the game might react, the possible ways in which the agents will respond to those stimuli, and some sort of reward for actions. The dynamic itself might be interpreted in a variety of ways—for example, it may be construed as a biological evolutionary process, in terms of differential reproduction of strategies, or it may be interpreted as a learning process (what we might call *cultural* evolution), in terms of, e.g., simple reinforcement processes. I will explore evolutionary dynamics in Section 1.1.3 and learning dynamics in Section 1.1.4 below. First, I present the underlying structure of the signalling game.

The most straightforward situation of this sort consists of two states of the world, s_0 and s_1, two messages (signals), m_0 and m_1, and two actions, a_0 and a_1. I will refer to this as a 2×2 signalling game, denoting to the dimension of the game in terms of state/act pairs and messages.[19] This can be extended in an obvious way to an $n \times n$ signalling game, where there are n possible states of the world, n possible messages, and n possible actions. Similarly, we can relax the symmetry of this structure by specifying an $n \times m$ signalling game. When $n > m$, the game gives rise to *informational bottlenecks*, and when $n < m$, the game gives rise to synonyms.[20]

[19]Sometimes, to disambiguate underlying assumptions about, e.g., the probability distribution over states, I will refer to a special type of 2×2 signalling game as an *atomic* 2*-game*. See Definition 1.4. This terminology is due to Steinert-Threlkeld (2016).

[20]See the discussion in Skyrms (2010a).

The basic signalling game has two players, whom we call the *Sender* and the *Receiver*. The sender and receiver want to coordinate upon a signalling convention. Assuming the underlying payoffs constitute a game of pure coordination, as was discussed in Section 1.1.1, the players have common interests. However, this symmetry can also be relaxed by altering the payoffs for the game.[21]

When agents play a signalling game, nature picks a state of the world at random. In the simplest case, we will assume that nature is unbiased—i.e., the probability distribution over the set of states is uniform. However, this symmetry can also be relaxed to account for different underlying probability distributions over possible states of the world. The sender observes the state of nature and must choose a signal to send to the receiver to indicate which state has occurred. The receiver sees the signal but cannot directly observe which state of the world obtains, making this a game of *imperfect information*. Thus, she must choose an action conditional upon which signal is sent. Each action is assumed to be a proper response to a particular state in the simplest case. We assume that a_0 is appropriate in s_0 and a_1 is appropriate in s_1; in the general $n \times n$ signalling game, a_i is appropriate in s_i. If the action matches the state, then the play is considered a success, and both the sender and receiver receive some positive payoff. Otherwise, the play is a failure, and they receive nothing (or perhaps some negative payoff, if we extend the game to include punishment for miscoordination). Since signalling (per Lewis (1969)) is a game of pure coordination or common interest, the sender has an incentive to convey information about the state to the receiver via some arbitrary signal; likewise, the receiver has an incentive to correctly interpret the signal by associating it with the correct action.

[21] For example, Donaldson et al. (2007) introduce asymmetric payoffs which they interpret as corresponding to a general-purpose action which is appropriate when the state of the world is unknown. Similarly, asymmetric payoffs may be interpreted as denoting some sort of similarity between states. The underlying game need not be one of pure cooperation. Partial conflict between players is discussed by Crawford and Sobel (1982); Maynard Smith (1991); Dickhaut et al. (1995); Huttegger and Zollman (2010); Zollman et al. (2012); Wagner (2012); Rubin et al. (2016); Ventura (2017), and pure conflict in a signalling context is discussed by Wagner (2012, 2014); Godfrey-Smith and Martínez (2013). Noukhovitch et al. (2020); LaCroix et al. (2020) discuss a spectrum of complete/partial conflict situations in the context of multi-agent reinforcement learning.

A sender *strategy* specifies, for each possible state, which signal to send. Similarly, a receiver strategy defines, for each possible message, which action to perform. As such, the 2×2 signalling game has four possible (pure) sender strategies and four possible (pure) receiver strategies.[22] The sender might send one or the other signal in one or the other state, or she might ignore the state and always send the same signal regardless. Analogous strategies are available to the receiver in choosing an action conditional on the message received. These strategies are summarised in Table 1.4. In general, there are n^n possible strategies for each of the sender and receiver in the $n \times n$ signalling game.

Sender Strategies		**Receiver Strategies**	
S_1:	m_0 if s_0, m_1 if s_1	R_1:	a_0 if m_0, a_1 if m_1
S_2:	m_1 if s_0, m_0 if s_1	R_2:	a_1 if m_0, a_0 if m_1
S_3:	m_0 if s_0, m_0 if s_1	R_3:	a_0 if m_0, a_0 if m_1
S_4:	m_1 if s_0, m_1 if s_1	R_4:	a_1 if m_0, a_1 if m_1

Table 1.4: Sender and receiver strategies for the 2×2 signalling game

The payoffs for all the possible combinations of (pure) strategies for the 2×2 signalling game is given in Table 1.5.[23] When the players coordinate, they receive full payoff; when they miscoordinate, they receive no payoff; when they partially coordinate, they receive full payoff half the time and no payoff half the time for an expectation of 1/2. In general, there are n^{2n} possible combinations of strategies in the $n \times n$ signalling game.[24]

Note that there are two possible ways in which the sender and receiver can achieve perfect coordination in this case. Lewis refers to these combinations of strategies as *signalling systems*. As such, the 2×2 signalling game has 2 possible signalling systems. These are shown in Figure 1.1.

[22] A *pure* strategy, in game theory, provides a complete characterisation of how a player will play the game. As such, it determines what move the player will make for every possible contingency that she might face. This is compared to a *mixed* strategy, where the player assigns a probability to each pure strategy.

[23] Since this is a game of pure coordination, the payoffs are entirely symmetric; thus, we can simply display one payoff in the table, rather than the usual notation (a, b).

[24] See Huttegger (2007a).

		Receiver			
		R_1	R_2	R_3	R_4
Sender	S_1	1	0	$\frac{1}{2}$	$\frac{1}{2}$
	S_2	0	1	$\frac{1}{2}$	$\frac{1}{2}$
	S_3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
	S_4	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Table 1.5: Payoff table for combinations of strategies in the 2×2 signalling game

(a) Signalling system 1 (b) Signalling system 2

Figure 1.1: The two signalling systems of the 2×2 signalling game

In general, there are $n!$ possible signalling systems in the $n \times n$ signalling game.[25] This means, in particular, that as the dimension of the game increases, the number of possible strategies increases significantly more quickly than the number of possible signalling systems, in the sense that

$$\lim_{n \to \infty} \left(\frac{n^{2n}}{n!} \right) = \infty.$$

If we interpret the *meaning* of a particular signal as giving some informational content as to which state obtains, then we can see, from the arbitrariness of the two signalling systems in the 2×2 game, the sense in which language is conventional—i.e., it depends entirely upon which signalling system the players end up adopting, and neither player has any preference over the two signalling systems. Instead, she prefers the one that accords with her partner, regardless of which one it is in fact. This sort of behavioural regularity is a convention in the sense of Lewis (1969); Skyrms (1996); Vanderschraaf (1995, 1998); Young (1998): everyone

[25] Though the symmetric case is not stated explicitly, Lewis (1969) proves that, in an asymmetric signalling game with m states and n signals ($n \geq m$), there are $\frac{n!}{(n-m)!}$ possible signalling systems. As such, when $m = n$, as we have here, it follows immediately that there are $n!$ possible signalling systems.

who acts has an interest in acting in accordance with the convention, though there exist alternative possibilities.

Through repeated plays of the game, actions come to have *precedent* in the sense that the sender and receiver might learn to coordinate their actions based upon past successes and failures. How this happens, and how we interpret it, is going to depend upon the underlying dynamic. The dynamic, essentially, is just a description of how the players in the signalling game differentially change their strategies, given their previous successes and failures. When the sender and receiver play the signalling game such that they are more successful than chance, we might say that they have evolved *efficient* communication—though the system of communication may still be suboptimal. Each of the signalling systems thus constitutes a *maximally* efficient communication convention. In the $n \times n$ signalling game at a signalling system, each state corresponds to a message, and each message corresponds to an action. Thus, there is a bijective mapping from states to actions via messages.

We can formally define a signalling system in a quite general way, as in Definition 1.2.

Definition 1.2. *Signalling Game*

Let $\Delta(X)$ be a set of probability distributions over a finite set X. A *Signalling Game* is a tuple,

$$\Sigma = \langle S, M, A, \sigma, \rho, u, P \rangle,$$

where $S = \{s_0, \ldots, s_k\}$ is a set of *states*, $M = \{m_0, \ldots, m_l\}$ is a set of *messages*, $A = \{a_0, \ldots, a_n\}$ is a set of *acts*, with S, M, and A nonempty. $\sigma : S \to \Delta(M)$, is a function from states to a probability distribution over the set of messages which defines a *sender*, $\rho : M \to \Delta(A)$ is a function from messages to a probability distribution over actions which defines a *receiver*, $u : S \times A \to \mathbb{R}$ defines a *utility function*, and $P \in \Delta(S)$ gives a probability distribution over states in S. Finally, σ and ρ have a

common *payoff*, given by

$$\pi(\sigma, \rho) = \sum_{s \in S} P(s) \sum_{a \in A} u(s, a) \cdot \left(\sum_{m \in M} \sigma(s)(m) \cdot \rho(m)(a) \right).$$

$\Diamond$

The payoff, $\pi(\sigma, \rho)$, for a particular combination of sender and receiver strategies gives us an expectation of the utilities of state-act pairs (given by $u(s, a)$) weighted by the relative probability of a particular state, provided by $P(S)$. This is referred to as the *communicative success rate* of the strategies σ and ρ.

Given this definition, the signalling systems of a signalling game can be defined formally as in Definition 1.3.

Definition 1.3. *Signalling Systems*

A signalling system in a signalling game is a pair (σ, ρ) of a sender and receiver that maximises $\pi(\sigma, \rho)$. $\Diamond$

Note that this definition of the signalling game is reasonably general and allows for a variety of utility functions and probability distributions over states (or signals or acts, at the outset). However, I have been making assumptions about these components for, what I have been calling, 'the simplest case'. It will be useful to introduce some terminology to disambiguate this particular type of game. Thus, following the notation of Steinert-Threlkeld (2016), we can introduce the further definition of an *atomic* signalling game—where states, messages, and actions are equinumerous, the utility function is 1 when the act matches the state and 0 otherwise, and nature is unbiased. See Definition 1.4.

Definition 1.4. *Atomic n-Game*

The *Atomic n-Game* is a signalling game, Σ, with the following particular restrictions:

1. $|S| = |M| = |A| = n$,

2. $u(s_i, a_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta, defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases},$$

and

3. $P(s) = \frac{1}{n}$ for all $s \in S$. ◊

The extensive form of the 2×2 signalling game is given in Figure 1.2.

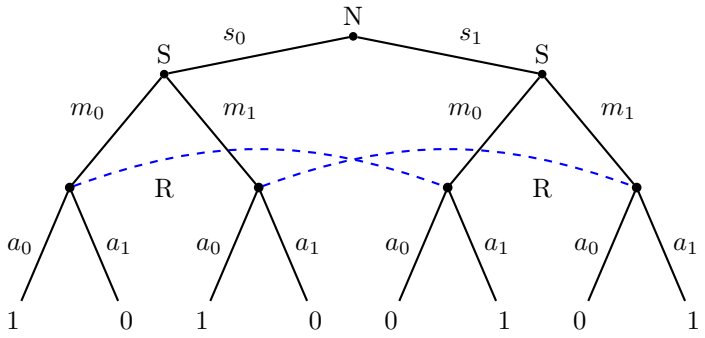


Figure 1.2: The extensive form of the simple 2×2 signalling game. Each node denotes a choice point for a given player, and each branch denotes the possibilities available to her at that point. The dotted lines indicate receiver's information set.

There is a close connection between Lewis' notion of a signalling system and the Nash equilibria of the underlying game structure. This is highlighted by Proposition 1.1.

Proposition 1.1. (Huttegger, 2007a)

Let Σ_n be the atomic n-game. Then the sender-receiver pair, (σ, ρ), is a signalling system if and only if (σ, ρ) is a *strict* Nash equilibrium. ◊

Skyrms (2006, 2010a), Argiento et al. (2009), Huttegger (2007a,b), and others have studied the 2×2 signalling game—and the atomic 2-game in particular—extensively via both simulation and analytic proof. It has been shown, in the atomic case, that under a variety

of dynamics (reinforcement learning and the replicator dynamic, for example), that perfect communication always arises when the states are equiprobable.[26] In the next two sections, I discuss a variety of useful dynamics for the evolutionary signalling game.

1.1.3 Evolution

The Darwinian notion of evolution via natural selection involves a process by which various traits, which affect the ability of an individual to reproduce, are inherited or transmitted. Some natural variation makes this a dynamical procedure, wherein traits that are more fit for a particular environment are more likely to lead to reproductive success. Thus, Darwinian evolution involves natural variation, differential reproduction, and inheritance.[27]

However, Skyrms (2010a, 50) points out that Darwinian processes as these lead to adaptive fitness when the environment is *fixed* (in most cases, at least). The situation is more complicated when the environment itself is dynamic, or when an individual's fitness depends upon other individuals in the population and their strategies, as in cases of vying for a mate or fighting for territory. This is the situation in the signalling-game context, as was laid out by Lewis (1969): the sender's payoff (interpreted now as biological fitness) depends upon what the receiver does, but this is also constituted by dynamic changes over time since the receiver's payoff depends upon what the sender does.

Hamilton (1967) first noted a conceptual analogy between the sex-ratio problem in biology and the theory of games of Neumann and Morgenstern (1944): '[i]n the way in which the success of a chosen sex ratio depends on choices made by the co-parasitizing females, this problem resembles certain problems discussed in the "theory of games". In the foregoing

[26] When the states are not equiprobable, pooling equilibria may be extremely efficient. See Huttegger (2007a) for details.

[27] See Darwin (1859, 1868) and the discussion in Fitch (2010, Ch. 2). Skyrms (2010a) highlights that the ideas of natural selection contained in these works were known to the ancient Greeks. (See especially his discussion in Chapter 4 on Evolution.)

analysis a gamelike element, of a kind, was present and made necessary the use of the word *unbeatable* to describe the ratio finally established' (486).[28] This is a gesture toward something like a stability concept in a biological context. Note, however, that Hamilton (1967) uses quotation marks and italics heavily when discussing the 'gamelike' features of sex-ratios.

The analogy between evolution and game theory is made precise by Maynard Smith and Price (1973); Maynard Smith (1978, 1979), where they introduce the notion of an *evolutionarily stable strategy* as a biological strengthening of the Nash equilibrium concept from classical game theory.[29] This is given in Definition 1.5.[30]

Definition 1.5. *Evolutionarily Stable Strategy (ESS)*

Let $E_J(I)$ denote the expected payoff of strategy I played against strategy J. I is an *evolutionarily stable strategy* if

1. $\forall J, E_I(I) > E_J(I)$, or
2. $\forall J, E_I(I) = E_I(J) \Rightarrow E_I(J) > E_J(J)$. $\Diamond$

The first condition requires that an incumbent strategy, I, does better against itself than a mutant strategy, J, does against I. The second condition requires that if both the incumbent strategy and the mutant strategy do equally against the incumbent strategy, then the incumbent strategy does better against the mutant strategy than the mutant strategy does against itself. Thus, '[i]f in a population adopting strategy I a mutant J arises whose expectation against I is the same as I's expectation against itself, then J will increase by genetic drift until meetings between two Js becomes a common event' (Maynard Smith and Price,

[28]This game-theoretic connection is built upon a Darwinian model of sex ratios given by Fisher (1930). Fisher's own account was already present in Düsing (1883, 1884), though no attribution is made there. See the discussion in Edwards (2000) and Jennions et al. (2017).

[29]In addition to Hamilton (1967), Maynard Smith and Price (1973) also cite MacArthur (1965) as inspiration of their model of the ESS concept. See also, Maynard Smith (1982) and Lewontin (1961).

[30]See Hines (1987) and Hofbauer and Sigmund (1988) for an extended analysis of the ESS concept.

1973, 17). Skyrms (2010a, 52) notes that if the first condition is satisfied, then mutants are driven out rapidly, whereas if the second condition is met, mutants fade away more slowly.

Though the ESS concept is a useful tool for studying the dynamics of natural selection, several authors note that this notion of stability does not fully capture the meaning of *dynamic* stability (Taylor and Jonker, 1978; Zeeman, 1980; Schuster and Sigmund, 1983, 1986; Foster and Young, 1990). Stability is a dynamic concept to the extent that rest states may be stable, but not *strongly* stable. In such a case, small perturbations to the system may subsequently *destabilise* these rest points. There are many such stability concepts in the study of dynamical systems.[31]

A standard evolutionary dynamic, due to Taylor and Jonker (1978), is the *replicator dynamic*.[32] This models a process of natural selection whereby the rate of growth of a particular strategy is assumed to be a linear function of its expected payoff, relative to the average payoff throughout the population of strategies. Payoffs, here, are interpreted as *fitness* in the sense of the reproductive success of a particular strategy, as opposed to preferences over outcomes.

For signalling, this can be modelled in two different ways. The *one-population* replicator dynamic takes advantage of the fact that the underlying game is a symmetric two-player game. Thus, the roles between the players are indistinguishable. Let x_i denote the proportion of the population playing strategy i (i.e., the proportion of individuals of type i in the

[31]For example, we might differentiate between a rest point that is *locally asymptotically stable* versus *globally asymptotically stable*. When trajectories that go near to an equilibrium point tend to stay near to that point, but do not necessarily converge toward it, the equilibrium is called *Lyapunov stable*. These notions apply to deterministic systems, but if our dynamic is stochastic, then we may also define a notion of *stochastic stability*, where, when perturbations to the system are reduced to zero, the probability that the population is in the vicinity of a stochastically stable state does not go to zero. However, throughout this dissertation, I will be concerned more with learning dynamics than deterministic dynamics, and I will be concerned more with 'medium-run' results, rather than limit results. Thus, I will not worry too much about the details of stability concepts, except when the necessity arises.

[32]For more details, see Hofbauer and Sigmund (1998). Taylor and Jonker (1978), Hofbauer et al. (1979), Zeeman (1980), and Hofbauer and Sigmund (1988) show that the ESS concept given by Maynard Smith and Price (1973) is a sufficient condition for local stability under the replicator dynamic.

population). $x = \langle x_1, \ldots, x_n \rangle$ is the distribution of strategies or types in the population, $f_i(x)$ is the fitness of strategy or type i, and $\phi(x)$ is the average fitness of the population, overall. Then the one-population replicator dynamic is given by the following difference equation (1.1), for discrete-time.

Discrete-Time Replicator Dynamic

(One Population)

$$x_i(t+1) = x_i(t)\left[\frac{f_i(t)}{\phi(t)}\right], \qquad \phi(x) = \sum_{j=1}^{n} x_j f_j(x) \tag{1.1}$$

If we consider instantaneous change, then we can derive the continuous-time version of the replicator dynamic for one population, which is given by the differential equation (1.2).

Continuous Time Replicator Dynamic

(One Population)

$$\dot{x}_i = x_i[f_i(x) - \phi(x)], \qquad \phi(x) = \sum_{k=1}^{n} x_k f_k(x) \tag{1.2}$$

In the two-population case, we separate the players into a population of senders and a population of receivers. Let x_i denote the proportion of senders in the sender population playing strategy i, and let y_j denote the proportion of receivers in the receiver population playing strategy j, and let everything else be interpreted as before. Then, we get the *system* of differential equations, given in (1.3).

Continuous Time Replicator Dynamic

(Two Population)

$$\begin{aligned} \dot{x}_i &= x_i[f_i(x) - \phi(x)], \qquad \phi(x) = \sum_{j=1}^{n} x_k f_k(x) \\ \dot{y}_j &= y_j[f_j(y) - \phi(y)], \qquad \phi(y) = \sum_{k=1}^{n} y_k f_k(y) \end{aligned} \tag{1.3}$$

Thus, the replicator dynamic models a situation wherein, when the fitness of a particular strategy is higher than the overall fitness of the population, the rate of change of the proportion of that strategy with respect to time is positive, and so the proportion increases; if a particular strategy is less fit than the overall fitness of the population, the equation is negative, so the proportion of that strategy or type within the population decreases over time.

For sufficiently large populations, these models can be interpreted as differential reproduction (genetic evolution), but Skyrms (2010a) notes that the replicator dynamic might also be understood as a model of differential imitation, or cultural evolution—i.e., individuals imitate strategies with probability proportional to their success. On such a model, when the population is large, imitation of this variety is just the replicator dynamic. However, the interpretation of the payoffs—which were initially understood in terms of Darwinian fitness—will depend upon the application, and this may or may not correlate with Darwinian fitness.

If we consider a two-population model—one population consisting of senders and one population consisting of receivers—for the atomic 2-game, as was given in Definition 1.4, and if we assume no pooling strategies, then the dynamic is a unit-square with the proportion of senders playing strategy S_2 on the x-axis, and the proportion of receivers playing strategy

R_2 on the y-axis. (Recall the strategies given in Table 1.4.) The dynamics for this situation are shown in Figure 1.3.[33]

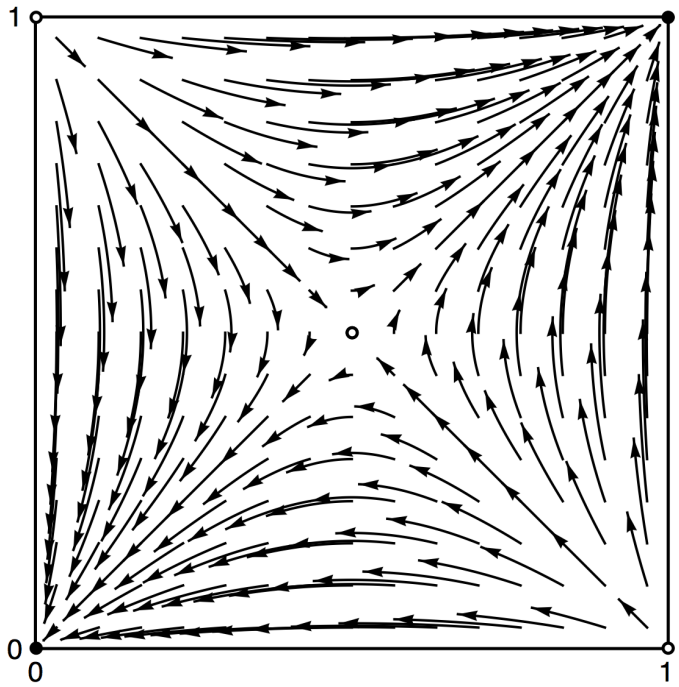


Figure 1.3: 2-population replicator dynamics for the atomic 2-game. The x-axis gives the proportion of the sender population playing strategy S_2; the y-axis gives the proportion of the receiver population playing strategy R_2.

There are five equilibrium points—the hollow circles in Figure 1.3 are dynamically unstable, whereas the solid circles are attractors for the dynamic. The stable equilibria in Figure 1.3 correspond to the two signalling systems of the atomic 2-game. Almost every combination of populations carries the state to one or the other signalling system under this dynamic—note that without mutation, if the populations started at, e.g., exactly the centre node, then it would remain there; however, a slight perturbation would carry them away from this point.

If we account for natural variation in terms of mutation, then we get the *Replicator-Mutator* dynamic (Hadeler, 1981; Hofbauer, 1985), given by the differential equation (1.4) (for the one-population continuous-time dynamic).

[33]This figure was created using the *Dynamo* notebook for Mathematica (Sandholm et al., 2012).

Continuous Time Replicator-Mutator Dynamic

(One Population)

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j(x) Q_{ji} - \phi(x) x_i, \qquad \phi(x) = \sum_{j=1}^{n} x_j f_j(x) \tag{1.4}$$

Here, Q_{ji} is a transition matrix, which gives the probability that type j mutates into type i. Assuming that mutations happen infrequently, it follows that strategies are inherited with high probability, as in the replicator dynamic; however, with some small probability, strategies may mutate into any other strategy. Skyrms (2010a, 61) notes that a cultural interpretation of the (mutation aspect of the) replicator-mutator dynamic might be imperfect imitation—this prevents an absolutely monomorphic culture.

The evolutionary dynamics of simple signalling games have been well studied, and there are several known properties and results, which are noted here. We have already seen that the signalling systems of the atomic n-game are logically equivalent to the strict Nash equilibria of the game. As such, we know that *if* individuals can arrive at a signalling system, then they will not have any incentive to deviate from their respective strategies. However, two questions immediately arise: (1) how robust are these equilibria? And, (2) how (under what circumstances) do individuals arrive at a particular equilibrium in the first place?

For the atomic 2-game, under the replicator dynamic, it turns out that signalling systems *always* evolve. (Recall that in this case, states are equiprobable.) This is given as a theorem in Huttegger (2007a). However, if states are not equiprobable, evolution toward signalling systems is no longer guaranteed, even in 2×2 signalling game—the system may evolve to totally-pooling equilibria. Completely pooling strategies are strategies in which the sender [receiver] simply ignores the state [message] and performs a particular action. For example, strategies S_1 and S_2 [R_1 and R_2] in Table 1.4 are bijective, whereas strategies S_3 and S_4 [R_3 and R_4] are totally pooling. See Figure 1.4.

(a) Total-Pooling Sender Strategy (b) Total-Pooling Sender/Receiver

Figure 1.4: Two examples of total-pooling strategies for the 2×2 signalling game

Part of the reason for this is because when one state is *very* likely, little to no information will be transferred by a signal. However, the individuals might do very well nonetheless. In the 2×2 signalling game with biased nature, we might ask how much bias is sufficient for pooling. Hofbauer and Huttegger (2008) show that there exists a bifurcation at $P(s_i) \approx 0.78, 0.79$. When the probability of each state is less than this, the pooling equilibrium is dynamically unstable; when it is more than this, the pooling point is an attractor. Even in this case, though, the replicator dynamic may lead to a signalling system, it is just no longer guaranteed. In fact, this situation gives a line of equilibria, which is structurally unstable. Uniform mutation collapses the line of equilibria to a single point.[34]

This assumes that the mutation rates for the sender and receiver are equivalent. However, if the receiver's mutation rate is twice that of the sender (i.e., if there is more experimentation on the receiver's end), then pooling equilibria are unstable. (Though, we cannot necessarily assume such favourable conditions.) Further, this analysis is under the assumption of symmetric payoffs. If a particular state is significant (so payoffs are asymmetric), then it may pay to learn a signalling system, even if the state is reasonably rare. As Skyrms (2010a) notes, '[p]redators may be rare, but it does not pay to disregard them' (67). However, this situation is made worse when we increase the dimension of the signalling game—in this case, we may get *partial* pooling equilibria where some, but not all, states (messages) pool to the same message (action).[35] Adding a correlation mechanism can help to destabilise partial pooling

[34]See the discussion in Skyrms (2010a).

[35]Pawlowitsch (2008) gives a complete characterisation of partial-pooling equilibria for simple signalling games under the replicator dynamic and shows that they have basins of attraction with positive measure. See also Huttegger (2007a) and the discussion in Section 1.1.4 below.

equilibria; however, there does not appear to be a pre-theoretic justification for doing so, as we could equally add an anti-correlation mechanism, which can re-stabilise partial pooling.[36]

1.1.4 Learning

The dynamic for the evolutionary signalling game can also consist of some learning procedure. Note that a key (interpretive) difference between evolutionary models, like the replicator dynamic, and learning models is that the former describes how rapidly a strategy multiplies, whereas the latter describes how individuals make choices—i.e., about their own strategies. Of course, 'choice' here need not be understood as a *rational* choice. Choice could be an innate biological (or chemical) behaviour which elicits some reward, as with quorum signalling in bacteria.[37]

There are several different processes by which individuals might learn to adopt a new strategy. In *imitation* learning, the individual may copy the behaviour of others. In this case, payoffs must be *observable* in some sense—for example, if the procedure by imitation is to 'imitate the best' (i.e., strategy), then individuals must be privy to others' payoffs, so that they can determine which strategy has the highest payoff, and thus imitate that strategy. This also presupposes some knowledge of the structure of the game—e.g., that it is a game of pure cooperation, so that the payoff that I would receive if I imitate another player's strategy is also as good as the payoff they received.

Best reply is a learning process by which individuals update their strategies based on some expectation about what others might do. This sort of learning procedure can range from very cognitively (or computationally) sophisticated—such as Bayesian updating—to very cognitively (computationally) minimal—such as taking advantage of a myopic best-response to one's own memory of previous play, referred to as *fictitious play*.

[36] See for example, the exchange between Skyrms (1996), D'Arms et al. (1998), and Skyrms (2000b).
[37] This example will be discussed further in Chapter 2; see Section 2.3.1.

Reinforcement learning describes a procedure by which individuals tend to do what has worked for them in the past. Thus, a simple reinforcement-learning model only minimally requires that the agent playing the game knows her own payoffs; however, she need not know the exact structure of the payoffs, she needs only to know the *relative magnitudes* of the payoffs—i.e., this action in this state produced a better outcome than this other action in the same state. She need not know the payoffs at the outset but may learn them by interacting with her environment. Further, she need not know any other structure of the game, and she need not even know that she is playing a game.

Simple reinforcement learning faces an exploration/exploitation trade-off that can be understood in the context of a 2-armed bandit problem: suppose we have two slot machines that pay off at different and unknown rates. The job of the reinforcement learner is to balance between exploration of the two machines—to not lock on to a bad strategy from early initial successes—and exploitation of the machine that pays higher, on average.[38] As March (1991) puts it:

> Exploration includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation. Exploitation includes such things as refinement, choice, production, efficiency, selection, implementation, execution. Adaptive systems that engage in exploration to the exclusion of exploitation are likely to find that they suffer the costs of experimentation without gaining many of its benefits. (71)

One such learning procedure is given by Bush-Mosteller reinforcement learning (Bush and Mosteller, 1955).[39] In this case, the dynamic is given by the following rule:

$$P_{t+1}(A) = (1-\alpha)P_t(A) + \alpha = P_t(A) + \alpha(1 - P_t(A)), \tag{1.5}$$

[38]See Schumpeter (1934), Robbins (1952), Holland (1975), Kuran (1988), for example. For a thorough introduction to reinforcement learning from a computational perspective, see Sutton and Barto (1998).

[39]See also Macy (1991), Börgers and Sarin (2000), Flache and Macy (2002), Macy and Flache (2002), and the discussion in Skyrms (2010a, Ch. 7).

where A is the action tried, and α is some learning parameter. Thus, 'the probability is incremented by adding some fraction of the distance between the original probability and probability one' (Skyrms, 2010a, 86). To ensure the equations remain tidy for interpretation, the probabilities of alternative actions are decremented so that everything still sums to 1. Note that there is no memory of accumulated reinforcements here.

Skyrms (2010a) points out that under Bush-Mosteller reinforcement learning, as it has been presented here, it is possible to get stuck 'playing the worse slot machine' (88) because learning is too fast—Bush-Mosteller reinforcement learning does not slow down over time. In spite of this theoretical shortcoming, Skyrms (2010a, 98) reports that simple signalling game simulations run under the Bush-Mosteller reinforcement learning model resulted in individuals learning to signal at a very high rate (0.999) for a range of learning parameters ($\alpha \in [0.05, 0.5]$) over $10,000$ iterations.

Whereas Bush-Mosteller reinforcement learning is too fast, an alternative learning model, called Roth-Erev (or *Herrnstein*) reinforcement learning, slows down over time. This gives it the 'Goldilocks property' (learning is *just right*) in that it always converges to playing the optimal machine with probability 1.[40]

This alternative reinforcement learning model has been empirically tested in a lab setting by Roth and Erev (1995); Erev and Roth (1998). They examine a family of simple dynamical models of learning to see how well they accord with real-world psychological data from previous experiments, which sophisticated rational choice fails to explain (Roth et al., 1991; Prasnikar and Roth, 1992). It is highlighted that an adequate model of learning should accord with two basic observed principles to maintain consistency with extant data on human and nonhuman animal learning. In particular, they point to the *law of effect*, whereby choices

[40]This is proved by Beggs (2005). See also Wei and Durham (1978); Pemantle (2007); Catteeuw and Manderick (2014).

that lead to success are more likely to be chosen in the future, and the *power law of practice*, whereby learning curves start steep and flatten out as time goes on (Blackburn, 1936).

These basic properties of learning have a long psychological pedigree. Thorndike (1905, 1911, 1927) showed, in a series of experiments, how animals learn behaviours. This is called the *law of effect* for the conditioning of stimulus-response relations by experience:

> Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation be weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond. (Thorndike, 1911, 244)

Namely, when the effect of an action is 'pleasing', it tends to strengthen the connection between the stimulus and that action, and when the effect is 'displeasing' it tends to weaken it. In this case, 'pleasing' might refer to the receipt of some reward upon acting, and 'displeasing' might refer to punishment as a result of acting.

Thus, it was said above that a learning model describes how individuals make 'choices', and that there are several degrees of computational complexity that might be inherent in a learning model. Further, it was suggested that reinforcement learning could be understood as a straightforward model, requiring only that the agents know their own payoffs. Given the relation of reinforcement learning to Thorndike's law of effect, we see a clear example of *how* 'knowing one's own payoffs' need not refer to any sort of cognitive knowledge about an underlying game structure. Instead, the individual need only be sensitive to positive and negative outcomes of her own actions, and possible the comparative magnitudes of these payoffs—for example, two different actions that produce positive rewards may differ in comparative magnitude; this requires further a notion of 'better' and 'worse' in addition to

'positive' and 'negative'. However, noticing a difference in magnitude does not require much complexity.

Herrnstein (1970) quantified this principle as the *matching law*, which states that the probability that a particular action is chosen is proportional to the accumulated reward received from choosing that action. Roth-Erev reinforcement learning is based precisely upon this idea. Thus, we begin with some initial 'inclination weights'—these may be equiprobable at the outset, so that action is initially random. The weights evolve by addition of rewards received for actions, and the probability that a particular action is chosen in the future is proportional to these weights.

There are a variety of ways in which these learning dynamics can be altered. For example, one can introduce negative payoffs to account for punishment, or one can add a 'forgetting parameter'. Similarly, one might change the response rule itself. For example, Blume et al. (2002) introduce a softmax function as an alternative response rule for Roth-Erev reinforcement learning. This approach defines an exponential response rule, where the probabilities (of choosing a particular action) are proportional to the exponential of the accumulated rewards rather than the accumulated rewards themselves. Generally, the accumulated rewards are multiplied by a constant, λ—the reciprocal of λ is sometimes called the 'temperature'. When λ is 0, the reciprocal is infinite, so the response rule is uniform among the possibilities. Similarly, when λ is arbitrarily large, the reciprocal approaches 0, and the act with the largest accumulated reward is chosen with the highest probability. LaCroix (2018) considers a model similar to softmax with λ close to 0 and interprets this as a sort of salience parameter.

There is a close relationship between evolution and learning in this context. Börgers and Sarin (1997) show that the mean-field dynamic for Bush-Mosteller learning is a version of the replicator dynamic, and Beggs (2005) and Hopkins and Posch (2005) show that the same is true for the mean-field dynamic of Roth-Erev learning—it is a version of the replicator

dynamic. We saw previously that the replicator dynamic updates ratios of strategy types in a population of various strategies by process of incrementally increasing, in subsequent generations, the representation of types that are more successful in the population overall. This is analogous to how simple reinforcement learning models update a single agent's dispositions over individual trials or interactions.

In the context of learning to signal, Roth-Erev reinforcement learning can be understood intuitively as an *urn-learning* procedure. I will consider, again, the atomic 2-game; see Definition 1.4. Assume now that the sender has an urn labelled s_0, and an urn labelled s_1. Similarly, the receiver has an urn labelled m_0 and an urn labelled m_1. At the outset of the game, each of the sender's urns is equipped with one ball for each possible message at her disposal—m_0 and m_1. Similarly, each of the receiver's urns contains a ball for each of her possible actions, which are labelled a_0 and a_1. See Figure 1.5.

(a) Sender Urns (b) Receiver Urns

Figure 1.5: Simple reinforcement learning model

Note that even this simple context presents a challenge for learning to signal successfully: the state-space is entirely symmetric, there are no saliences, and this updating procedure is (computationally) one of the most straightforward possible learning strategies.

On each play of the game, the state of the world is chosen at random. The sender then selects a ball at random from the urn corresponding to the state of the world and sends that message to the receiver. The receiver then chooses a ball at random from the urn corresponding to the message received. If the action matches the state of the world, then the sender and the receiver both reinforce their behaviour by returning the ball to the urn from which it was chosen and adding another ball of the same type to the urn from which the original ball was

chosen. If the action does not match the state, then each player simply returns the drawn ball to the urn from which it was drawn. In the most basic case, there is no penalty for miscoordination, though it is possible to model punishment by discarding a ball when it led to a failure. The game is then repeated for a newly chosen state.

The dynamic shifts strategies to the extent that adding balls to an urn for a successful action shifts the relative probability of picking a ball of that type on a future play of the game. Adding balls to a particular urn changes the conditional probabilities of the sender's signals (conditional on the state) and the receiver's acts (conditional on the signal). Thus, the conditional probabilities of the sender's signals and the receiver's actions change over time, and the players become more likely to perform previously successful actions.

More formally, in the signalling-game context, Roth-Erev reinforcement learning can be modelled by keeping track of the accumulated rewards given by the function $ar_{\sigma,t} : S \times M \to \mathbb{R}$ for the sender's reward at time t, and $ar_{\rho,t} : M \times A \to \mathbb{R}$ for the receiver's reward at time t, with some initial inclination weight for $t = 0$. In the simplest case, the weights are all equivalent to start. The accumulated rewards build up according to the following update rules for the sender and receiver:

$$\begin{aligned} ar_{\sigma,t+1}(s_i, m_j) &= ar_{\sigma,t}(s_i, m_j) + u(s_i, a_k), \text{ and} \\ ar_{\rho,t+1}(m_j, a_k) &= ar_{\rho,t}(m_j, a_k) + u(s_i, a_k), \end{aligned} \tag{1.6}$$

where s_i, m_j, and a_k are the state, message, and act that were played at time t. The sender and receiver thus choose their rewards according to the law of effect, whereby their propensities to pick a particular message or act are proportional to the accumulated rewards for that message or act:

$$\begin{aligned} \sigma_{t+1}(s_i)(m_j) &\propto ar_{\sigma,t}(s_i, m_j), \text{ and} \\ \rho_{t+1}(m_j)(a_k) &\propto ar_{\rho,t}(m_j, a_k). \end{aligned} \tag{1.7}$$

This gives the dynamic for simple reinforcement learning in the signalling game. This formalism is slightly more general than the urn-learning metaphor since it does not require that the utility function has only integer values. Simple reinforcement learning of this form is standard in behavioural psychology.[41]

Skyrms (2010a, 94) notes that, though this reinforcement learning procedure is straightforward, it is still very effective for learning how to signal. Consider, for example, the 2×2 signalling game, with unbiased nature, unbiased initial propensities, and reinforcement for success equal to 1—i.e., the atomic 2-game. On simulation, after 100 iterations, the communicative success rate for the sender and receiver is $\pi(\sigma, \rho) \approx 0.8$, on average; after 300 iterations, the communicative success rate for the sender and receiver is $\pi(\sigma, \rho) \approx 0.9$, on average. The limiting results of this case give the following theorem for signalling under simple reinforcement learning.

Theorem 1.1. (Argiento et al., 2009)

In the *Atomic* 2*-Game* (i.e., the 2×2 signalling game, with unbiased nature, unbiased initial propensities, and reinforcement for success equal to 1),

$$\lim_{t \to \infty} \pi(\sigma_t, \rho_t) \to 1.$$

Furthermore, the two signalling systems are equally likely to occur; that is, with probability 1/2,

$$\lim_{t \to \infty} \frac{\sigma_t(s_0, m_0)}{\sigma_t(s_0, m_1)} = \lim_{t \to \infty} \frac{\sigma_t(s_1, m_1)}{\sigma_t(s_1, m_0)} = \lim_{t \to \infty} \frac{\rho_t(m_0, a_0)}{\rho_t(m_0, a_1)} = \lim_{t \to \infty} \frac{\rho_t(m_1, a_1)}{\rho_t(m_1, a_0)} = 0,$$

and with probability 1/2, the limits of the reciprocals of these fractions are equal to 0. ◊

[41] See, for example, Bush and Mosteller (1955); Suppes and Atkinson (1960); Arthur (1993); Roth and Erev (1995, 1998); Börgers and Sarin (1997, 2000); Bruner et al. (2018).

However, Theorem 1.1 only holds for the atomic 2-game, which was the most straightforward scenario. In more complex games—e.g., games of higher dimensionality or with unequal initial propensities—suboptimal equilibria may develop and prevent uniform convergence to perfect signalling. A large body of subsequent work is devoted to finding out how far these positive results generalise. When $n = 2$, assuming only pure strategies, we see that the strategies can either be bijective or totally pooling. However, when $n > 2$, it is possible to have *partial* pooling strategies. See Figure 1.6 for an example of partial pooling in the 3×3 signalling game.

(a) Partial-Pooling Sender Strategy (b) Partial-Pooling Sender/Receiver

Figure 1.6: Two partial-pooling strategies for the 3×3 signalling game

It was mentioned in Section 1.1.3 that Pawlowitsch (2008) gives a complete characterisation of these partial pooling equilibria and shows (assuming the replicator dynamic) that the partial pooling equilibria have basins of attraction with positive measure. We might imagine, given the connection mentioned between the replicator dynamic and reinforcement learning, that the same is true under the learning dynamic. Indeed, this is the case. Some evidence for this is given in terms of simulation results. Table 1.6 shows run failure-rates for the atomic n-game, for a variety of n, with $n > 2$.[42]

Model	**Run Failure Rate**
$n = 3$	0.096
$n = 4$	0.219
$n = 8$	0.594

Table 1.6: Run failure rates for the atomic n-game under simple reinforcement learning

[42]More details are given in Barrett (2006).

In this case, there are 10^3 runs for each model, with 10^6 plays per run. A 'failure' occurs when the players' success rate is less than 0.8 at the end of a run. In every case, the players learn to communicate better than chance, so Barrett (2006) takes these conventions to be 'more or less effective'.

Addition of mutation to the replicator dynamic, we saw, will destabilise total pooling equilibria and also appears to destabilise partial pooling equilibria by reducing the size of the attracting region (Huttegger et al., 2010). Similarly, in the learning model, the addition of some extra machinery—such as a correlation mechanism (Skyrms, 2010a; Barrett, 2006) or a salience parameter (LaCroix, 2018), for example—can help individuals learn how to signal.

However, this is still the simplest case, and several different parameters can be altered at the outset. The choice of initial parameters can significantly alter the outcome of the game. For example, in the atomic n-game, it is assumed that all states are equiprobable. This is unlikely to be true in nature. As with the evolutionary dynamic, under the learning dynamic when some states are improbable, the receiver may ignore the signal, and simply perform the same action every time, for a reasonable payoff. However, in the case of extremely asymmetric state probabilities, it may also be the case that payoffs are asymmetric.[43]

1.1.5 Some Generalisations

We have already seen that the atomic 2-game can be extended in several ways by varying the initial parameters. For example, we can increase the dimension of the game; relax the symmetry between the number of states, signals, and actions; change the payoff structure or the underlying dynamic; in an evolutionary context, we can add correlation and mutation; in a learning context, we can add punishment or salience; etc. There are other ways of

[43]See Skyrms (2010a); Donaldson et al. (2007).

extending or generalising the atomic n-game in addition to altering these initial structural parameters.

For example, signalling may occur in structured networks with a well-defined topology. Zollman (2005) examines learning to signal with neighbours on a grid, using imitation dynamics, where each individual can observe each of her eight neighbours and imitates the best strategy that she sees. (The topology of the network is a 100×100 grid, placed on a torus so that each of the $10,000$ individuals on the grid has eight neighbours.) In this case, it is found that alternative signalling systems can co-exist, but occupying different regions of the topography. This has been likened, in the literature, to different regional 'dialects'.

Wagner (2009) extends this analysis and compares neighbourhood interactions with more complex network structures. Wagner (2009) shows that the behaviour of the system—i.e., whether and when populations converge to signalling systems under a variety of starting assumptions—depends significantly on the topological structure of the network itself. One of the main results of his argument is that the topological structure of so-called 'small-world networks' is very conducive to the efficient evolution of meaning.[44] This is significant because many real-world social interactions, *in fact*, take place in small-world networks.

Mühlenbernd (2011) also examines the evolution of signalling in a structured spatial society. This is similar to the networks discussed in Zollman (2005); Wagner (2009), except the agents can 'choose' to interact with more distant neighbours in the community. The choice is established by a 'degree of locality' parameter, whereby an individual chooses to interact with a neighbour with some probability, determined by the (Manhattan) distance of that neighbour from herself and the degree of locality.[45] This parameter fills the gap between

[44]'Small world network' is a technical term characterised by a graph with a certain set of properties—e.g., high clustering coefficient (of nodes), short average path length (between nodes), etc. For example, many forms of the underlying architecture of the internet are small-world networks. See Watts and Strogatz (1998); Humphries and Gurney (2008); Telesford et al. (2011)

[45]The Manhattan distance, also called the taxi-cab distance, between two points is measured along axes at right angles (as a taxi cab driving through Manhattan).

Zollman (2005) and Wagner (2009), whose models are at the extremes of the scale that Mühlenbernd (2011) introduces. (For example, the higher the degree of locality, the more probable that an agent will choose to interact with her immediate neighbours.)

There are many different ways of generalising the simple signalling-game context to explain various phenomena in question, or to attempt to have one's models accord more realistically with real-world situations under which signalling might evolve. One way to do so, as we have seen in Sections 1.1.3 and 1.1.4, is to change the dynamic so that it accords with empirical evidence.

Additionally, there are several parameters in the basic signalling game that might be varied. We have already seen some results from a variety of situations in which, for example, the states are not equiprobable. We have also seen how an increase in dimensionality in the $n \times n$ game affects the learning outcome. However, there is no reason why we should expect the states of the world, acts, and possible signals to be equivalent. There may be asymmetry here: if there are more state-act pairs than there are signals, then we have a situation wherein informational bottlenecks occur, whereas when there are more signals than state-act pairs, synonyms might arise.[46] There is also no requirement that there should be the same number of appropriate acts as there are states, in which case a signal might carry *disjunctive* information about the state.[47]

It is worth noting that 'evolution of a signaling system is *evolution of a system of categories*' (Skyrms, 2010a, 109). As such, two distinct things are happening here: the sender and receiver must simultaneously partition nature into distinct *kinds* and further code for representing/interpreting those kinds using available signals (Barrett, 2007). This is true in the symmetric signalling game; however, it is more evident in asymmetric versions. In

[46]For more signals, see Wärneryd (1993). For more state-act pairs (though with asymmetric payoffs) see Donaldson et al. (2007). For a more general analysis of information transfer in the case of informational bottlenecks, see LaCroix (2020a). See also, Hu et al. (2011).

[47]See Skyrms (2010a, Ch. 9).

this way, it is possible to model a signalling game to represent more or less coarse-grained information. For example, it is well known that vervet monkeys have distinct alarm calls for distinct predators: snakes, eagles, and leopard.[48] In this case, we might think of more coarse-grained as separating an aerial predator from a terrestrial predator, instead of separating each particular predator. Thus, we have different categories of varying specificity. The states are now 'eagle', 'leopard', 'snake', and 'snake or leopard', for example. The evolution of systems of categories, Skyrms (2010a) notes, can happen 'without any complex rational thought' (116-117)—instead, it is a consequence of the dynamics in question.[49]

It has been suggested (Skyrms, 2000a, 2004, 2010a) that the Lewis signalling game might be modified slightly to account for the evolution of certain *logical* notions, such as inference. Consider the alarm call system of vervet monkeys, where there are three possible states ('leopard', 'eagle', or 'snake'), and appropriate actions for each. Suppose there are two senders in such a signalling-game context. If one sender—with partial information about the state—can send a disjunctive signal (e.g., 'snake or leopard'), and the other sender—with different partial information about the state—can send a negative signal (e.g., 'no leopard'), then the receiver, upon hearing *both* of these signals, might learn to perform the correct action—in this case, 'stand tall scan the ground'.[50]

In a two-sender, one-receiver signalling game, each signal conveys perfect information about a *coarse-grained* partition of nature, and the combination of signals (can) convey perfect information about a *fine-grained* partition of nature—for example, $\{s_0, s_1\}$ partitions nature in a coarse sense, and so does $\{s_0, s_2\}$. However, their combined information partitions nature in a fine-grained sense—i.e., $\{s_0, s_1\} \cap \{s_0, s_2\} = \{s_0\}$.

[48] This is discussed in more detail in Chapter 2.

[49] See also, Skyrms (2000a).

[50] Skyrms (2010a) calls a signal that encodes the disjunctive state *proto-truth-functional*. Whether such a signal is functionally compositional or taken as atomic is not examined here. It is at least possible for negative signals to arise in nature, however. In the initial work, Skyrms assumes that the proto-truth functions were already evolved; however, in later work he proposes models for seeing how and whether they will evolve.

Barrett (2007, 2009) shows how the two senders and the receiver can interact to simultaneously and spontaneously partition nature and code for the categories, thus partitioned.[51] This can be re-interpreted as a *syntactic* signalling game, with one sender who has the option of sending more than one signal (Skyrms, 2010a). This game is discussed in more detail in Chapter 4. Skyrms' overarching point is that logic should be seen as a way of processing information. The currency of the signalling game is information transfer; therefore, we should be able to discuss at least some basic logical principles in terms of the signalling-game context.[52]

1.1.6 Rule Following and Dispositions

Signalling games thus model simple communicative interactions between individuals where some coordination is necessary. It is in this sense that the theory of signalling games emphasises the social aspects of language. In an evolutionary context, signalling games can provide some fundamental insights into how conventionally meaningful communication might *possibly* arise in a natural way and without an antecedent language. As such, though the Lewisian story of natural salience giving rise to precedent is sufficient for these purposes, it is not necessary. It is important, of course, to note that we do not (and likely cannot) know how these things actually came about. However, without the signalling-game framework, we might find scepticism in line with Rousseau (1755), Russell (1922), or Quine (1967) appealing[53]

[51] See also, LaCroix (2020a).

[52] See also, Steinert-Threlkeld (2014).

[53] Of course, this sort of scepticism has a long philosophical tenure; for example, the Cratylus dialogue (Plato, 1921a) discusses whether the sounds of words have some sort of *natural*, necessary connection to their meanings, or whether such mappings are arbitrary. The diversity of ways to denote the same thing in different languages seems, *prima facie*, evidence for the latter view. For example, a domesticated carnivorous mammal that typically has a long snout, an acute sense of smell, and a barking, howling, or whining voice might be denoted by any of the following: 'Chien' (French), 'Kutya' (Hungarian), 'Inu' (Japanese), 'Köpek' (Azerbaijani), 'Chó' (Vietnamese), 'Hond' (Afrikaans, Dutch), 'Hund' (German, Norwegian, Danish, Swedish), 'Koira' (Finnish), 'Mbwa' (Swahili), 'Car' (Tajik), 'Eey' (Somali), 'Cachorro' (Portuguese), 'Pies' (Polish), 'Caine' (Italian), 'Madraí' (Irish), 'Dog' (English), etc. However, given that several of these words have obviously identical roots, some particulars may be explained away via some sort of drift—that is, perhaps the original word *did* have some 'natural' association with the thing in question, and over time as

The main problem with this sort of philosophical stance is that it ignores fundamental questions about how communication gets off the ground in the first place. On the subject of meaning, the sceptic is left with a violent circularity that appears to support the sceptical stance—however, this ignores the apparent conjunction of empirical facts that language exists and that, at some prior point in history, language did not exist. Therefore, something must have given rise to the ability to communicate. Indeed, as we will see in Chapter 2, some form or other of communication is utterly ubiquitous in nature.[54] The game-theoretic framework thus provides a precise mathematical language with which to discuss and, perhaps more importantly, test such philosophical intuitions.

This further highlights that it makes no sense to study language as some sort of structure that exists out in the world and independently of anything else. Linguistic phenomena depend inherently upon the context in which they are used. So, to study language philosophically means to understand the use of a language within a community of speakers and the circumstances under which it is used. Language does not occur in a vacuum. Formal language theory, in linguistics, treats languages as sets of sentences, each of which is an independent object, and a grammar is supposed to characterise the members of that set without any reference to actual contexts of use.[55] Philosophers tend to do worse.

Thus, the theory of signalling games reminds us that language serves a purpose: namely, to communicate. Therefore, by accounting for communication conventions, this sort of philo-

---

dialects became varied and evolved into new languages, these became *slightly* different. However, many of these languages do not have the same roots—if there were a *necessary* connection between the thing in the world and the sound, then even disparate languages should use similar sounds.

[54]Of course, the idea of the conventionality of signals and the importance of social relations in the study of communication is not new. This has arisen repeatedly throughout the history of philosophy; see, for example, (Barnes, 2001; Morgan, 1967; Hume, 1739; Smith, 1761, 221).

[55]For a criticism of this approach, see, for example, Givón (2002a): 'there is something decidedly bizarre about a theory of language (or grammar) that draws the bulk of its data from ... out-of-context clauses constructed reflectively by native speakers' (74-75). Hurford (2012) argues there are at least some cases, in the study of syntax, where discourse does not matter (183); however, he also points out that a lot of the most interesting data that is used in theorising about syntax is derived from communicative use (186) and that 'many facts of central interest to syntacticians exist for discourse reasons, because of the communicative purposes to which language is put' (190).

sophical approach can safeguard against unnecessary abstraction. Further, rather than being based solely upon intuitions about natural language, or theorising about the 'actual' underlying structure of ordinary language, the signalling game approach provides a rigorous set of tools for examining language *use* in a population of speakers and the conditions under which they might arise. This can account for meaning, in the context of information transfer (Skyrms, 2010a,b), meta-linguistic notions of truth (Barrett, 2016, 2017), and reference. Thus, the view of Dretske (1981)—that epistemologists should not spend their time on 'little puzzles' or rehashing ancient arguments about scepticism (Skyrms, 2010a, 33)—applies equally well to philosophers of language. The theory of signalling games provides a method for the philosophical study of language without reference to intuitions, and without dependence upon little puzzles.

1.2 An Aside About Applicability

In the introduction, I mentioned some virtues of modelling and simulations for the study of language evolution.[56] As we have seen, explicit evolutionary game-theoretic models can shed light, for example, on the evolution of vocabulary and meaning. Many researchers who utilise formal models acknowledge the limitations of these tools. As Progovac (2019) highlights, computer simulations 'provide a novel empirical way of testing plausibility of evolutionary hypotheses, even when they cannot themselves directly confirm or refute such hypotheses' (61). It is crucial, given the multi-component approach advocated by Fitch (2010, 2017), to use such methods in conjunction with empirical evidence.[57] Here, I address some specific concerns regarding the use of the signalling-game framework in particular and computer simulations in general.

[56]See Cangelosi and Parisi (2002). For an overview of simulation studies in the evolution of language, see Steels (2011); Jäger (2008).

[57]See also Toya and Hashimoto (2015).

Evolutionary game theory, in general, has become a standard tool used in evolutionary biology, and the utility of such models has been widely accepted (Maynard Smith and Szathmáry, 1995). In spite of this, Fitch (2010) points out that analyses of this sort have been 'surprisingly rare' in discussions of the evolution of language (51). Indeed, several works that aim to be interdisciplinary have come out in the last decade. Many of these urge for more collaboration between researchers in disciplines relevant to the evolution of language. This includes, at the very least, philosophy, modern biology (including neo-Darwinian evolutionary theory, developmental and molecular genetics, and neuroscience) and the contemporary language sciences (including theoretical linguistics, psycholinguistics, and comparative linguistics) psychology, cognitive science, neuroscience, anthropology, sociology, ethology, etc. (Hurford, 2007; Fitch, 2010; Platt, 2018).[58] However, very few of these discuss evolutionary models for testing hypotheses—though Hurford (2012, Sec. 2.4) spends some time discussing criticisms of this form of analysis for the evolution of language. While such analysis may be fruitful when done correctly, it is worthwhile to consider these objections briefly, or at least to keep the explanatory shortcomings of this type of modelling technique in mind as we proceed.

At the outset, it was pointed out that the theory of signalling games gives a possible answer to the question of how a convention arises from scratch with minimal presuppositions. This requires social coordination, agreement, and continued use within a population, without (initially) having any communicative means for agreeing upon the conventional code. Hurford (2012, 138) points out that, minimally, these sorts of models take the following for granted:

[58] On the subject of urging such interdisciplinary work, here is a small sample: 'clearly more interdisciplinary dialogue is needed' (Hurford, 2007, 287); 'The data that can help resolve the perennial issues of debate in language evolution come from so many different disciplines ... Thus researchers must cooperate with others to achieve a broader and more satisfactory picture. Answers to the difficult questions about language evolution, which include some of the deepest and most significant questions concerning humanity, require interdisciplinary teamwork of a sort that remains regrettably rare in this field. My central goal ... has been to increase the potential for such collaboration by providing access to the insights from many relevant fields to any interested reader' (Fitch, 2010, 3-4); 'The ideas summoned in this brief, yet powerful, book endorse the hypothesis that we will answer this, and other challenging questions, only through interdisciplinary dialogue and investigation' (Platt, 2018, 6).

1. A population of individuals capable of internal mental representations of the meanings to be expressed, i.e. prelinguistic concepts, from some given set;

2. An assumed willingness to express these concepts, using signals from a predetermined repertoire, to other members of the population;

3. An ability to infer at least parts of meanings expressed by others from an assumed context of use;

4. An ability to learn meaning-to-form mappings based on observation of their use by others.

The last assumption (4) simply describes the dynamic in question. Without assuming *some* kind of dynamical process, it is unclear what explanation we might be able to give of how conventions of communication are socially transmitted within and across generations in the population. Further, the wording that Hurford (2012) uses here seems to imply that learning is the only dynamic process that he takes such evolutionary game-theoretic models to involve. We have seen in this chapter that this assumption is, at best, false and, at worst, disingenuous. Whether and how quickly a signalling convention evolves might depend upon the dynamic in question, but by exploring a variety of dynamics, and noting the differences that these make for the outcome of the signalling game, we can make stronger claims.

Assumption (3) seems fine since 'meaning' can just be understood as 'information' in the sense of Skyrms (2010a,b). There need not be any assumption of inference here, in any case. Such abilities may arise merely by association. There is no outlandish requirement here (see, for example, Skinner (1948)), and these abilities are empirically well-grounded. The simplest models of evolutionary signalling—e.g., simple reinforcement learning—start with the most basic requirements to see what is at least possible at such a level of complexity. A bona fide explanation of complex communication will likely require some more cognitive machinery, but then these modelling choices will need to be justified by empirical work.

Assumption (2), I take it, is two possible issues: the first is with willingness, the second is with 'predetermined repertoire'. In the case of willingness, there is a slight tension between actual payoffs in the underlying context and the theorised symmetry of payoffs in the game. That is, in many actual-world cases, signals are not cost-free. For example, in sending an alarm call, the receiver has the benefit of learning of a predator's presence, without the cost of alerting the predator to her location. However, there are several notions involved here, many of which have been the subject of specific research programmes in game theory. These include reciprocal altruism, kin selection, and inclusive fitness, to name a few. To communicate, individuals need to cooperate; in order to cooperate, individuals need to trust one another. However, underlying all of these facts—cooperation, trust, altruism, etc.—is the existence of a social group.

Given that the purpose of communicative abilities is *to communicate*, the existence of a (stable) social group is a *necessary* condition for the evolution of signalling.[59] Hurford (2007) himself points out that 'a crucial precursor to the appearance of these proto-linguistic abilities was not in itself a specifically linguistic change, but rather a shift in the normal social relationships between individuals in a group' (244), and further,

> Given the emergence, somehow, of such trusting and cooperative social arrangements, the ancestors of modern humans found themselves in an environment where learned meaningful signals were advantageous, and there would have been pressure for the shared vocabularies to grow culturally, which in turn would have exerted pressure for a capacity to learn larger vocabularies to evolve by biological natural selection. (Hurford, 2012, 113)

So, it seems that the willingness to socialise is not a theoretical assumption that needs to be justified. It is required for the phenomena in question.

[59]See, for example, Jackendoff (2007); Seyfarth et al. (2005); Cheney and Seyfarth (2007); Fitch (2010), and the discussion in Chapter 2 below.

Concerning the predetermined repertoire issue, we should note that there is nothing that *requires* there to be a fixed or predetermined repertoire of possible messages—indeed, Skyrms (2010a) addresses this 'artificial limitation' (118). Of course, the availability of even potential signals is going to be constrained in some way by biological limitations—i.e., the physical ability to create specific phonemes, in the case of spoken communication. This ability is significantly reduced in apes and monkeys as compared with humans, due to anatomical constraints placed on sounds production.[60] Moving further down the line, bacteria do not use sounds to signal, but rather biochemical processes; therefore, these signals are significantly constrained by molecular biology. Within these constraints, it is still possible to create new signals, so long as there is evolutionary pressure at the outset. Thus, there is no need to assume a fixed repertoire—Skyrms (2010a, Ch. 10) offers a simple first pass at how we might model something like the invention of signals.[61]

The urn-learning process we examined in Section 1.1.4 is a Pólya urn process, wherein the urns are initially populated with our fixed vocabulary (or fixed set of actions). However, a Hoppe-Pólya urn model (Hoppe, 1984) is one in which we start with only a black ball in each of the urns—this is the *mutator*; when a black ball is chosen, it is returned to the urn, and a ball of a new colour (or type) is added to the urn—this corresponds to the invention of a new signal.[62] Thus the receiver always has the additional action—*send a new signal*—available to her. Skyrms (2010a) points out that, in general, even for 'quite a large number of trials, the expected number of categories is quite modest' (126)—around 12 categories for $100,000$ trials, for example. If we modify the Hoppe-Pólya urn model by adding differential reproduction, this results in reinforcement learning with invention.

[60]For example, 'voice' in humans takes advantage of the diaphragm, chest muscles, ribs, abdominal muscles, lungs (for regulation of air pressure, causing the vocal folds to vibrate); the larynx and vocal folds; and the pharynx, oral cavity, and nasal passage. For example, the tube, which air passes through from the glottis to the lips, is curved in apes but forms a right angle in humans—these different physical shapes place limitations on the types of sounds an individual can produce.

[61]See also Alexander et al. (2012).

[62]In this metaphor, the balls are 'labelled' by colour, rather than the label of our fixed repertoire of signals, m_0, m_1, etc.

In the case where there is some fixed number of signals at the outset, plus the possibility for invention, the players might either proceed as normal, or they may try to use a new signal without success, in which case the game progresses as usual; or, they may try to use a new signal with success, in which case the game moves from a signalling game with m signals to a signalling game with $m+1$ signals. Skyrms (2010a) points out that if we can model such a game, then there is no principled reason why we should not start with $m = 0$ signals available to the sender—so, there is no predetermined repertoire. In a 3-state, 3-act signalling game with no signals, the players ended up with anywhere between 5 and 25 signals, with a mode of 13, after $100,000$ iterations.[63] Twenty-five signals might seem excessive to represent three states of nature; but, the players invent synonyms, which help them to avoid polymorphic traps, so the possibility for the invention of signals makes signalling *systems* a more robust phenomenon.[64] Many of the synonyms invented do very little work in the signalling system thus evolved. However, it is also possible to add a 'forgetting' parameter to prune less useful signals. Skyrms (2010a) points out that there are certain ways of modelling this that 'can be remarkably effective in pruning little-used signals without disrupting the evolution of efficient signaling. Often, in long simulation runs, we get close to the minimum number of signals needed for an efficient signaling system' (134-135). So, invention can help efficacy, and forgetting can help efficiency.[65]

Note also that redundancy in signals exists in nature. The translation of RNA is a non-isomorphic process wherein sequences of 3 adjacent bases—called *codons*—are converted into amino acids. While there are 20 amino acids to code for, there are $4^3 = 64$ possible combinations for coding them—two-base codons would not provide enough possibilities to code for all 20 amino acids ($4^2 = 16 < 20$). Thus, the genetic code constitutes a redundant system (out of structural necessity). There are in fact 'synonyms' in this code—e.g., the

[63] Further data is given in Skyrms (2010a, 130).

[64] Note further that on this process, in the limit we would have an infinite number of signals. Twenty-five is quite a bit less than this.

[65] See also the discussion in Barrett and Zollman (2009).

codons TTT and TTC both code for the amino acid phenylalanine. Fitch (2010) points out that this redundancy means that some mutations in the DNA do not affect the particular protein for which that DNA codes. As such, these 'silent-substitutions' are invisible, in a sense, to selection. Nonetheless, their accumulation over time affects molecular phylogeny to the extent that such accumulations provide a 'random mutational record of the history of that particular chunk of DNA' (210).[66]

Regarding point (1), Hurford (2007) seems to advocate for this possibility, at any rate. I will discuss this further in Chapter 2. See also LaCroix (2019b) for more details on the adequacy of signalling-game models applied to this problem. Further, as we shall see in Chapter 2, internal representations of the meanings of signals are not necessary in the simplest case, as with, e.g., quorum signalling in bacteria.

Fundamentally, while some of the assumptions of these types of models may not be wholly (empirically) justified, this sort of criticism does not seem detrimental enough to warrant abandoning this methodology. Indeed, many interesting results can arise from modelling in this way—even if they require a bit of hedging about 'how actually' versus 'how possibly' claims. Such hedging is effectively built into any discussion of language origins. This is why multiple approaches are necessary. If we take account of empirical results, then there is no reason why a carefully crafted and well-justified model should be discounted outright.

Finally, we might note an interpretive issue concerning the content of the (simple) signals thus obtained. They appear to be doing some dual work, as both imperatives and declaratives. That is, in the signalling system shown in Figure 1.1a for example, does m_1 *mean* the declarative statement 's_1 obtains', or does it *mean* the imperative 'do a_1'? Note that this is just a version of the indeterminacy of translation (Quine, 1960). That is, the idea that an anthropologist in a remote village interviewing a native cannot jump to the conclusion that

[66]For further discussion and examples of redundancy in the genetic code, see Pearce et al. (2004); Enns et al. (2005); Kafri et al. (2006).

the exclamation *Gavagai!* means *rabbit*: it could well mean *undetached rabbit parts*, or *the rabbit I saw five minutes ago behind the stone at the top of the hill* (Mithen, 2005, 172), or any other thing relevant to the circumstance.[67] Harms (2004a,b) suggests that we treat such holophrastic signals as having 'primitive content' (in a similar sense to the *pushmi-pullyu* teleosemantics of Millikan (1995)) such that we simply leave it uninterpreted, or interpret it as a simultaneous conjunction of both the indicative and the imperative. Huttegger (2007b) shows how we might model the signals so that such a distinction can be made, and Zollman (2011) presents a different model that does not give rise to some interpretive issues that he points out are present in Huttegger (2007b). However, there are two senses in which this interpretive/translational issue should not be of any serious concern for the theory of signalling games.

On the one hand, human children (and human adults) may sometimes use holophrastic, one-word phrases in this dual imperative, indicative sense: 'there is X!'/'do something about X!'.[68] On the other hand, part of the apparent problem comes from the fact that the messages in the signalling game (our object language, in this case), and the language that we are using to reason about the signals (English—our meta-language), are two entirely different language games with altogether different expressive capacities. Even the English-language phrase 's_1 obtains' is more complicated, in some sense, than 'm_1'—whatever that may consist in. For example, 'm_1' might be the string 's_1 obtains', or it might be a guttural yelp—I take the latter to be simpler than the former.

Finally, as with all evolutionary models, there is an interpretive issue that should be flagged regarding timescale. In the learning model, interactions are discrete events. If we have a large population of individuals, and interactions are thus commonplace, thousands of individual transactions may take place in a short period of 'real' time. Young (1998) takes this to mean

[67] See also the discussion in Wittgenstein (1953, §33) on ostensive definition: 'Point to a piece of paper.—And now point to its shape—now to its colour—now to its number (that sounds queer).—How did you do it?—You will say that you "meant" a different thing each time you pointed'.

[68] For example, '*FIRE*!'.

that it does not make sense to talk about 'short-run' and 'long-run' behaviour in these model types, at least without some sort of metric for explaining to what these phrases correspond in real-time.

The issue of evolving communication or language in terms of evolutionary time also arises as an objection to gradualist assumptions more generally—in particular, those who argue for a saltationist stance in linguistics suggest that there simply would not have been enough time for adaptation to do its work in the context of communication. Progovac (2015) nicely addresses this issue, which is worth quoting at length:

> Tiny selective advantages are sufficient for evolutionary change; according to Haldane (1927), a variant that produces on average 1% more offspring than its alternative allele would increase in frequency from 0.1% to 99.9% of the population in just over 4,000 generations. This would still leave plenty of time for language to have evolved: 3.5 to 5 million years, if early Australopithecines were the first talkers, or, as an absolute minimum, several hundred thousand years (Stringer and Andrews, 1988), in the event that early *H. sapiens* were the first. Moreover, fixations of different genes can go in parallel, and sexual selection can significantly speed up any of these processes. The speed of the spread depends on how high the fitness of these individuals was relative to the competitors. According to e.g. Stone and Lurquin (2007), if relative fitness is high, it can take just a few dozen generations for the variant frequency to increase tenfold. (19)

Tobias (1987) argues that fundamental steps toward human language had been made in *Homo habilis*, 2 million years ago, though this does not indicate *full* human-level language capacities. Note that Berwick et al. (2013) suggest that the capacity for language to have evolved approximately 100,000 years ago.[69] This is still sufficient time, according to the analysis above.

[69]Fitch (2010, 257) suggests we *know* that know that language had evolved to its 'modern state' (presumably meaning with the generative capacities noted above) by the time 'anatomically modern Homo sapiens [AMHS]' left Africa, 50,000 years ago; see also Chomsky (2002). This is taken to be the 'last plausible moment at which human linguistic abilities like those of modern humans had evolved to fixation in our species' (Fitch, 2010, 273). See also Mellars (2006).

1.3 Summary

In this chapter, I have examined the question *how ought we to study language and its origins?*—at least, in a philosophical sense. The philosophy of language, historically, has gone about studying language—i.e., in terms of concepts like intentionality, truth, reference, etc. while ignoring the fundamental *purpose* of language: to communicate with others. Thus, taking a quasi-Wittgensteinian bent—focusing on the use of language—we have seen how the signalling game provides a coherent way of understanding language as conventional interactions in a community of language-users in particular contexts that warrant such use.

In this discussion, I have introduced a good deal of technical terminology which will be of significant use throughout the rest of this dissertation. Further, I have outlined several extant results from various extensions of the simplest version of the signalling game—both in an evolutionary context and a learning context (what has been called *cultural evolution*). Finally, I suggested that the theory of signalling games provides many clear and coherent insights clearly and coherently that are exceptionally fruitful for the discipline as a whole. I ended by noting some caveats about modelling assumptions, and why and in what ways these assumptions may or may not be justified.

In the next chapter, I move beyond the simple signalling-game framework and start to ask questions about how more complex signalling and more complex linguistic phenomena might arise. Also, we will see the conditions under which we might expect signalling to appear in nature and the conditions that might be required for more complex signalling behaviour to arise, by examining animal communicative and cognitive mechanisms in some detail.

Chapter 2

Communication and Language

If a lion could speak, we could not understand him.
— Wittgenstein, *Philosophical Investigations*

A tiger's anatomy should allow it to produce the point vowels /i/, /a/, and /u/.
— Fitch, *Evolution of Language*

In Chapter 1, I introduced the signalling-game framework. The simple communication systems that are well modelled by the signalling game are disparate in obvious ways from full-fledged *languages*. In this chapter, we will begin by examining in detail the salient differences between communication and language (Section 2.1). The gradualist perspective on language origins posits an intermediary stage between simple communication and language, called a *protolanguage* (see Section 2.1.1). To determine the appropriate 'targets' for an evolutionary explanation—i.e., what it is that evolved—we can survey the distinctions between communication and language. The most famous such account is presented in Section 2.1.2. Some alternative distinctions are offered in Section 2.1.3; however, we will see that what is

common to all of these is a notion of *compositionality* in the form of productivity, openness, generative capacity, or hierarchical structure.

Indeed, several models of signalling games have been built to try to account for the emergence of compositional signalling—what we might call *proto-compositionality*. These are surveyed in Section 2.2. In section 2.3, I examine the converse question of what animal communication systems are capable. This analysis includes quorum signalling (2.3.1), functionally referential alarm calls (2.3.2), openness in terms of continuous expressivity (2.3.3), combinatorial signalling (2.3.4), and compositional signalling (2.3.5). What we shall see is that there is scant evidence that compositionality exists in nature; further, where it does appear to exist, I argue that it cannot possibly serve as a precursor to compositionality in human languages.

The salient differences between language and communication provide the possible explanatory targets for an evolutionary account—namely, once we know what the differences are, we can suggest what would have had to evolve to move from a system of communication to a system of language. By examining empirical data, we can restrict the possible explanatory targets to a smaller set of *plausible* explanatory targets. The purpose of this chapter is to give a first argument that compositionality is not the correct target for an evolutionary explanation of how language evolved. In Section 2.4, I discuss some other considerations of which an evolutionary account must maintain sensitivity, including biological, cognitive, and social constraints.

The main contribution of this chapter is, therefore, negative. However, I offer an alternative account (the positive contribution of this part of the dissertation) in Chapter 3.

2.1 Communicative versus Linguistic Capacities

Given that signalling is ubiquitous in nature, but languages are often taken to be unique to humans, the question naturally arises: *What is the difference between simple signalling behaviour (or simple/complex communication systems) and language?*

Since this dissertation is couched in the signalling-game framework, and evolutionary game-theoretic explanation in general, I take for granted that the gradualist approach to language origins is the correct one. Given this fact, I also take for granted that the *primary* purpose of language is to communicate—a necessary condition of the gradualist approach is that language and communication differ only in degree, not kind.[1]

One might object to this insofar as language and communication are different *kinds* of things: the latter is something that individuals *do*, whereas the former is a tool that individuals *use*—e.g., to communicate, to express thought, etc. Even so, if the primary *use* of this linguistic tool is to communicate, then communication and language are best understood as being only different in degree.

By way of analogy, fishing is a thing that one does, whereas netting is a tool that one uses—but this tool makes the action of fishing more efficient or more effective. Similarly, hunting (e.g., hare) is a thing that one does, but cooperation is a tool that one can utilise to hunt (e.g., stag) more effectively. I believe this analogy is helpful in the sense that, e.g., 'netting' is not a *type* of fishing, nor is 'cooperation' a type of hunting; nonetheless, *fishing-using-netting* and *hunting-in-groups* are types of things of the relevant category. Similarly, *communicating-via-language* is a genuine type of communication, though language itself is perhaps best conceptualised as a tool. There is no category error here.

[1] See also the discussion in Lewis (1975).

This need not be taken for granted. As was mentioned in the Introduction, authors like, e.g., Chomsky (1980b); Bickerton (1990); Wray (1998) suggest that the *primary* purpose of language is/was not communication, *per se*, but the expression of *thought*. However, even here, it is informative to examine what it is that animals *can* do and what they cannot, to proceed with a comparative analysis. It is (at the very least) *theoretically* fruitful to understand these modes as differing merely in degree. Thus, let us suppose that language and communication lay on a spectrum with (as far as we are concerned) human languages at one extreme and the simplest form of communication via atomic signalling at the other extreme.

2.1.1 On Protolanguages

Recall that the gradualist picture of language origins posits that there must have existed, at some point in evolutionary history, a *protolanguage* (sometimes called 'pre-language'). This serves as a theoretical bridge between modern human-level linguistic capacities and the communication systems that would have been available to non-linguistic hominin ancestors. As such, a conceptual clarification of protolanguage, and in what protolanguage inheres, is crucial in the study of language evolution.

The distinction between language and animal communication has long been noted. For example,

> What is it that man can do, and of which we find no signs, no rudiments, in the whole brute [i.e., animal] world? I answer without hesitation: the one great barrier between the brute and man is Language. Man speaks, and no brute has ever uttered a word. Language is our Rubicon, and no brute will dare to cross it. (Müller, 1864, 367).[2]

[2]Indeed, this type of argument, from human capacities for language, was seen in Darwin's own time as a weakness of his theory.

Protolanguage is a theorised intermediary between these two apparently disparate systems. The notion of a protolanguage, used in the context of a biological stage of human evolution, was introduced by Hewes (1973). However, the concept of such an intermediary was already present in early theories—for example, the onomatopoetic or the expressivist (or interjectionist) theories of word-origins (Herder, 1772).[3] Today, there are several perspectives that one might take on the constitution of a protolanguage; these views can be (approximately) divided into three camps, which include theories of *lexical-*, *gestural-*, and *musical* protolanguages.[4]

Theories of lexical protolanguage are perhaps the most common and the most intuitive of the three classes.[5] Such theories begin with words or lexical items that are unconnected by (simple or complex) syntax. The primary question, then, is how syntax evolves and becomes complex. This model is adopted implicitly by several researchers, though it is most explicit in Bickerton (1990); Jackendoff (2002). Note that this allows for many different mechanisms or capacities to be included in one's actual theory. Bickerton (1990) argues that the evolution of syntax was 'catastrophic', whereas Jackendoff (2002) suggests it was incremental. In both of these cases, they assume that it is beneficial for individuals to share information. Most such accounts in linguistics and anthropology assume cooperation as an underlying aspect of human behaviour.

Fitch (2010) highlights that theories that posit a lexical protolanguage take a lot for granted—with the main issue being voluntary control of vocal expression. One alternative to lexical protolanguage is a theory of gestural protolanguage, which posits that the primary com-

[3]These are (in)famously dismissed as the 'bow-wow' and 'pooh-pooh' theories by Müller (1864). In the 1864 publication of the lectures (first delivered in 1861), Müller (1864) initially apologises in a footnote for these dismissive titles: 'I regret to find that the expressions here used have given offence to several of my reviewers. They were used because the names Onomatopoetic and Interjectional are awkward and not very clear. They were not intended to be disrespectful to those who hold the one or the other theory' (372). However, 10 years later, Müller (1873) argues that he felt *certain* that 'if this theory were only called by its right name, it would require no further refutation' (189).

[4]For a detailed chapter-length overview of each of these positions and their critics, see Fitch (2010, Ch. 12–14).

[5]See, for example, Lieberman (1984, 2000); Bickerton (1990); Givón (1995); Jackendoff (2002).

municative modality that predates vocal/auditory speech was not itself vocal/auditory, but rather manual/visual.[6] Such theories have the benefit that protolinguistic 'fossils' are readily apparent in modern language, in terms of hand-gestures (while speaking), pantomime, and signed languages. For example, humans often make gestures while speaking, even when their interlocutor is not able to see them. Accordingly, an explanation of the evolution from (gestural) protolanguage to (primarily vocal) natural language requires an account of the shift between these modalities. In fact, a theory of gestural protolanguage needs to explain two things: how the modality of communication shifted *and* how complex syntax arose. Furthermore, did complex syntax arise in terms of gestural communication, and then the mode switched to oral, or did the mode switch to oral communication and later became complex? (Note that the latter of these seems to collapse into a lexical theory of protolanguage.)

Darwin (1871) offers an incredibly dense 10-page treatment of language origins (53-62). He takes a multi-component approach to language, which recognises the necessity of several distinct mechanisms (rather than a single 'key' feature of language that needed to evolve). In particular, Darwin (1871) notes the innateness of the human language *faculty*, in conjunction with the necessity of *vocal learning* for human language, and draws an analogy between human language and bird song: 'all the members of the same species utter the same instinctive cries expressive of their emotions; and all the kinds that have the power of singing exert this power instinctively; but the actual song, and even the call-notes, are learnt from their parents or foster-parents' (55). On this 'musical protolanguage' (Fitch, 2006), there is a general increase in sophisticated cognitive capacities, followed by sexually selected attainment of the specific capacity for complex vocal control which gives rise to song. In the final stage of language evolution, these complex songs became endowed with meaning, which both affected and was affected by, further increases in cognitive capacities. On this musical protolanguage account, song-like protolanguages already include complex phonology

[6]See, for example, Mandeville (1723); de Condillac (1747); Hewes (1973); Rizzolatti and Arbib (1998); Corballis (2002); Arbib (2005); Call and Tomasello (2007); Tomasello and Call (2007).

and even complex syntax, but such a protolanguage lacks *semantic meaning*.[7] In such a case, the main question is how propositional meaning might have arisen out of song.

Note that, because the signalling-game model is abstract, we can remain agnostic about whether a signal denotes a sound, as in a lexical item; a string of sounds, as in a song; or some non-verbal cue, as in a gesture. As such, a 'signal' in the signalling game might be lexical and take advantage of the vocal/auditory channel, or it might be gestural and take advantage of the manual/visual channel, or it might be a string of sounds that is interpreted holophrastically—i.e., as an atomic whole. Thus, at this stage, the signalling-game framework is general and abstract enough to maintain theoretical neutrality about protolanguage. The critical question, for now, is what elements of language are unique to linguistic systems. This provides the explanatory target for an account of protolanguage.

2.1.2 Design Features

The question of the salient differences between language and communication is by no means new. Herder (1772)—a contemporary of Goethe and Kant—highlights the distinction clearly:

> I cannot conceal my astonishment at the fact that philosophers . . . can have arrived at the idea that the origins of human language is to be found in . . . emotional cries. All animals, even fish, express their feelings by sounds; but not even the most highly developed animals have so much as the beginning of true human speech . . . Children produce emotional sounds like animals; but is the language they learn from human beings not an entirely different language? (24)

Thus, it has long been accepted that there is a salient distinction between language and simple (animal) communication systems, but wherein might this difference lie? Perhaps the most influential attempt to distinguish between animal communication and linguistic

[7] See, for example, Darwin (1871); Jespersen (1922); Livingstone (1973); Richman (1993); Brown (2000); Merker (2000); Mithen (2005); Fitch (2010).

communication was proposed and developed by Charles F. Hockett in the 1950s and 1960s (Hockett, 1958, 1959, 1960a,b, 1963; Altmann, 1962, 1967; Hockett and Altmann, 1968).

Hockett (1958, 1959) initially proposed a list of 7 'key properties of language': *duality*, *productivity*, *arbitrariness*, *interchangeability*, *specialisation*, *displacement*, and *cultural transmission*.

1. *Duality of Patterning.* Many meaningful signals (e.g., words or morphemes) are produced from meaningless sounds (e.g., phonemes or features).[8] The function of these meaningless units is to distinguish the meaningful units from one another.

2. *Productivity/Openness.* A potentially infinite number of different meaningful messages can be produced by combining the elements of the language in different ways. Almost all animal communication systems have a small, finite number of possible messages; thus, they constitute closed systems. Human languages are potentially infinite, and so are open systems of communication.

3. *Arbitrariness.* The relation between a signal and its meaning is arbitrary. Many signals in language (outside of onomatopoeia) are not iconic, as opposed to many animal signals; e.g., a dog baring its teeth to bite.

4. *Interchangeability.* Utterances that are understood (e.g., by a receiver) can be thus produced—i.e., the role of sender and receiver is interchangeable. Competent speakers of a language can act as both sender and receiver, as opposed to, e.g., birdsong, which is typically produced by the male only, or the waggle dance of *worker* honeybees, which is understood by queens and drones but is not reproduced by them.

5. *Specialisation.* Signals thus produced are specialised for communication, and not as a byproduct of another behaviour. The distended belly of a female stickleback fish is a

[8] See, e.g., Hockett (1960b, 6–8) for details. See also the discussion in Ladd (2012).

signal that she is ready to breed; however, this is a byproduct of the development of roe—communication in language is not a byproduct in this way.[9]

6. *Displacement.* Language allows one to communicate about things that are not immediate in time or space. Many animal communication systems can only be used to 'refer to' things *here* and *now*. Hockett (1958) defines displacement in terms of antecedents and consequences: '[a] message is displaced to the extent that the key features in its antecedents and consequences are removed from the time and place of transmission' (579).

7. *Traditional/Cultural transmission.* Systems of language are culturally transmitted between generations. Human language is not innate but learned.[10] Many animal communication systems at least appear to be hardwired—for example, Winter et al. (1973) show that young squirrel monkeys raised by muted mothers will produce the full range of the calls of their species in the appropriate contexts.

The presentation in Hockett (1958, 1959) is couched in Shannon's (1948) model of communication as a transmission of information from a sender to a receiver, but this is consistent with many accounts of animal communication in the biological literature—including sociobiology (Wilson Jr., 1975), ethology (Hailman, 1977), and behavioural ecology (Krebs and Dawkins, 1984), for example.

This list of 7 'key properties' of language was extended to a list of 13 'design features' of language, which additionally included the *vocal-auditory channel*, *broadcast transmission and directional reception*, *rapid fading*, *total feedback*, *semanticity*, and *discreteness*.

[9]In fact, this is not a signal at all, but rather a 'cue'; I will ignore this distinction for now, though see Section 2.3.

[10]Note, the *capacity* for language seems to be innate, but the particular language that is spoken by any particular person is learned.

8. *The vocal-auditory channel.* The auditory system perceives sounds emitted from the mouth. That is, signal modality involves the production and perception of sound.[11] This is supposed to distinguish (spoken) language from, e.g., chemical or electrical signals.[12]

9. *Broadcast transmission and directional reception.* In general, signals travel to any potential receiver (i.e., within earshot), and the acoustic properties of these signals can help to determine the originating source. This distinguishes vocal signals from, e.g., olfactory signals, as when an animal marks its territory.

10. *Rapid fading (transitoriness).* Signals last a short time, as sounds fade quickly. Disregarding modern recording means, the transitory nature of a vocal signal is different from, e.g., a chemical signal, which may persist for some time.

11. *Total feedback.* A sender also perceives the message she sends. Auditory feedback of one's own (vocal) signal is frequent in human language, but mating signals of, e.g., stickleback fish consists in a change in belly and eye colour—neither of which can be perceived by the sender of the signal.

12. *Semanticity.* There is a fixed relationship between a signal and its meaning. Signals in language are associative, and some may be referential or have a denotation.

13. *Discreteness.* Language is built up from discrete units, such that altering one of the units of a linguistic item can change the meaning of that item. The discrete parts of language can be recombined to create new meanings. In general, animal communication systems are either continuous (as in the waggle dance of honeybees), or they are not semantically recombinable (as in birdsongs).

[11]Note that this feature excludes, e.g., signed languages as languages, since, at the time Hockett was writing, it was not yet widely acknowledged that signed languages *were* comparable to spoken languages in many relevant respects. However, it is now accepted that signed languages are fully complex grammatical languages on a par with spoken languages (Stokoe, 1960; Klima and Bellugi, 1979). So, *spoken* language is not the only system adequate for language.

[12]This also, as it happens, differentiates spoken language from written language; however, Hockett (1960b) understands written language as derivative of spoken language.

This list of 13 is the standard presentation of Hockett's design features of language (Crystal, 1987; Hauser, 1996). These additional features are present in 'all' human languages—though some are apparently 'trivial', individual animal communication systems may lack them.

Wacewicz and Żywiczyński (2015) highlight that the additional six design features can be derived from the original seven key properties. On the one hand, Hockett (1958, 1959) already insists that the vocal-auditory channel is the prototypical manifestation of linguistic behaviour; but such a channel then includes characteristics such as broadcast transmission, rapid fading, and total feedback. On the other hand, semanticity and discreteness are logical consequences of duality.

Later, Hockett (1963); Hockett and Altmann (1968) shift their focus from comparative concerns—i.e., between human and animal 'languages'—to the particular properties of human languages. They add three additional features for a final list of 16; these include *prevarication*, *reflexiveness* and *learnability*.

14. *Prevarication.* Signals can be false, deceptive, or meaningless. Many systems of communication are incapable of deception (e.g., quorum signalling in bacteria).

15. *Reflexiveness.* Languages can be used to communicate about languages. There are certain limitations on the *types* of things about which animals can 'talk'.

16. *Learnability.* A speaker of one language can learn another language. Related to innateness, birds who sing cannot, for example, learn the songs of another (though similar) species.

Though these are supposed to be specific to natural languages, Hockett notes that prevarication relies upon semanticity, to the extent that messages must first be meaningful for them to be false; displacement, to the extent that a *successfully* false message needs to refer to

something outside of 'here' and 'now'; and productivity, which would guarantee the ability to produce false messages in the first place.[13]

Whatever individual components end up being relevant to language, the key thing to note is that all are taken to be necessary, and none alone will be sufficient: while many of these features are contained within animal systems of communication, Hockett believed that human languages were the only forms of communication that satisfied *every* item on this list. However, there are obvious problems that arise once we have defined such a list; For example, one might point out that the vocal-auditory channel criterion excludes signed languages, and one might further point out that many animal signals utilise the vocal-auditory channel. Thus, we have a feature that is absent in certain *bona fide* human languages, but which is present in some animal communication systems.[14] I will not delve into a detailed of criticism of Hockett's criteria since there are many such efforts in the literature.[15] In any case, a *key* feature of language, on Hockett's account, which is not derivative of other features, is productivity/openness, which is accounted for by compositionality.

[13]Wacewicz and Żywiczyński (2015) note these points specifically; however, it is not clear that prevarication actually requires all three of these features. False or deceptive signals certainly rely upon semanticity to the extent that falsehood is parasitic on truth (i.e., a signal needs to have an established conventional meaning in order for it to be used in a false way). However, the necessity of displacement for prevarication seems also to be derivative of pre-established conventional meaning; for example, since a receiver of a signal can have incomplete information about the state of the world, if a vervet monkey signals that a leopard is present when there is no leopard present, then it is because of the pre-established meaning (which itself may require displacement) that the receiver acts. Similarly, it is not clear why productivity is necessary to the extent that a deceptive signal requires only a pre-established signal and not a novel signal per se. For more on prevarication in a signalling context, see Skyrms and Barrett (2018).

[14]Note that Hockett (1960b) argues that the *particular* acoustic features of human languages—e.g., vowel colouring—really are unique to humans.

[15]See, for example, Bradshaw (1993), Anderson (2004, Ch. 2), Everett (2005), Wacewicz and Żywiczyński (2015). The list itself has been modified several times: Altmann (1967), Hockett and Altmann (1968), Ristau and Robbins (1982), etc. Similar criteria have been alternately proposed by, e.g., Brown (1973), Limber (1977), Chomsky (1979).

2.1.3 Alternative Distinctions

Wacewicz and Żywiczyński (2015) suggest that Hockett's design features are mostly incompatible with modern research in the evolution of language since they focus on the features of language and communication systems rather than the language faculty. Even so, it is still useful in finding possible explanatory targets. In this section, I survey some other candidates for uniquely-human features of language. We will see that they all have one salient thing in common: compositionality.

Hurford (2012) alternatively posits the following 'universally learnable' features of human language: *storage capacities*, *hierarchical structure*, *word-internal structure*, *syntactic categories*, *long-range (syntactic) dependencies*, and *constructions*. Not all of these features will be used, or used extensively, by all individuals in every language; nor are these necessarily human-specific, in the sense that some derivative or simplified version might appear in animal communication, but the uniqueness to humans is that *all* of these items are in principle learnable by any human.[16] Instead, they are in the 'centre of the distribution of features that languages have' (Hurford, 2012, 373)—i.e., there may be some statistical outliers.

Regarding storage capacity, Goulden et al. (1990) estimate that 'well-educated adult native speakers of English have a vocabulary of around $17,000$ base words' (321). In contrast, Diller (1978) calculated that high school teenagers knew, on average, approximately $216,000$ words. Part of the discrepancy here arises from assumptions made about what constitutes a unique word and what constitutes knowledge—i.e., whether we measure active or passive vocabulary (Cooper, 1997). In either case, this significantly outpaces (by several orders of magnitude) any known lexicon in animal communication systems.

[16]The fact that these are taken 'in principle' is evident from the following remark: 'any normal human child, born no matter where and to whichever parents, can acquire any human language, spoken anywhere in the world. Adopt a baby from deepest Papua New Guinea, and bring it up in a loving family in Glasgow, and it will grow up speaking fluent Glaswegian English' (Hurford, 2012, 260-1).

Hierarchical structure has not only to do with the *meronomic* nature of natural languages (concerning part-whole relations)[17] but also the *dependency relations* between parts within an expression. Speakers put sentences together in hierarchical ways, and listeners also deconstruct sentences in hierarchical ways.[18] Hierarchical structure is closely related to compositionality insofar as many languages present hierarchical structure both in terms of sentence composition and in terms of morphological composition—i.e., word-internal hierarchical structure. Hurford (2012) points out that though English is relatively impoverished with respect to morphology, some languages—agglutinating languages—have a vibrant internal morphological structure to their words; e.g., Inuit dialects can often involve a dozen morphemes bound together to comprise a single-word sentence.[19]

Long-range dependencies might include, e.g., grammatical agreement between subjects and verbs. The following example, from Hurford (2012), is grammatical, readily intelligible, and may well occur in a typical English conversation:

> '**The shop** that sold us the sheets that we picked up from the laundry yesterday morning **is** closed now'.

[17]This is not to be confused with the concept of *meronymy* in lexical semantics. A *meronom* is an object which is a part of a whole; whereas, a *meronym* is the *name* of a part. Meronymy is thus a relationship between words. Meronomy is a relationship between parts and sub-parts, which is compared with a *taxonomy*, whose categorisation is based on discrete sets.

[18]On the hierarchical nature of sentence-assembly in real time in the brain, see, for example, Garrett (1975, 1982), Levelt (1989, 1992), Dell et al. (1997), Smith and Wheeldon (1999), Wagers and Phillips (2009).

[19]For example,

(1) *ayagciqnillruyugnarquq*

means 'He probably said he would go', which can be compared with

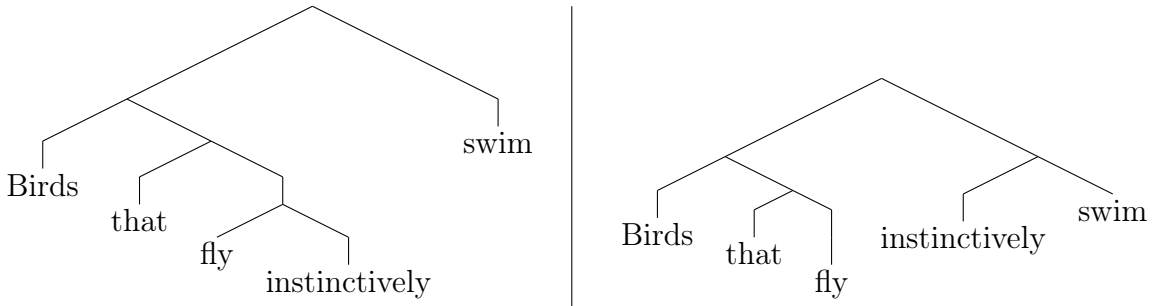
(2) *ayagciqsugnarqnillruuq*

meaning 'He said he would probably go'. Note the structural difference between ayagciqnillru**yugnarq**uq and ayagciq**sugnarq**nillruuq; this shows that 'it is the position of morphemes within words in Inuit, not of words themselves, that corresponds to syntactic positioning in a language like English' (Compton and Pittman, 2010).

Here, the grammatical dependency concerns an agreement between the subject—*the shop*—and the copula—*to be*—which occur syntactically far from one another. Linear dependencies of this sort are closely related to hierarchical structure and can be partly explained by such structures; see Figure 2.3. Berwick and Chomsky (2016); Chomsky (2017) highlight that the following sentence,

'Birds that fly instinctively swim',

is ambiguous insofar as the adverb could be understood to modify either of the verbs; see Figure 2.1.



(a) Adverb modifies *to fly* (b) Adverb modifies *to swim*

Figure 2.1: Ambiguous syntax trees

However, the sentence

'Instinctively, birds that fly swim'

is unambiguous. What is perhaps initially puzzling is that, in this case, the adverb modifies the verb 'to swim', which is further away from the adverb than 'to fly', with respect to linear order. However, the adverb is closer to the verb 'to swim' with respect to *structural* (i.e., hierarchical) order—'swim' is embedded one level down from 'instinctively' whereas 'fly' is embedded two levels down. See Figure 2.2

At any rate, Hurford (2012) points out that

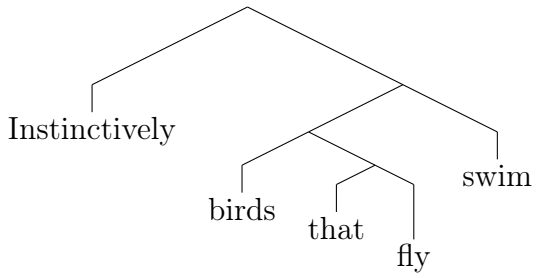


Figure 2.2: Unambiguous syntax tree

> humans are able to learn systems demanding a certain degree of online memory during processing of a sentence. Putting it informally, when a word is heard, the processor is able to store certain grammatical properties of that word in a temporary 'buffer' and wait until another word comes along later in the sentence with properties marking it as likely to fit semantically with the stored word. (342)

Finally, constructions are taken to be linguistic constructions 'of any size and abstractness, from a single word to some grammatical aspect of a sentence, such as its Subject-Predicate structure' (Hurford, 2012, 348). A speaker's knowledge of her language consists of a large inventory of such constructions. This ability arises, in part, from the use of *variables* in a language. Thus, this too is directly related to the notion of (functional) composition.[20]

2.1.4 The Preeminence of Compositionality

The focus on recursion or composition as *the* defining feature of human languages is pervasive, if often implicit. By way of example:

> Apes, but also dogs, have 'lexicons' that can attain a few dozen words (Premack, 1971, 1986). However, such abilities are insufficient to enable non-human animals to construct a grammar comparable to that of humans. (Mehler et al., 2006, 254)

[20]See Fried and Östman (2004).

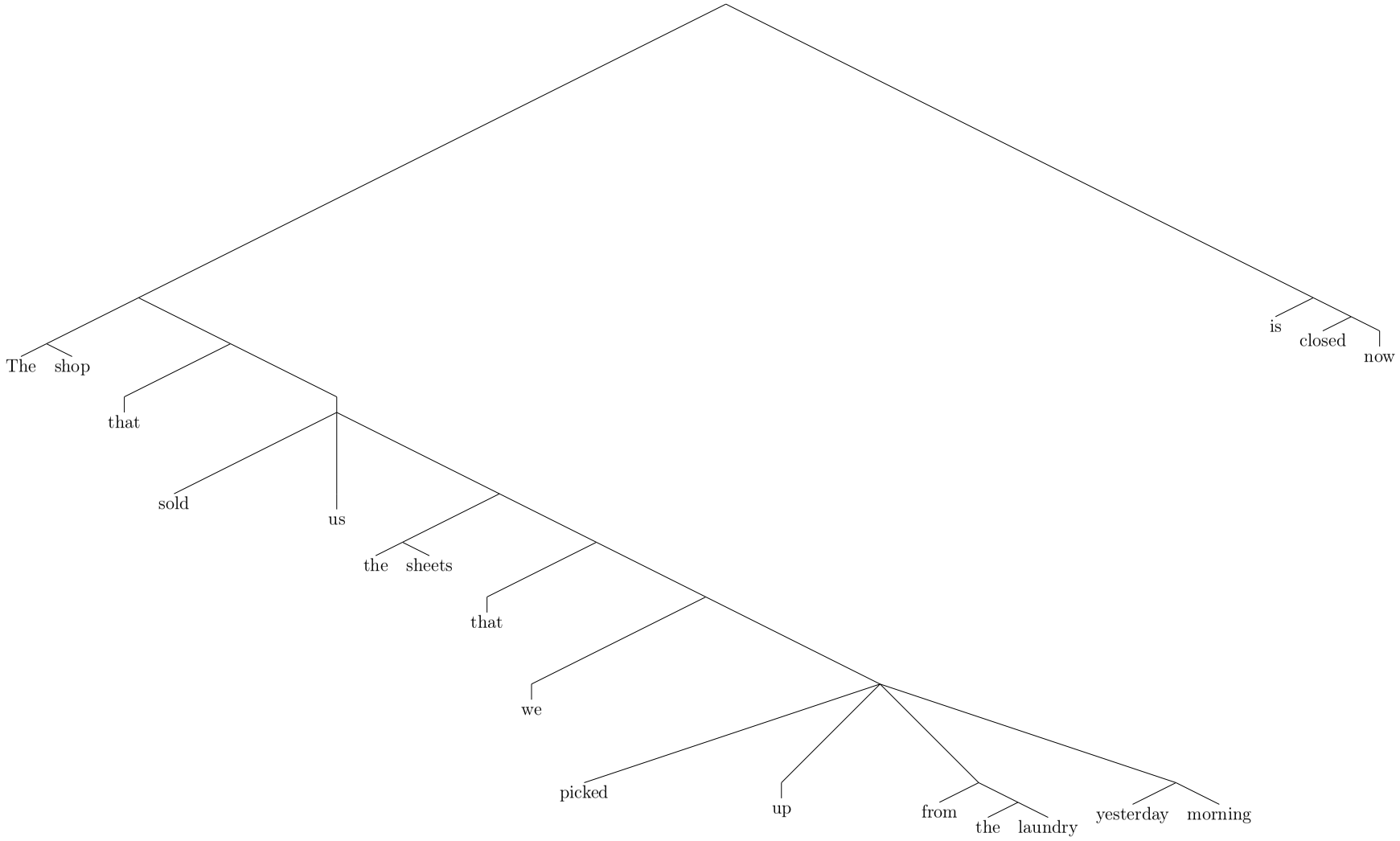


Figure 2.3: Long-range dependencies can be explained by hierarchical structure

> Despite intensive searching, it appears that no communication system of equivalent power [to human language] exists elsewhere in the animal kingdom. (Fitch, 2010, 1)

> Our best current evidence suggests that no other living species has a communication system that allows it to do what we humans do all the time: to represent and communicate arbitrary novel thoughts, at any desired level of detail. Indeed, our current data suggest that even a rudimentary version of this ability (to communicate some novel thoughts) is lacking in other species. (Fitch, 2010, 26)

> No non-human has any semantically compositional syntax, where the form of the syntactic combination determines how the meanings of the parts combine to make the meaning of the whole. (Hurford, 2012, 96)

> [Combinatorial communication] is rare in nature, and where it does exist it is, with one salient exception, simple and limited. . . . [The] one extreme exception to the norm of non-combinatorial communication: human linguistic communication. (Scott-Phillips and Blythe, 2013, 1, 5)

Hauser and Fitch (2003) argue that syntactic recursion is the only *uniquely* human language-trait. Similarly, Hauser et al. (2002) say that recursion is found solely in human language. They also claim that the defining feature of human language is recursion.[21] Compositionality is often taken to be one of (if not *the*) cornerstone(s) of productivity in language. Compositionality contributes significantly to openness, flexibility, and learnability, which are taken to be characteristics unique to languages, as we have seen. Thus, it stands to reason that the most important target phenomenon for explaining the evolution of *language* (out of simple communication systems) will involve an explanation of how compositionality might arise.

Many researchers do not argue for their emphasis on syntax, but instead, assume that this is the correct explanatory target of an evolutionary account. For example, Berwick and

[21] However, Hurford (2012) points out that a system of two constructions with a combinatorial rule that is limited only by working memory (and thus recursive in their sense) would not constitute a human language; rather, the 'combinatorial ability [of human language] is as impressive as it is because we have massive stores of constructions to combine' (537).

Chomsky (2016) suggest the following three *key* properties of the syntactic structure of human language: (1) syntax is hierarchical, (2) the particular hierarchical structures associated with sentences affect their interpretation, and (3) there is no upper bound on the depth of relevant hierarchical structure. Though hierarchical structure is a syntactic property of language, this notion depends implicitly and inherently upon compositional semantics—this is prototypically demonstrated by Chomsky's (1957) example of colourless green ideas, which sleep furiously.

Compositionality is a crucial feature of explanations that depend upon the signalling-game framework as well. Lewis (1969), in his discussion of signalling conventions, attempted to show how such conventions might be understood as a rudimentary language, $\mathcal{L}$, consisting of possible moods and truth conditions. Lewis calls the signalling language rudimentary because it lacks the following features that, he believes, we should expect from a fully realised language:

1. Compositionality. There only exists a closed finite set of sentences—the domain of $\mathcal{L}$. It is not possible to create a new sentence, along with its truth condition, out of old parts.

2. There is no idle conversation. The sentences of $\mathcal{L}$ are used only for a particular activity.

3. There is little choice on the part of the actors. A sender observing a certain state of affairs who wishes to tell the truth must send a particular signal: 'He has no choice whether to speak or be silent; no choice what to talk about; no choice even how to phrase his message' (160).

4. Signals (in the indicative case) represent facts about the world; they cannot express beliefs or hypotheses.

5. Indicative sentences of $\mathcal{L}$ can only express facts about the occasion of utterance of the sentence.

6. The language $\mathcal{L}$ contains only two moods: indicative and imperative. A full-blown language should at least be able to account for interrogative, commissive and permissive moods. (Lewis notes that some of these are reducible to indicatives of imperatives, but that these reductions are problematic.)

There are several other ways in which the language, $\mathcal{L}$, might be impoverished. However, this list is significant enough to specify some of the main differences between language and communication. It is of interest that Lewis himself prioritised compositionality as a key feature of language that is absent in signalling. Thus, we might take compositional signalling as our explanatory target to bridge the gap between communication and language. Compositionality takes centre stage in most arguments of language origins both within and without the signalling-game framework. Indeed, several signalling-game models for compositional communication have been proposed in recent years. The subsequent section outlines these models; however, as we shall see, the inverse question of what animals are capable of doing is often ignored or presupposed in these discussions.

2.2 Complex Signals: Models of Compositionality

We have seen, in Chapter 1, how simple signalling might arise under a variety of circumstances. The signalling-game framework gives a robust set of models for examining conditions under which we should expect simple communication to appear in nature. However, simple signalling of this sort is a far cry from the complex structures present in human language. We have seen in the previous section that one of the most salient features in discussions of the differences between simple communication and language is some sort of generative capacity in general—in particular, a principle of compositionality. It stands to reason that we should be able to modify the signalling-game framework to account for some (simple) notion of (proto-)compositionality. It is commonly assumed in the signalling literature that

an explanation of compositional signalling behaviour will be necessary for explaining the generative complexity present in human languages. Thus, an account of how compositional signalling might evolve is going to be significant. Several attempts at showing how compositionality might arise in a signalling context have been offered recently. These models vary in their construction; here, I give a brief overview of each of these.

2.2.1 Signal-Object Associations

Nowak and Krakauer (1999) examine how compositionality might emerge by way of natural selection on signal-object associations. Signals are interpreted as unique sounds. Each individual in the population communicates with every other individual, and rewards are summed. The rewards are interpreted as the fitness of strategy so that a higher payoff implies higher fitness. After 20 rounds of play (for a population of 100 individuals), a suboptimal, but evolutionarily stable, state, where each sound is associated with one object, emerges.

Taking account of the fact that early signals were likely noisy, Nowak and Krakauer (1999) introduce the possibility for error in perception by having possible available sounds range on a linear spectrum between 0 and 1. They highlight the fact that the number of objects that can be differentiated by unique sounds is inherently limited by the sounds available. However, though adding new sounds increases the number of objects that can be represented, this comes at the cost of accuracy, since the probability of making mistakes is also increased—they call this the 'linguistic error limit'. Thus, the ability to transfer information does not improve.

Being able to form sounds into words is a way of overcoming the linguistic error limit, in the sense that the combinatorial capacities of words with sound-length l allows individuals to create unique signals while avoiding the noise of a spectrum of single sounds. In particular, a

listener need not have an entire lexical list available (for all the possible words in the lexicon) but only needs to be able to identify the individual phonemes that comprise a word correctly. In the case of 100 unique objects, their population obtains maximum payoff when unique sounds denote the 28 'most valuable' objects, and the rest are ignored. In the case of simple word formation, with words of length $m = 2$ to $m = 10$, the maximal payoff is achieved for a word length of $m = 7$, which gives rise to 49 objects described by 49 words.

For an analysis of compositionality, the most compelling case is in their third model, where grammar emerges as a way of combining words into phrases to convey more information. They describe the conditions under which grammar is going to be more fit than non-grammar for signalling and show that 'a grammatical system is favored only if the number of relevant sentences... exceeds the number of words that make up these sentences' (8031). Grammar is thus understood as a 'simplified rule system that reduces the chances of mistakes in implementation and comprehension' (8031). It is in this sense that it will be favoured by natural selection in a world where mistakes are possible. To show how such a system might evolve gradually via natural selection, they consider a state space consisting of pairs of objects, each with two properties, giving rise to four possible combinations. Their strategy space is constituted by the probability p that players use atomic words and the probability $1 - p$ that players use grammatical constructions. They show that $p = 0$ and $p = 1$ are the only evolutionarily stable strategies. Further, their evolutionary dynamic evolves to use the grammatical rule with probability 1.

However, it is essential to note that in their discussion of the emergence of compositional language, Nowak and Krakauer (1999) only analyse whether atomic versus compositional signalling makes it easier to arrive at a signalling system. Furthermore, they only analyse subject-predicate structures rather than genuinely *functionally* compositional signals. Thus, their model cannot explain the emergence of logical function words, such as 'not' or 'of'.[22]

[22]See also the discussion in Steinert-Threlkeld (2016).

2.2.2 Syntactic Signalling

Barrett (2006, 2007, 2009) considers a signalling game where there are two senders, each of which can send one of two possible messages, and there are four state-act pairs. In this case, there is an informational bottleneck in the sense that no signal alone can adequately partition nature; however, the two senders together can completely partition nature. Skyrms (2010a) reinterprets this situation as a signalling game in which one sender sends two signals in a particular order, giving rise to a *syntactic* signalling game. (Mathematically, these two models are equivalent.) The sender and receiver communicate perfectly when they learn a bijective mapping between state-act pairs and sequences of signals. Note that the sequence length, in this case, is fixed. The receiver then needs to interpret the correct action as being given by the intersection of the two signals.

Again, no signals are functionally compositional in this case. Franke (2014) points out that the syntactic signal is more parsimoniously interpreted as a single atomic signal: '[a]lthough we can describe the situation as one where the meaning of a complex signal is a function of its parts, there is no justification for doing so. A simpler description is that the receiver has simply learned to respond to four signals in the right way' (84). So, the receiver treats the 'complex' signal as though it were atomic. However, Barrett (2006) himself does not use the phrase 'compositionality' to describe what is going on in the syntactic signalling game that he presents. Instead, the syntactic signalling game in this context could not be compositional because, as Barrett (2006) notes, '[e]ach [atomic] signal is *independent* in the sense that neither sender knows what the other sent' (229).

Similarly, Barrett (2006) explicitly states that 'the two signals together may be considered to be a single length-two message'. However, the order of the signals matters for the complex signal to transmit the right information, so this is a case of, what we might call, syntactic composition without semantic composition. This is similar to how birds use syntactically

well-ordered combinations of notes to signal, e.g., sexual attractiveness or threat.[23] However, the individual notes do not mean anything—thus, this is a case, in nature, of phonological syntax without compositional semantics (Hurford, 2012).

2.2.3 Spill-Over Reinforcement and Lateral Inhibition

Franke (2014, 2016) uses *spill-over reinforcement* (similar to a mechanism in O'Connor (2014)) and *lateral inhibition* (similar to a mechanism in Steels (1995)) in a model of simple reinforcement learning to try to give an account of, what he calls, *creative* compositionality. Here, we find complex signals of the form m_{AB}. Formally, these are new atomic signals. However, they bear a similarity to basic atomic signals by the distance $s = d(s_{AB}, s_A) = d(s_{AB}, s_B)$. On Franke's account, 'spill-over' affects the reinforcement of non-actualised state/message pairs proportional to their similarity to the actualised state/message pair (and *mutatis mutandis* for the receiver and message/act pairs). 'Lateral inhibition' lowers the accumulated rewards for non-actualised pairs when the actualised pair was successful. The 'creative' portion of *creative compositionality* has to do with the fact that a new (complex) signal is more likely to be used when a new (complex) state arises—that is, the sender chooses a compositional signal with some likelihood, though she has never seen the complex state before.

This is supposed to provide an account of the emergence of compositional signals; however, the complex signals in this model do not have a genuine syntactic structure which then gets compositionally interpreted. Brochhagen (2015) points out that one requirement of compositionality is that the relations between complex signals and their constituent (simple) parts need to be generalisable to obtain a productive compositional system. In the case of

[23] See, for example, Hailman et al. (1985), Tempelton et al. (2005), and the discussion in Skyrms (2010a). For example, Snowdon (1990) points out, with respect to a particular type of birdsong, that '[t]he main limit of this complex grammatical system is that there is no evidence that any of the 362 sequences documented has any *functional* significance' (228, emphasis mine). Note however, that some experimental evidence for compositional syntax in birdcalls (Suzuki et al., 2016). This will be discussed in more detail in Section 2.3.

spill-over reinforcement, players have a propensity to increase constituent-based association (via the spill-over mechanism); however, elements of the language which are structurally analogous are not taken into account: '[p]layers are not sensitive to the overall architecture of their communicative system' (Brochhagen, 2015, 286).

2.2.4 Functional Negation

Steinert-Threlkeld (2016, 2017) introduces a type of functional compositionality in the form of his 'negation game'. The game is like an $n \times n$ signalling game, except there are $2n$ possible states and acts; thus, the sender has n atomic signals, $m_1, \ldots, m_n$, but the sender can also send signals of the form $\boxminus m_i$ for $1 \leq i \leq n$, as a sort of 'minimal negation' (in Steinert-Threlkeld's words).[24] This sort of behaviour seems empirically plausible given, e.g., the 'boom boom' prefix to alarm signals noted by Zuberbühler (2002) (discussed in section 2.3 below).[25]

The mathematical notion here is that of a derangement $f : [2n] \to [2n]$—namely, a bijective function with no fixed points—for $[n] = \{1, \ldots, n\}$. Further, f is applied to both the states and the acts. So, $f(s_i) \coloneqq s_{f(i)}$ and $f(a_i) \coloneqq a_{f(i)}$.

The model for minimal negation that Steinert-Threlkeld (2016, 2017) employs has much structure built-in. However, he is less concerned with the question of how compositional signals might arise as he is with the question of *why* compositional signals might arise. The intuitive answer is that it is more efficient to learn to compose signals, functionally, than it is to evolve new signals from scratch. In the example of alarm calls of vervet monkeys, he points out that having minimal negation would seem to make learning more straightforward: 'once signals for the three predators are known, the signals for their absence are also

[24] This is minimal negation in the sense that it captures some minimal intuitions about how negation should work: (1) every state has a negation, (2) the negation of a state is distinct from the state, and (3) distinct states have distinct negations.

[25] See Schlenker et al. (2014) for a semantic analysis of this type of signalling.

automatically known by prefixing with the negation signal. By contrast, with only atomic signals, three new unrelated signals would need to be introduced to capture the states corresponding to the lack of each predator' (388). He finds that for $n = 2, 3, 4$ (4, 6 and 8 states), agents communicating with an atomic language learn to communicate more successfully (with statistical significance); however, for more sophisticated 'worlds' where there are 14 or 16 states ($n = 7, 8$), agents learning a functional language perform better (again, with statistical significance).[26]

As such, several extant models purport to explain the emergence of compositional signalling. However, as will become apparent, I believe that this is the wrong way to explain complex signalling as it may have evolved in nature—this is not to say that describing how compositionality might arise is neither relevant nor interesting. I will discuss this in more detail in Chapters 3 and 4; for now, in the context of the salient distinctions between language and animal communication systems, one significant problem is that *genuinely* compositional signals are scarce, if nonexistent in nature. There is *very little* empirical evidence thus far for *compositionally* meaningful call sequences. Very little, however, does not mean none. Most current data that suggests compositional signalling in nature comes from Zuberbühler's study of several species of African forest monkeys (*Cercopithecus*). However, Fitch (2010) points out that such combinatorial phenomena are 'currently known only in African *Cercopithecus* monkeys, and nothing similar is known in other well-studied monkey species or any great ape. Thus, these provide little evidence of a "precursor" of syntax in the LCA [Last Common Ancestor]' (185).

In examining the salient differences, we have found a potential explanatory target. However, few consider the converse question of what animals are indeed capable. This possibly restricts the plausibility of our explanatory target. In the next section, we shall see some particular features of animal communication systems, and I will suggest that these preclude composi-

[26]The difference between the two languages for $n = 5, 6$ are not statistically significant.

tionality as a correct explanatory target. Thus, we require a new one. I will spend the rest of this chapter discussing empirical restrictions for compositionality, and I will present an alternative explanatory target in the next chapter.

2.3 Communication in Nature

I have said that communication is found everywhere in nature. In nonhuman animals, every taxon that has been investigated has displayed the existence of some type of communication system (Lishak, 1984; Lugli et al., 2003; Marler and Slabbekoorn, 2004; Belanger and Corkum, 2009; Houck, 2009; Mäthger et al., 2009; Bruschini et al., 2010; Costa-Leonardo and Haifig, 2010; Haddock et al., 2010; Wyatt, 2010; Thiel and Breithaupt, 2011). Communication can occur between both conspecifics and heterospecifics (Rabin et al., 2003; Magrath et al., 2007; Lea et al., 2008; Pope and Haney, 2008; Touhara, 2008; Shabani et al., 2009; Bruschini et al., 2010) and potentially across phyla, depending upon one's definition of communication (Schaefer et al., 2004; Gera and Srivastava, 2006; Raguso, 2008).

Considering *how* animals communicate, we see that signalling occurs across virtually every possible modality. Chemical signals, taking advantage of pheromones, are the most common in both aquatic and terrestrial environments (Ayasse et al., 2001; Belanger and Corkum, 2009; Houck, 2009; Harder and Jackson, 2010; Wyatt, 2010; Thiel and Breithaupt, 2011; Zhang et al., 2011). Rosenthal (2007) examines how visual signals take advantage of variations in light, symmetry, size, and coordinated movements; some species have dedicated physical structures for delivering signals to other animals (Wilkinson and Dodson, 1997). Neither tactile nor electrical signalling is particularly well understood; nonetheless, tactile signalling has been documented in a variety of species, including deer mice (Terman, 1980), ants (Pratt, 2005), and between cleaner fish and client reef fish (Bshary and Würth, 2001), and electrical signalling has been observed to play a role in dominance relations (Fugère et al.,

2011). Finally, vibratory signals include signals produced by an acoustic apparatus, such as a larynx or syrinx, or other multi-purpose morphological features, such as beaks (Wilkins and Ritchison, 1999) or feet (Rose et al., 2006; Delaney et al., 2007).

Signals may be used for a variety of purposes and interactions, including (though not necessarily limited to) mate choice, resource competition, predator-prey encounters, parent-offspring dynamics, and social group cohesion (Bradbury and Vehrencamp, 2011).[27] Bradbury and Vehrencamp (2011) argue that the essential feature of a signal, regardless of its modality, is that it is conspicuous against background noise since a signal which cannot be detected in the first place will be of no use in an evolutionary setting. As was noted above, under the signalling-game framework we can remain agnostic about what constitutes a signal—e.g., whether it is verbal, gestural, chemical, etc.—all that is required for something to be a signal is that it be a pattern of stimulation produced by an individual and to which another individual can respond.[28] There is a good discussion of the 'currency' of linguistic discourse in Hurford (2012, Sec. 3.4), which is taken to be a 'sentence-like' unit. Despite the difficulty in defining, exactly, what a signal is—for example, whether it is indicative or imperative—it is possible to simply specify that it is the basic unit of discourse and say nothing further.[29]

[27] On the signalling-game framework, it was assumed that communication requires coordination, which in turn requires cooperation. However, Hurford (2007) notes that territorial calls are, by definition, anti-social, to the extent that they deter further or future contact (186). Even so, the cooperative aspects of the signalling game can be relaxed extensively—for example, Wagner (2012) shows that communication (in the sense of information transfer) is possible even when communicators have completely opposed interests, as in a zero-sum game.

[28] When the individual producing the signal has no control over its production, it is called a *cue*; when the individual receiving the signal has no control over its response (to the signal), this is called *sensory manipulation*—I will discuss this in more detail in Chapter 3.

[29] See, for example, Harms (2004a,b), Huttegger (2007b), Zollman (2011), and Millikan (1995). Note that Grafen (1990) surveys the difficulties in defining *what* a signal is in the first place, and comes to the conclusion that he is 'unable to offer a formal definition of signals in terms of game theory' (536). See also Hurford (2007, Sec. 6.1).

Signalling might be understood as the *collaborative sharing* of information.[30] Implicit in this assumption is that reliable signals benefit *both* the sender and the receiver and in the same way. If this were true, then signalling behaviour could be selected for in an obvious way. However, this view was challenged by Dawkins and Krebs (1978), taking a gene-centred, and individualistic view of evolutionary advantage. They argue that signalling is better understood as the manipulation of receivers by senders, rather than collaborative sharing of information between senders and receivers. However, Hinde (1981) points out that this ignores the fact that the receiver, too, can extract information for her own benefit from an unwitting sender. Krebs and Dawkins (1984) highlight that there is a tension between the sender and receiver of a signal, and for such a system to be evolutionarily stable, it must provide a *net benefit* to both. That is, it may pay more for a sender to exaggerate her signal, but the receiver will eventually ignore unreliable or uninformative signals. Honest signalling might evolve when the signal is costly for the sender to send. One such theory involves the *handicap principle* (Zahavi, 1975; Grafen, 1990; Zahavi and Zahavi, 1997).

In fact, there are two types of costs that can be associated with such handicaps. *Efficacy costs* refer to the costs that are associated with the ability to generate a signal such that it can be perceived effectively by its intended recipients—for example, signalling to attract a mate in a sparsely dispersed species. *Strategic costs*, on the other hand, are whatever costs are incurred over and above the signalling itself—the function of which is supposed to be to ensure honesty. The handicap principle only examines strategic costs.

It is pointed out that, for a handicap-based signalling system to be stably reliable, the strategic cost for low-quality signallers needs to be greater than the cost for high-quality signallers.[31] Thus, signals must be differentially costly for such a system to evolve in the

[30]Indeed, this assumption is somewhat built in to the signalling-game model, as it was presented in Chapter 1, in the sense that Lewis (1969) presupposed that the signalling game was arbitrarily close to the cooperation end of the spectrum suggested by Schelling (1960).

[31]See Enquist (1985); Pomiankowski (1987); Grafen (1990); Johnstone (1995).

first place.[32] Fitch (2010) points out that any plausible theory of the evolution of language is strongly constrained by the fact that 'human language appears, at least superficially, to have evaded all known theoretical routes to honest signaling' (196), as is evidenced by the feature of languages that humans 'freely and continuously use language to share accurate information with unrelated others'.[33] In some cases, the evolutionary fitness of (honest) signalling is apparent: for example, antler size provides an reliable signal of male phenotypic quality in roe deer (Vanpé et al., 2007), similarly for roaring in red deer (Clutton-Brock and Albon, 1979; Reby and McComb, 2003).

Signals in nature require a type of lexicon. When signalling is holophrastic—atomic signals express unique ideas, concepts, states, etc.—an increase in state-complexity gives rise to a correspondent increase in lexicon size. However, if there exist a small number of rules for combining simple signals into complex signals, then fewer resources are required for storing individual signals. That is, more expressions can be made without a correspondent increase in lexicon size. Many animals employ 'complex' signals to maximise the diversity of their communicative ability without excessively inflating their lexicon (Hebets and Papaj, 2005). Complex signals have components which may or may not be delivered simultaneously, which may or may not occur in the same sensory modality, and which may or may not elicit independent or distinct behavioural responses when sent individually (McCowan et al., 2002; Candolin, 2003; Hebets and Papaj, 2005; Kulahci et al., 2008; Gordon and Uetz, 2011). The complexity of a signal may be learned, or it may be innate, though many species require a period of learning before they can correctly perform or deliver complex signalling behaviour (Rosenthal, 2007).

Systems of communication in nature range from extremely simple—as with quorum signalling in bacteria or pheromone release in moths—to extraordinarily complex and multi-modal—for

[32]See the discussion in Maynard Smith and Harper (2003).

[33]There is a word for this in German—*Mitteilungsbedürfnis*—which describes the apparent *need* or drive in humans to share thoughts and feelings.

example, communication systems utilised by *C. l. familiaris* involve olfactory signals, visual cues, and a wide range of vocalisations.[34] Several complex systems have been well studied. Human communication, at the extreme end of the spectrum, involves several signalling modalities. Some of these are innate, and some of these are learned—for example, it has been shown that smiling (a visual mode of signalling) is common to all human cultures and can be observed in blind newborns (Eibl-Eibesfeldt, 1973).[35]

To begin, I will examine several communication systems that occur in nature and which are often invoked in discussions of compositional signals. *Bona fide* compositionality is often taken to be rare or nonexistent in animal communication systems, as we have seen. A few examples in nature that might be compositional exist; however, the fact that there is disagreement about whether or not these communication systems are *actually* compositional, as opposed to holistic, is evidence that we need a clear understanding of what it means for a communication system to be compositional, as opposed to say complex or combinatorial—this will be a focus of the discussion in Chapter 4.

I have already suggested that there is an apparent adaptive advantage for combinatorial communication in a communication system: namely, fewer elements are required to be stored in memory to produce the same possible number of messages, thus allowing for more efficient communication.[36] Nonetheless, several suggestions have been made for why combinatorial communication is so rare in nature. For example, Zuberbühler (2002); Ouattara et al. (2009)

[34]Note that, although these examples pertain primarily to conspecifics, dogs are also extremely adept at interpreting signals from humans. Contrary to popular belief, dogs appear to respond to (human) word *meaning* rather than merely intonation (Andics et al., 2014, 2016) and are sensitive to subtle visual cues, such as attentional focus (Call et al., 2003; Virányia et al., 2004). Additionally, dogs are able to maintain a reasonably sized lexicon of distinct words (Pilley and Reid, 2011).

[35]Fitch (2010) suggests that the use of words like 'innate' or 'instinct' in discussions of the evolution of language is best avoided, in the sense that it is not clear what they contribute to the discussion, but they almost certainly fuel fruitless disagreement—e.g., the *nature* versus *nurture* debate (Tinbergen, 1963; Lorenz, 1965). Even so, it is sometimes relevant to keep in mind the distinction between the different capacities that individuals might have. For example, all humans have the *capacity* to spontaneously acquire language—the idea of a *universal grammar*. I will sometimes have occasion to distinguish whether behaviour is innate, learned, or has the capacity to be learned as we see a variety of examples of communication systems in nature.

[36]See, Nowak and Krakauer (1999); Nowak et al. (2000).

suggest that it is cognitively more challenging than simple communication. Alternatively, Nowak and Krakauer (1999); Nowak et al. (2000) indicate that the advantages of combinatorial communication only exist after the number of signals in a communication system exceed some threshold. (This is consistent with the conclusions of Steinert-Threlkeld (2016); Kottur et al. (2017).) Finally, Lachmann and Bergstrom (2004) argue that a combinatorial system of communication is more susceptible to dishonest signalling.

In this section, I examine several animal communication systems, chosen for the particular properties they exhibit. Quorum signalling in bacteria (2.3.1) is the paradigm example of a signalling system in nature; the signalling game well models this. Moving up in complexity, the oft-cited case of alarm calls in, e.g., vervet monkeys (2.3.2) highlights several properties of simple signalling behaviour, including functional reference. The 'waggle dance' of honeybees (2.3.3) is a contender for complex signalling insofar as messages are encoded in two dimensions, and are potentially infinite, like human languages (though as we shall see they are not in fact). Finally, the two most cited cases for complex signalling, concerning combinatorial signals (2.3.4) and possibly compositional signals (2.3.5), are presented.

In the previous section, the explanatory target of an analysis of language origins has been narrowed down by examining how human languages are disparate from simple communication systems. The subsequent section takes the converse (and often ignored) approach of looking toward what possible precursors to this unique ability might demonstrably exist in nature. My conclusion is that there are *no such plausible precursors in nature*; therefore, compositionality is not the correct explanatory target.

2.3.1 Simple Communication: Quorum Signalling

At the far end of the communication spectrum, where the simplest forms of communication lie, we have *quorum signalling*. Bacteria provide a nice example of the strengths of the

signalling-game framework: the extent to which communication in bacterial organisms is well modelled by the most straightforward sort of signalling games is evidence in favour of how easy it is to communicate, regardless of cognitive sophistication, as Skyrms (2010a) argues. Whiteley et al. (2017) suggest that 'microbes are highly gregarious communicating organisms and that bacterial communication can modulate a range of behaviours that are important for fitness (reproductive success)' (313).

Quorum-sensing, which was initially discovered by Nealson et al. (1970), is the regulation of gene expression in response to fluctuations in cell-population density (Bassler, 1999; Miller and Bassler, 2001).[37] Nealson et al. (1970) observed that the bioluminescent bacterium *Vibrio fischeri* produce a luminescent enzyme (*luciferase)* only if cultures had reached a threshold population density. When the bacteria population is at low cell concentration, they do not express this luciferase gene.

Subsequent research (Eberhard et al., 1981; Engebrecht et al., 1983; Engebrecht and Silverman, 1984, 1986, 1987) revealed that the actual 'autoinducer' used by *V. fischeri* is an *acylated homoserine lactone* (AHL) signalling molecule. Though AHL-based quorum sensing was initially thought to be isolated to certain marine bacteria, it was shown that, e.g., *Erwinia carotovora* and *Pseudomonas aeruginosa*—two species of non-marine bacteria—possessed a quorum-sensing system very similar to *V. fischeri* (Gambello and Iglewski, 1991; Bainton et al., 1992; Ochsner et al., 1994; Pearson et al., 1994; Latifi et al., 1995). It is now known that a wide range of organisms possess homologues of the luxI and luxR genes from *V. fis-*

[37]Bioluminescence was being studied in the late 1960s; it was observed that *Vibrio fischeri* cultures only produced light when a large number of bacteria were present. The initial suggestion for why this should be the case was that the culture medium contained an inhibitor which was removed when a large number of bacteria were present (Kempner and Hanson, 1968). However, it was later discovered that the luminescence was initiated not by the removal of an inhibitor but rather by the accumulation of an activator molecule, which was termed an 'autoinducer' (Nealson et al., 1970; Eberhard, 1972). (Skyrms (2010a) misidentifies the year of discovery as 1977.) Note that '*auto*inducer' turned out to be a misnomer, since bioluminescence can be induced across species (Greenberg et al., 1979). To avoid confusion, the term 'quorum sensing' was introduced by Fuqua et al. (1994). See also Turovskiy et al. (2007).

cheri and take advantage of a quorum-sensing signalling system; see, for example, Turovskiy et al. (2007).

In the signalling-game model, the relevant states of the world are *quorum* or *not-quorum*, the signals are chemical signalling molecules (such as AHL), and the actions vary depending upon the species—for example, to control bioluminescence, biofilm formation, virulence, or spore formation. Thus, this situation is well modelled by a 2×2 signalling game (though it may not be atomic given, e.g., the actual distribution over the state space). For more on the communicative aspects of quorum sensing in bacteria, see Schauder and Bassler (2001); Taga and Bassler (2003); Bauer and Mathesius (2004) and the discussion in Skyrms (2010a).

2.3.2 A Classic Example: Alarm Calls

One function of animal signals is to express the *internal* (e.g., emotional, intentional, or motivational) state of the animal. Chimpanzees, for example, display abundant abilities for communicating emotional states and altering social interactions in addition to some limited referentiality; however, they lack the unlimited generative ability of human language to convey novel thoughts (Seyfarth and Cheney, 2005; Slocombe and Zuberbühler, 2005). Signals may also transmit information about *external* states or events. In this case, a signal is *functionally referential* (Marler et al., 1992; Macedonia and Evans, 1993; Hauser, 1996; Zuberbühler, 2000).[38] Macedonia and Evans (1993) suggest that alarm calls exist on a continuum, with 'response-urgency' on one end and functional-referentiality on the other.[39] A signal is said to be referential to the extent that it refers to something (i.e., in the outside

[38]This terminology was suggested by Marler et al. (1992) to make clear that, though such context-specific alarm calls are 'word-like' in their function, the concept of functional reference was to remain 'neutral about the underlying mental processes' (67). The machine-learning literature on emergent communication uses the term 'grounded communication' to denote functional reference; however, 'grounding' is often less clearly defined and is used less precisely.

[39]Fitch (2010) notes that there are actually two continua here: '[a] call may be functionally informative about the outside world, from the listener's viewpoint, while for the signaller it is simply an expression of its current emotional state' (191). See also the discussion of separating the sender and receiver in Godfrey-Smith (2018).

world), and *functionally* referential to the extent that it is *context-dependent*—the contextual feature defines the referent—but such signals are *stimulus-independent*: they elicit an appropriate response, even in the absence of the referent. This is opposed to, e.g., high- or low-urgency alarm calls, which do not refer to any particular thing (Blumstein, 2007).

Alarm calling is common to several social species with predators. There are well-documented cases of alarm-call signalling in a variety of species, including vervet monkeys (Garner, 1892; Seyfarth et al., 1980a,b; Cheney and Seyfarth, 1990), Putty-nosed monkeys (Arnold and Zuberbühler, 2006b, 2013), Campbell's monkeys (Zuberbühler, 2001), Diana monkeys (Zuberbühler et al., 1999), black-fronted titi monkeys (Cäsar et al., 2012; Berthet et al., 2018a,b), white-faced capuchin monkeys (Digweed et al., 2005), pale-winged trumpeters (Seddon et al., 2002), and domestic chickens (Evans et al., 1993; Karakashian et al., 1988), for example.

Cheney and Seyfarth (1990) offer some evidence for the functionally-referential nature of vervet alarm calls. They show how vervets can become desensitised to false alarm calls. When a vervet (or, a loudspeaker in the case of their experiments), sends a 'leopard' signal several times, when there is no leopard present, the group will start to ignore the call. However, if a *different* vervet sends the leopard alarm call, the vervets will respond again. This appears to single out an individual as being more or less reliable. Vervets also give *wrrr* and *chutter* calls when they meet another group. When the *wrrr* call is (falsely) sent several times, again, the vervets will stop reacting to it (no group is seen coming into view). However, in this case, they will also ignore the *chutter* call, since these two calls seem to pick out the same thing. This is not true of the alarm calls: when vervets become desensitised to a 'leopard' alarm call to the point that they ignore it, they will still react to an 'eagle' or 'snake' alarm.

Vervet monkeys have distinct alarm calls for different predators: a 'bark', a 'cough', and a 'chutter' to warn of leopards, eagles, and snakes, respectively. In each case, a particular state

of the world (a specific predator being present) has a uniquely appropriate action. For a leopard, the best response is to climb a tree and out onto a branch where the leopard cannot follow; however, this action would be *inappropriate* for an eagle's being present. When an eagle is present, the most appropriate response is to dive into a bush—again, this would be inappropriate for the snake's being present, given that the snake may well be in or near the bush, and this would be similarly inappropriate for a leopard's presence. For a snake, the best response is to scan the ground and move away (or, for a group of vervets, to stand tall and mob the snake). Thus, we have a 3×3 signalling system of the form given in Figure 2.4.

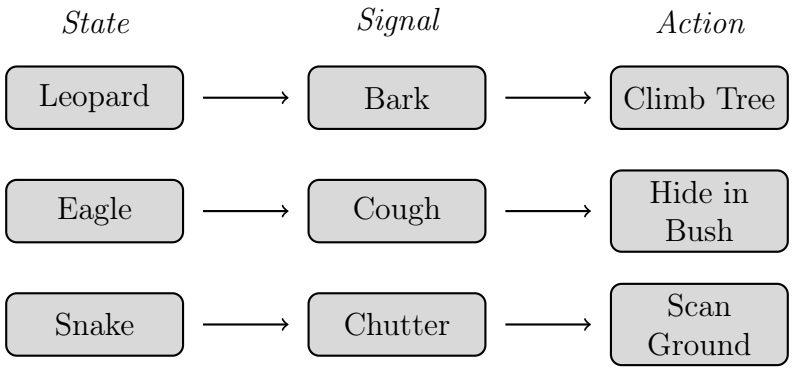


Figure 2.4: Simple vervet monkey signalling system

However, as is often the case, the real world contains much more noise than this simplified picture suggests—though the example is often presented in this form, this is not entirely accurate. In fact, the leopard call is a bark, but if a female leopard is spotted, the vervet will elicit a high-pitched squeal; this call generates the same response in the receiver as a bark, and so might be interpreted as constituting a synonym. Additionally, the data presented in the original studies (Seyfarth et al., 1980a,b) indicated substantial variation in responses, and the reactions are probabilistic.[40]

Almost a century before the original playback experiments provided by Seyfarth et al. (1980a,b), Garner (1892) performed playback experiments using an Edison phonograph and noted that alarm and food calls elicited appropriate responses. This latter point highlights

[40]Note also that Darwin (1871) reported that, when exposed to a stuffed snake, three monkeys 'uttered sharp signal cries of danger, which were understood by the other monkeys'.

another way in which the picture is slightly more complicated: vervets do not just vocalise for alarm calls. They also call when they find food, in aggressive confrontations, and during sexual activity, among others. Vervets additionally vocalise via grunting in a variety of circumstances: (a) when a submissive meets a dominant individual, (b) when a dominant meets a submissive individual, (c) when one vervet goes out into an open area, and (d) when a vervet comes across an out-group conspecific (Cheney and Seyfarth, 1982, 1990).

Though these other contexts of vocalisation are often ignored when the vervet signalling system is presented, the alarm call system indeed *roughly* fits the model of a simple signalling game—i.e., when we talk about states of the world and actions pertaining to predators as a *context* that is independent of, e.g., mating or foraging for food. Further, alarm calls provide a particularly salient example, in the sense that it is simpler to demarcate *what* the relevant states of the world and the appropriate actions are.

Another well-documented, and perhaps slightly more complicated, case can be observed in the signalling behaviour of ring-tailed lemurs (primarily reported in Macedonia (1993) but see also Bolt and Tennenhouse (2017)). Ring-tailed lemurs produce several different types of alarm calls when they perceive a threat: *gulps* are produced upon perception of carnivores, quickly moving humans, and other potentially threatening objects; *rasps* are produced in response to seeing large airborne birds, like hawks; *shrieks* are produced in response to low-flying birds (higher urgency); *yaps* or *barks* are produced in the presence of potentially dangerous mammals (the response here is to mob the animal, to scare it off).

Additionally, as with vervets, ring-tailed lemurs vocalise non-alarm-call signals in a reasonably diverse—though well-differentiated—assortment of circumstances, including affiliative and agnostic vocalisations. They produce single or serial open- and closed-mouth clicks, which appear to indicate concern, with the open/closed difference relating to the degree of concern. In their forest habitat, where individuals in a group cannot always be seen, these click signals are used in the absence of a predator to maintain contact with the group—to

keep close together, or when a mother calls her offspring, for example. Further, ring-tailed lemurs produce alternating choruses of moans or mews, and wails or howls, which in group alternation can help a lost group rejoin the main group, or in solitary alternation can serve for attention/response of alternating males and females. Finally, a distinct set of calls (*yips*, *cackles*, *squeals*, *twitters*, *plosive barks*, and *chutters*), are utilised to maintain dominance relations. (Females dominate males, and, within the sexes, there is a definite hierarchy that is maintained.) Thus, the ring-tailed lemur has a repertoire of more than 20 acoustically distinguishable signals, each with distinct conditions of use.

Ring-tailed lemurs also make use of chemical signals which are received via the olfactory system (Kappeler, 1998); however, the available documentation on this system is significantly more sparse than the literature on vocal signals. In general, chemical signals of this sort are of less interest for this particular project—focusing on complex signals—because the scent has no internal structure.[41]

2.3.3 Communication in Honeybees: The Waggle Dance

The idea that bees used something 'like language' to communicate food sources was a contested issue in the 18th and 19th century. That bees dance has been known possibly since Aristotle (or earlier). The oft-cited passage from *History of Animals, Book IX*, reads

> On each expedition the bee does not fly from a flower of one kind to a flower of another, but flies from one violet, say, to another violet, and never meddles with another flower until it has got back to the hive; on reaching the hive they throw off their load, and each bee on his return is accompanied by three or four companions. One cannot well tell what is the substance they gather, nor the exact process of their work. (Aristotle, 1995a, 241)[42]

[41] Though I admit that this statement might well be anthropocentric, there appears to be no possibility for syntax here, since the signal is sent (and perceived) as a unitary whole—which is not to say that animals with more well-refined olfactory abilities are not able to pick up on *very* subtle distinctions.

[42] In a brief article, Haldane (1955) calls into question the translation of the sentence 'on reaching the hive they throw off their load ...' [ὅταν δ' εἰς τὸ σμῆνος ἀφίκονται ἀποσείονται ...]. In particular, he points out

This dance was rediscovered, and noted, by Spitzner (1788): '[w]hen a bee has come upon a good supply of honey anywhere, on her return home she makes this known in a peculiar way to the others. Full of joy she twists in circles about those in the hive. ... for many of them soon follow when she goes out again',[43] and independently by Unhoch (1823), who notes that '[w]ithout warning, an individual bee will force its way suddenly among 3 or 4 motionless ones, bend its head toward the surface, spread its wings, and shiver its raised abdomen a little while. ... The dance mistress often repeats her dance four or five times in different places. ... What this dance really means I cannot yet comprehend'.

Thus, the purpose of the dance was well noted, but not yet understood. Lubbock (1874) noticed that bees (and ants) often find food sources with some degree of accuracy after another bee [ant] has already been there; however, he points out that a 'simple sign' would likely suffice for this behaviour—though, at the time, no evidence is found for this conjecture. Maeterlink (1901) and Lineburg (1924) performed experiments with bees and concluded that the honeybee dance must serve the simple purpose to attract the attention of other bees. It was further suggested that the reason the bees are so successful in finding the original food source is because of odour perception, rather than information transmission from one bee to another. For an overview of several experiments, see Gould (1975).

von Frisch (1967) famously decoded the dancing behaviour of the honeybee (*Apis mellifera*), which can be understood as a signal comprised of two components. The first component—distance—is signalled by the duration of the dance; the longer the dance, the further the food source. The second component—direction—is signalled by the angle to the vertical of the dance; the angle encodes the direction to which the food lies, relative to the sun's position. Since the location is described in terms of two distinct components, and receivers

that 'σμῆνος' might be translated as either hive or *swarm*—the latter is relevant in this context given the fact that bees do not perform their dance until a reasonable number of spectators are there to observe it. He further points out that, though 'ἀποσείειν' is 'to shake off', the form in which it is used here—ἀποσείονται—is also used by Aristotle in *History of Animals, Book VI* to describe the post-copulatory movements of hens, 'which are certainly not shaking anything off themselves' (24).

[43]Quoted in von Frisch (1967).

reliably respond to both of these components by flying a certain distance in a particular direction, this system of communication is arguably compositional: location is composed of both distance and direction (Hurford, 2012).

This is interesting because human languages are effectively open, whereas animal communication systems are usually closed systems—they consist of a (generally very small) fixed set of possible messages and no compositional syntax. The waggle dance appears to be an open communication system on the technicality that there is a potentially uncountable number of 'distinct' dances at an extremely fine-grained level, which could 'refer' to a potentially uncountable number of locations with an arbitrary degree of specificity. This is because the space being referred to—distance and direction—and the dance itself are *continuous* rather than discrete.

However, this is not the case, in fact. Schürch and Ratnieks (2015) have analysed the informational content of the two vector components (direction, distance) encoded by the bees' waggle dance, and estimate that they convey approximately 2.9 and 4.5 bits of information, respectively.[44] That the dance only encodes, as Schürch and Ratnieks (2015) claim, up to 7.5 combined bits of information, implies that states of nature are being partitioned into coarse-grained chunks. For example, the direction component of the state vector is a continuous 360 degrees; however, the direction component of the message that encodes the state-space consists of 4.5 bits, meaning that the compass is being divided into a couple of dozen distinct values, approximately 15.9 degrees each.[45]

[44]This is an extension of original work by Haldane and Spurway (1954), using data from von Frisch (1946) (note that they only calculate the information for the direction component). However, it has been pointed out that their data was biased in favour of 'good' dancers, so it underestimated systematic errors in communication (Schürch and Couvillon, 2013; Schürch et al., 2013). Schürch and Ratnieks (2015) take advantage of more accurate recent data (Couvillon, 2012; Couvillon et al., 2012).

[45]Assuming all directions (states partitions) are equiprobable, 22 state partitions results in approximately 4.4594 bits of information and 23 state partitions results in approximately 4.5236 bits of information.

2.3.4 Complex Signalling: Putty-Nose Monkeys

Arnold and Zuberbühler (2006a,b, 2008, 2013) suggest that putty-nosed monkeys have two distinct alarm calls, one for each of two predators: leopards (a 'pyow' sound) and eagles (a 'hack' sound)—though they are not functionally referential. Each of these signals sent on its own produces an appropriate evasive action in the receivers of the signals. For leopards, the action is climbing up a tree, and for eagles, the action is hiding in bushes. Putty-nosed monkeys can combine these two calls into longer sequences, with pauses that break strings into discrete blocks. However, the resultant action is not a combination of evasive action for both predators; instead, the combined call seems to signal the movement of the group to a new location. These 'pyow-hack' combinations reliably instigated group-movement in the monkeys (Arnold and Zuberbühler, 2006a,b). They suggest that this is parallel to the linguistic combination of two distinct signals A and B into a signal AB whose meaning is distinct from either of the individual signals.

Scott-Phillips and Blythe (2013) use this as an empirical example of 'composite' signalling. Though they use the words 'composite' and 'combinatorial', they clearly have in mind something like compositionality, given that they claim that there exists 'one extreme exception to the norm of non-combinatorial communication: human linguistic communication' (5). However, it is not clear how their differentiation of combinatorial and non-combinatorial communication can be used to clarify what we mean by compositional communication. For example, their model does not capture sensitivity to the syntactic structure, which is apparent in complex signals in bird song and whale song. Nor does it capture a notion of semantic composition—the meaning of a fully composite signal pair need not have anything to do with the meaning of its parts. This applies, particularly, to the combinatorial 'pyow-hack' call of the putty-nosed monkey. Given that 'pyow-hack' has no structural relation to its constituent parts, it could be treated mathematically by a brand-new signal. That being said, the ability to combine signals to create new 'atomic' signals (similar to Barrett (2009))

gives an advantage to the extent that limited lexical or computational resources can be used to express more, and so to avoid bottlenecks.

2.3.5 Compositional Signalling: Campbell's and Diana Monkeys

The most cited case in the signalling literature is that Campbell's monkeys emit a low-pitch 'boom' note preceding an alarm call in contexts when the danger is not immediate. When the threat is urgent, they produce the alarm call in isolation. Playback studies with Diana monkeys suggest that the receivers of such signals interpret the 'boom' call as a modifier for the regular alarm call by reacting less to the subsequent alarm call than in contexts where the alarm call was sent alone (Zuberbühler, 2002).

Note that several of these studies make claims about the signals being *functionally referential*. Zuberbühler et al. (1999) attempt to answer the question of whether predator alarm-calls really are functionally referential in the following way. Female Diana monkeys both elicit alarm calls upon viewing a predator first-hand and respond to alarm calls of male Diana monkeys by repeating the call. In their experiments, Zuberbühler et al. (1999) played-back various kinds of pairs of stimuli—for example, a matching pair, as in an eagle alarm call followed by the characteristic shriek of an eagle, and a mismatched pair, as in an eagle alarm call followed by the signature growl of a leopard. In each case, the pairs of stimuli were separated by an interval of 5 minutes of silence.

In the experiment, the female monkeys displayed less concern upon hearing, e.g., the characteristic shriek of an eagle five minutes after the eagle alarm call, and they showed more concern upon hearing a characteristic leopard growl five minutes after hearing the eagle alarm-call. This intuitively makes sense because, in the first case, the shriek of the eagle should be expected, and it affords no new information, whereas in the second case the

growl of the leopard contains further information (which is surprising, hence the concerned response).

They conclude that the alarm calls do not just serve to trigger (i.e., behaviourally) an evasive response, since the Diana monkeys have an 'idea' of the relevant predator in mind for at least five minutes following the initial alarm call—as is elicited by the response upon hearing the predator itself. Hurford (2007) points out that this behaviour indeed meets the criteria for a call being 'functionally referential'.[46]

Steinert-Threlkeld (2016, 2017) uses this 'boom' prefix as an empirical justification for his model of functional negation; however, there are two things to note about that justification. First, he admits that the use of a negation-like call, that relies upon the fact that the other vervets, for example, already know how to respond to barks 'seems prima facie superior' to the individuals having to learn a brand-new signal 'since it leverages the existing signalling behavior' (385). Though he cites the work of Zuberbühler (2002), he admits that his proposal is 'purely speculative'; even so, he suggests this research 'makes it plausible' that his model will accord with theories of the gradual emergence of compositionality.

However, let us consider whether the proto-syntactic signalling of Campbell's monkeys really is a plausible candidate for a proto-compositional precursor of human linguistic capacities. The most recent last common ancestor (LCA) between the great apes (including *Homo*) and old-world monkeys (including *Cercopithecus*) existed approximately 25 MYA (million years ago), during the Paleogene period. Let us assume that the LCA to great apes and old-world monkeys *did* have functionally proto-compositional syntax. This requires that proto-compositional communication evolved at least 25 MYA. In this case, we should expect that most old-world monkeys *and* great apes would show signs of using proto-compositional syntax. However, as far as we can tell, they do not—except for the few mentioned *Cer-*

[46]See also Marler et al. (1992).

copithecus species, no other known related species utilises such apparently compositional capacities.

Assuming these dispositions evolved in the LCA, 25 MYA, and given that they do not appear in most decedents of the LCA, this implies that this disposition was lost—and lost more readily than it was held onto—in almost all other *Catarrhine* species. Meanwhile, these dispositions in few *Catarrhine* species remained utterly unchanged for more than 20 million years, and these dispositions were lost in *almost all* great apes and old-world monkeys.

Furthermore, the most recent common ancestor of humans and chimpanzees, for example, is theorised to have lived 10-4 MYA; *Australopithecine* species evolved *after* this break, with *Homo* likely evolving approximately 2 MYA. While it is controversial whether Neanderthals had language ($0.4 - 0.04$ MYA), let alone whether *H. heidelbergensis* had language ($0.7 - 0.2$ MYA), it is generally accepted that *Australopithecine* species ($3.9 - 2.9$ MYA) did *not* have language (Fitch, 2010). This implies that this proto-compositional disposition evolved into fully-fledged natural language on the order of $0.2 - 2.0$ MYA.[47] Therefore, assuming that the LCA of old-world monkeys and great apes had proto-compositional syntax implies that for the vast majority of the evolutionary history more than 100 species, such abilities were generally lost, while for one or two species, the complexity of these abilities grew exponentially into natural language in a relatively short period.

Though this is not technically impossible, this seems improbable. Of course, Jackendoff (2010) highlights that just because something is hard to imagine, this does not make it impossible. Even so, the scant empirical evidence for compositional precursors to human language comes paired with an improbable evolutionary history. Thus, it stands to reason that this syntactic disposition was not present in the last common ancestor of *Homo* and *Cercopithecus* and Campbell's monkeys. However, if such a proto-compositional disposition

[47] Berwick and Chomsky (2016) give a more conservative estimate, between $0.06 - 0.2$ MYA, corresponding to the first anatomically-modern humans and the last exodus from Africa.

was not present in the LCA, then it cannot serve as a precursor to human compositional language. The more plausible alternative is that there is *no* empirical evidence for any precursor to human language in nature.[48] It appears that combinatorial syntax in *Cercopithecus* is more accurately described as analogous, rather than homologous, to syntax in humans.

That said, Steinert-Threlkeld (2016) does not purport to show *how* a rudimentary proto-compositional form of signalling did indeed evolve; rather, he was interested in showing the conditions under which compositional signalling might be advantageous from an evolutionary perspective. What he finds is that such complexity is only beneficial in a suitably complex world. Thus, more complex dispositions, such as binary operators or partially recursive syntax, would require a larger state-space. This at least provides a tentative answer to the question of why compositional signalling is rare, though communication is universal in nature.

2.4 Alternative Accounts

In the preceding discussion, it was suggested that a necessary condition for explaining the emergence of language from simple communication systems was compositionality, in the sense of it being an apt explanatory target; in the previous section, I suggested that this is not so: there is scant evidence of compositional precursors in nature, and what evidence there is cannot be appealed to. However, the following fact was ignored: several factors likely led to the evolution of complex communication. Even so, it is not entirely apparent that the evolution of complex syntax is the foremost antecedent condition for language. On the one hand, gestural cues have been offered as an alternative explanation for an an-

[48]Note that an account which posits the sudden emergence of compositional syntax does not fall prey to these criticisms since it requires no proto-compositional precursor to compositional language. This is a virtue of the so-called *saltationist* perspective on language origins. However, this view is not without its own problems (LaCroix, 2020b). Therefore, *if* gradualism is the correct approach to language origins, then compositionality *cannot* be a correct target.

tecedent protolanguage, rather than a lexical protolanguage which forms mappings between state/act pairs and signals.[49] Gestural signals might come in a variety of types. For example, *pantomime* is an iconic gesture. Another iconic gesture might involve the spatial representation of size. Somewhat less iconic, though not necessarily entirely conventional, are *deictic* gestures—canonically, pointing with one's finger.[50] Finally, altogether conventional are *emblematic* gestures—e.g., *thumbs up*—which have conventional meanings that depend upon the culture—indeed, several conventional gestures that have a positive connotation in western, English-speaking cultures are deeply offensive elsewhere in the world.

Meaningful gestures are often learned in infants earlier than meaningful vocal signals and are often interchangeable in the early stages of development with spoken words. Fitch (2010) points out that 'at a crucial point in development, coinciding closely with the onset of two-word phrases, children combine vocalizations and gestures more synergistically: for example, denoting actions with words and objects with pointing' (435). Further, similar combinations of gestural and lexical tokens are observed in language-trained bonobos (Savage-Rumbaugh et al., 1993).

In studying the evolution of complex signalling behaviour, there is a question of what we mean by complexity. Hurford (2012, Sec. 5.3) discusses at length what it means for one language to be more complex than another, from the point of view of linguistics, and points

[49]However, as has been mentioned, we might note that there is nothing inconsistent about assuming that the signal in the signalling game is a *gestural* signal. Indeed, given the abstract framework, a signal could be anything, as long as it is able to come to represent, or be associated with, a state/act pair.

[50]This might be understood as iconic, in line with the idea that '*Quand le doigt montre le ciel, l'imbécile regarde le doigt*'. (This idiom has a variety of instantiations, and is attributed, variously, to Confucius or the Shurangama Sutra; a similar formulation—'When you show the moon to a child, it sees only your finger'—is said to be a Zambian proverb.) However, even a gesture of pointing—naturally salient, though it may be—can be interpreted as conventional, as in the discussion of Wittgenstein (1953, ¶85): '[a] rule stands there like a sign-post. — Does the sign-post leave no doubt open about the way I have to go? Does it shew which direction I am to take when I have passed it; whether along the road or the footpath or cross-country? But where is it said which way I am to follow it; whether in the direction of its ringer or (e.g.) in the opposite one? — And if there were, not a single sign-post, but a chain of adjacent ones or of chalk marks on the ground—is there only *one* way of interpreting them? — So I can say, the sign-post does after all leave no room for doubt. Or rather: it sometimes leaves room for doubt and sometimes not'. see also Wittgenstein (1953, ¶454).

out that 'it could be claimed that complexity in one part of a language balances simplicity in another part: either you have to learn complex verbal morphology, or you have to learn a bunch of separate words. Either way, what you learn takes about as much effort' (380) so that 'complexity in one subsystem should compensate for simplicity in the other' (383), and vice-versa.

However, the task is greatly simplified when we consider simple signals versus complex signals. While it may be difficult to compare the relative complexity of a particular language as a whole—say, English—with another—say, French—it is relatively straightforward, given the basic nature of the simplest signals in the simplest signalling game, to say whether a different type of signalling behaviour is more or less complex. Thus, functionally compositional signals are going to be more complex than simple atomic signals. So too, signals that are concatenated to greater length are going to be more complex than their atomic constituents—AB is 'more complex' in this sense than either A or B alone.

The complexity of signals might also be fruitfully understood in terms of information theory, though Hurford (2012) points out that '[l]inguists typically don't mention Information Theory in the vein of Shannon and Weaver (1949) ... the intuition is fundamentally information-theoretic' (385-6).[51] Whether or not human languages can be compared using information-theoretic notions, such as bit-complexity, simple communication systems certainly can.[52]

Furthermore, though full-blooded compositionality might not be the most appropriate avenue for discovering the evolution of language out of communication, recursion and composition may still have some part to play in our explanation. We might take, as a simple *idea* (not a definition, per se) of recursion, the following: any word (signal) that is combined with something that has already been built up by combining words (signals), counts as recursion.[53]

[51]Different evaluation metrics might appeal to Bayesian inference, Kolmogorov complexity, or minimal description length. See Rissanen (1978, 1989).

[52]See, Skyrms (2010a), Skyrms (2010b), LaCroix (2020a).

[53]See Nevins et al. (2009) and the discussion in Hurford (2012, Sec. 5.4).

Even this relatively straightforward or simple conception of recursion is fruitful to the extent that it encompasses the recursiveness of human-level languages, including fringe cases like Pirahã, while still capturing something that animals *cannot* do—consistent with the claims of Hauser et al. (2002), among others.

2.4.1 Function Words and Universals

In the case of language, function words do not generally arise in a new language until after a foundation of expressions with juxtaposed content words has been built up (for example, in pidgin or creole languages).[54] Hurford (2012) puts the point nicely: '[t]he meanings of *star* or *twinkle* can be demonstrated ostensively or paraphrased with other words. But what are the "meanings" of *of*, *the*, *are*, *what*, and *but*?' (326). Further, the fact that such a question is comprehensible to any speaker of English seems to strongly imply that these words must help to convey meaning in *some* way, even though the meaning cannot be pointed to or paraphrased (easily).[55] Hurford (2012) points out that function words can be distinguished from content words across three different dimensions:

1. The central uses of function words are to signal pragmatic force and aspects of grammatical structure, whereas content words, as the label implies, contribute mainly to the propositional content of a sentence.

2. The classes of function words are minimal sets, unlike the classes of content words, for example, sets of nouns or verbs, which are practically open-ended.

3. Function words are phonetically reduced in many ways.[56] (328)

[54]Function words, to be contrasted with *content words* are items belonging to a (small) closed class of language which generally indicate grammatical function. These might be conjunctions, determiners, or auxiliary verbs. Content words, on the other hand, are nouns, verbs, and adjectives.

[55]In the linguistics literature, 'function words' may also be referred to as 'functional items', 'closed class items', or 'grammatical items'. The function/content demarcation was present in Aristotle.

[56]See Shi (1996) and Shi et al. (1998).

As a result, function words do not appear in new languages until a solid base of expressions with juxtaposed content words has been built up—for example, creoles typically have fewer function words than non-creoles; creoles which evolved from pidgins usually have fewer function words in their earlier stages; and pidgins generally lack function words altogether (Siegel, 2008; Bickerton, 1999).

In the several attempts to model compositional signalling that we have seen, we begin with extremely basic reinforcement learning. Nowak and Krakauer (1999) show how compositional signalling might make it easier to arrive at a signalling system—it was said that they assume too much, but this is perhaps a virtue. Barrett (2006, 2007, 2009) shows how syntactically complex signals might arise in a basic reinforcement-learning structure. This accounts for how perfect communication can occur in situations where there are fewer signals than state-act pairs.

Having fewer signals than states of the world and appropriate actions is almost certainly going to be more representative of real-world evolutionary environments than assuming symmetry in the dimensionality of the signalling game, so this contribution, whether or not it produces bona fide compositionality, is important in its own right. Indeed, in such cases where there are informational bottlenecks, the system tends to naturally favour the outcome that allows for the highest level of information transfer.[57] That being said, there are further complications that may help or hinder information transfer and evolution toward signalling systems, as we have seen.

Thus, the analysis of the evolution of compositionality from the very few assumptions of the reinforcement-learning model of signalling are likely going to be too simplistic—compositional signalling does not arise in these simple cases, but it also does not arise in nature. Steinert-Threlkeld (2020) proves that the following assumptions jointly imply *trivial* compositionality:

[57]See LaCroix (2020a); Barrett and LaCroix (2020) for further details.

(A1) Agents communicate about a fixed set of states.

(A2) Optimal communication consists in correctly identifying the true member of the state space.

(A3) Messages are fixed-length sequences of signals from fixed sets.

Where compositionality is trivial just in case complex expressions are always interpreted as intersection of the parts—i.e., *brown dog* is the intersection of *brown* and *dog*.[58] The story of how compositional signalling gets off the ground from simple signals is a necessary one to tell, but it is going, in all likelihood, to be one of the later developments. Even so, there are good arguments that certain features of language evolved later than other, simpler features—again, complex language must evolve from simpler languages. Though few things are universal to *every* human language, there are many universal *implication generalisations*: if language X has feature Y, then language X has feature Z.

For example, Greenberg (1963, Universal 34) posits that '[n]o language has a trial number unless it has a dual. No language has a dual unless it has a plural'. The evolutionary interpretation is that the plural is evolutionarily prior to the dual: a language that contains duals can only evolve out of a language that already contains plurals—therefore, plurals are upstream, as it were from duals.[59] The idea, therefore, is that languages include layers of features which leave some impression of their evolutionary priority.[60] For example,

> In each language, we find vestigial one-word expressions and proto-syntactic (2-, 3-word) constructions keeping company with more fully elaborate syntax. Most languages have the possibility of conveying propositional information without the benefit of syntax. English speakers use a single word, *Yes* or *No* and pragmatic inference identifies the particular proposition which is being confirmed or denied. Few languages lack such devices. (Hurford, 2012, 376-7)

[58] This will be discussed in more detail in Chapter 4.

[59] Hurford (2009) discusses the fact that not all implicational universals have such an evolutionary interpretation.

[60] For a nice example, Hurford (1987) discusses the evolution of numeral systems. See also, Hopper (1991), Nichols (1992), and Fennell (2001).

In a similar example, we note that the indefinite article is often derivative of the numeral 1—for example, 'un' and 'une' in French, or 'eins', 'ein', 'eine' in German. Further, Hurford (1987, 2012) points out that it takes a while for a language to develop a numeral system. So, if numeral systems are relatively late-developing, and indefinite articles are derivative of a numeral system containing 'one', then we should expect a definite article to arise earlier in the evolutionary stage. Indeed, Dryer (2008) surveyed 473 modern languages and reports that 81 of them have a definite article, but no indefinite article, whereas the reverse is never true. This is consistent with Hurford (2012).

Presumably, simple constructions of this sort would have been the first to evolve—as we have seen, simple signals corresponding to single units of propositional content are standard in animal communication systems, and the theory of signalling games tells us how such things might arise.

> Putting words together without any explicit marker of their semantic relation to each other (as in word soup) is, we can reasonably suppose, a primitive type of syntax. Grammatical constructions with dedicated function words indicating how they are to be interpreted came later. Mere juxtaposition is found in compound nouns, like boy wonder, village chief, and lion cub. (Hurford, 2012, 374)

Though it is trivial *qua* compositionality, there is something to be said for 'mere' concatenation, in the sense of compounding by juxtaposition, when we consider the evolution of language in this way: Jackendoff (2002) points out that '[t]he facts of compounding thus seem symptomatic of protolinguistic "fossils": the grammatical principle involved is simply one of concatenating two nouns into a bigger noun, and the semantic relation between them is determined by a combination of pragmatics and memorization' (250).[61]

[61] The idea of 'living fossils' of language was first introduced by Bickerton (1990, 1999). See also, Jackendoff (1999, 2002). A 'living fossil' in biology denotes species that have changed very little from their fossil ancestors, such as the lungfish—See Ridley (1993).

Progovac (2009a, 209) offers the following such examples of compounding by juxtaposition: 'Him retire!?', 'John a doctor?!', and 'Her happy?!' (interrogatives); 'Me first!', 'Family first!', and 'Everybody out!' (imperatives); and 'Class in session', 'Problem solved', 'Case closed', and 'Me in Rome' (declaratives). These examples consist of 'Root Small Clauses', which lack (for example) tense-markers and subject-verb agreement. Her point is that root small clauses of this sort 'instantiate/approximate a grammar of an earlier stage of syntax, protosyntax, which was measurably simpler' (Progovac, 2009a, 209). Indeed, a necessary condition for something to be a (syntactic) linguistic fossil is that 'it has to be theoretically proven to be measurably simpler than its more complex/more modern counterparts, and yet show clear continuity with them' (Progovac, 2015, 3). This sort of 'layering' is common to all evolved systems (Progovac, 2009a, 2015), including the brain 'with the recently evolved neocortex co-existing with more ancient diencephalon and basal ganglia' (Hurford, 2012, 378), and cities, for example.[62]

Note that the focus on function words, and thus syntax, highlights *only* the internal structure—i.e., the linguistic components—of language origins. However, it is necessary to account for (or at least, to not ignore) several *external* components that are thought to be relevant to the evolution of complex communication and language.

2.4.2 Biological Components

Thus, there is a further biological component that might play into the story of the evolution of language. For example, compared to our closest relatives, the base position of the human larynx is significantly lower in the throat, with the pharynx above it. This feature of human anatomy has been widely discussed, and many have argued that the modern human vocal

[62] 'Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses' (Wittgenstein, 1953, 18).

tract is necessary for the production of speech.[63] Also different is the position of the velum—separating the oral and nasal cavities—and the placement and shape of the tongue body as compared with the cavities of the vocal tract. All of this together means that humans are capable of producing significantly varied and more numerous sounds than other primates. The wide variety of sounds are helped by the pharynx being easily manipulated—most vowel sounds ([a], [i], [u]) and many velar consonant sounds ([k], [g]) are beyond the physical capacity of, e.g., chimpanzees. The position of the velum, in addition, makes it physically impossible for chimpanzees to make distinct nasal and non-nasal sounds (e.g., [m] compared to [b]).[64]

2.4.3 Cognitive Components

In addition to physical components that may affect the ability to produce language, there are cognitive components that may play a significant role. Tulving and Markowitsch (1998) point out that '[m]any animals other than humans, especially mammals and birds, possess well-developed knowledge-of-the-world (declarative memory) systems, and are capable of acquiring vast amounts of flexibly expressible information' (202). Given that animals can remember, Hurford (2007) points out that they can be in mental states relating to the past.[65] Even more impressive, as we have seen, honeybees can remember and code for where they found nectar (von Frisch, 1967).

Some species of monkeys apparently have a rudimentary hierarchy in their conceptual knowledge about the world to the extent that they can distinguish a category, FOOD, from other

[63]See, Lieberman et al. (1972); Laitman and Reidenberg (1988); Donald (1991); Pinker (1994); Carstairs-McCarthy (1998), as well as the discussion in Fitch (2010, Ch. 8). Note that one might argue that such an apparatus was selected for ease of communication; however, there is a clear selective disadvantage, since a lowered larynx increases the likelihood of choking to death (Darwin, 1859)—unless of course, complex sound production offered some advantages over and above the increased risks.

[64]However, chimpanzees can eat or drink and breathe at the same time.

[65]For empirical studies that examine several species of bird with resepct to memory of, e.g., hidden food, see Biebach et al. (1989), Healy and Suhonen (1996), Clayton et al. (1997), and Griffiths et al. (1999).

categories, and within that category can further differentiate between high- and low-quality food, based on colour (Hauser, 1998; Hauser and Marler, 1993). This hierarchy translates, semantically, to their different calls. In the view of Dummett (1993a), animals are capable of having *proto-thoughts*, but they are not capable of considering propositions, which are associated with language. However, Hurford (2007, Ch. 4) argues that animals are capable of *proposition-like* cognition.[66]

2.4.4 Social Components

In addition to biological and cognitive constraints, there are social considerations to be taken into account. Studies of *creolization*—the evolution of a creole language out of a pidgin communication system—show that isolated individuals, on their own, do not spontaneously create new languages; instead, a social 'critical mass' is necessary for the emergence of a creole language (Hall Jr., 1966; Mühlhäusler, 1997). Once such a community is formed, however, syntactically sophisticated languages can develop quite rapidly (Kegl, 2002; Senghas and Coppola, 2001; Senghas et al., 2005).[67] Though biological factors undoubtedly constrain the instinct to learn language, Fitch (2010) suggests that cultural transmission, along with a need to communicate, plays an essential role in triggering the biological capacity.[68]

Humans have certainly learned the (cultural-dependent) ability to suppress laughter or crying in socially inappropriate situations. Similarly, while some signalling in animal communication systems might be innate, *and* have an automatic response, some animals can learn to suppress this behaviour. With proper training, several species can inhibit (or initiate) the production of signals, though the structure itself is innate. This includes many mammals, such as cats,

[66]He admits that conceding this point depends largely upon how one defines a proposition in the first place. It is always possible to define the word so broadly that it includes animal cognition.

[67]Note that, whereas pidgins lack syntactic complexity and are generally not considered languages in the linguistic community, Creoles are *bona fide* languages in the sense that they exhibit the same syntactic complexity (e.g., phrase structure, function words, negation, quantification, etc.) as other human languages.

[68]See also, Bickerton (1981); Singh (2000); Mufwene (2001).

dogs, and guinea pigs (Myers, 1976; Adret, 1993), as well as all primates thus tested—e.g., lemurs (Wilson Jr., 1975), capuchin monkeys (Myers et al., 1965), rhesus monkeys (Sutton et al., 1973; Aitken and Wilson Jr., 1979), and chimpanzees (Randolph and Brooks, 1967). Though withholding is difficult for animals in the wild, it is still possible. For example, Goodall (1986) and Townsend et al. (2008) have observed wild chimpanzees cover their mouths to avoid vocalisation.

Similarly, there is evidence against solely reflexive vocalisation due to *audience effects*—several mammals will not produce alarm calls when no conspecifics are present, as in vervet monkeys (Cheney and Seyfarth, 1985). Similarly, ground squirrels will only produce calls when kin are present (Sherman, 1977). At an even higher level of sophistication, male chickens appear to be more likely to produce alarm calls when with their mate, chicks, or other familiar birds are present than when unfamiliar birds are present, and they only produce food calls when hens are present, not when alone with another cockerel (Marler et al., 1991; Evans and Marler, 1994; Evans and Evans, 2007).

Most aspects of complex behaviour arise from a combination of environmental inputs subject to genetically based constraints. The language capacity in humans might be appropriately described as an *instinct* to learn (Marler, 1991). Further, it is well known that apes in a lab setting can learn (at least) 125 distinct referential signals; however, nothing close to this lexicon size has been observed in a natural environment—this is a difference by order of magnitude. This strongly implies that apes have a *latent*, though unexpressed, *cognitive capacity* to acquire a reasonably large lexicon. Further, Fitch (2010) points out that, since this ability would have been present in the last common ancestor of humans and chimpanzees, this fact is relevant to models of language evolution to the extent that 'any mutations that

increased referential signal *production* would already have found *listeners* able to make sense of these signals' (165).[69]

In so-called 'Kaspar Hauser'[70] experiments, young squirrel monkeys are raised by muted mothers, and so are not able to hear, and so are not able to learn prototypical conspecific vocalisations. In spite of this, such squirrel monkeys will produce the full range of the calls of their species, and in the appropriate contexts (Winter et al., 1973). Similarly, Owren et al. (1993) show how 'cross-fostered' macaques that are raised among a different species will learn to interpret and respond appropriately to the calls of the parent-species but will nonetheless produce calls that are typical of their own species.

Hurford (2007) suggests that the crucial precursor for the jump from proto-linguistic to linguistic abilities was not necessarily a linguistic change, but rather a change in the social relationships between groups:

> clear and strong relationships between social group size, grooming time and vocal repertoire size have emerged in our analyses. Independent contrast analyses revealed that evolutionary changes in repertoire are a strong predictor of both changes in group size and changes in grooming time among non-human primates. . . . [Though the direction of causality cannot be inferred from correlational analyses,] our findings are consistent with the hypothesis that the vocal communication system may facilitate (or constrain) increases in group size and levels of social bonding within primate social groups. (McComb and Semple, 2005, 3)

The species with the most extensive vocal repertoire (38 signals) also happened to have the most abundant groups (approximately 125 individuals, on average).[71]

[69]Note that dogs and parrots are also capable of acquiring large vocabularies, on a par with trained primates. Thus, this is not a primate-specific cognitive ability. See Pepperberg (1990); Kaminski et al. (2004).

[70]Kaspar Hauser was a young German who lived in the early 19th century, until he was murdered in 1933. From a young age, it is alleged, he was held in a dungeon with no social contact. When he was discovered, he was only able to say his name and the phrase 'I want to be a horseman like my father'.

[71]Data can be found in Bermejo and Omedes (1999).

Therefore, an explanation that accounts solely for syntax is going to be inadequate. We will additionally require some consideration of the effects of social structure and cognition on the evolution of language. Syntax, as an explanatory target, is too self-contained.

2.5 Summary

There is some difference between human languages and communication systems that arise in nature. This leads naturally to the question, elucidated in Wittgenstein (1953): 'Where do we draw the lines between human and the rest of the world? Humans have a set of behaviors and a language that seems to differentiate them from the animal world. Is this a difference in complexity, in self-awareness, or some other unnamed quality?' (281). In general, trying to find necessary or sufficient conditions for clearly demarcating language from communication leads to conceptual and practical difficulties. However, it is informative to examine—i.e., to find a specific target for our explanation—what are often taken to be the features of communication systems that are unique to linguistic systems of communication.

Fitch (2010) defines language as 'a system which bi-directionally maps an open-ended set of concepts onto an open-ended set of signals' (173). In a similar vein, Berwick and Chomsky (2016) suggest the following 'basic' property of language: '[a] language is a finite computational system yielding an infinity of expressions, each of which has a definite interpretation in semantic-pragmatic and sensorimotor systems' (1). As we have seen, the open-endedness of language is often cited as *the* main differentiating feature between human languages and animal communication systems.[72] The prototypical instantiation of this open-endedness is

[72]Fitch (2010) further points out that there is no logical or empirical reason to assume that language *must* have evolved from a preexisting communication system (i.e., one which was present in the LCA)—citing the fact that no known animal communication system can be considered a language in the above sense. That is to say, novel aspects of human language might have evolved from *cognitive* as opposed to linguistic precursors. However, some communication abilities do appear to provide *potential* precursors to mechanisms that are involved in languages. For example, the ability to interpret signals as meaningful is likely a necessary condition for language to arise, whereas innately learned vocalisations may not be necessary.

compositionality. Several authors argue that compositionality is a uniquely human communicative capacity. For example: '[n]atural lexicoding [compositionality] appears to be a purely human phenomenon. The only animals that do anything remotely similar have been tutored by humans' (Marler, 1998, 11); 'there is no compelling evidence for any semantically compositional learned signalling in wild animals' (Hurford, 2012, 18), and further that '[h]uman language is a unique naturally occurring case of learned and arbitrary symbolic communication, about objects and events in a shared external world' (Hurford, 2007, 184).

However, we also saw that, when it comes to trying to model an extended signalling game that accounts for how compositionality might arise, these models tend not to be sensitive to empirical data. Thus, requiring an explanation of compositional *syntax* will be necessary, but further down the road of the evolutionary story. We should not expect a minimal model, such as reinforcement learning, to give rise to compositionality while at the same time noting that compositionality is extremely rare in nature.

What we have seen up to this point is that communication abounds in nature. Furthermore, some communication systems are quite complex. The difference between communication and language was assumed to be a difference in degree, rather than kind. Thus, what we require is an explanation of how *complexity* can arise out of simple signalling systems, rather than linguistic compositionality *per se*. Shifting the focus from compositionality to complexity requires shifting the focus of our evolutionary models: rather than trying to explain how complex signals arise, we should examine how complex *structures* might naturally occur. This complexity may well give way to compositional communication downstream.

This is still a bit of a simplification, as we have seen. Minimally, the evolution of language capacity in humans will be a complex process, involving the *co-evolution* of both speech-production (which requires biological evolution of the requisite anatomy for producing the range of sounds which humans can produce) and speech perception—the fact that this requires a co-evolutionary process is evident from the fact that production and perception

of language utilise separate macro-mechanisms in the brain—primarily Broca's area, in the former case, and mainly Wernicke's area in the latter case. However, though Broca's area is implicated in motor function and thus voluntary control of vocal production in humans, lesion studies in monkey vocalisations show now corresponding function (Jürgens, 2002).[73] Nonetheless, Hurford (2012) argues that

> no phenomena, either in apes or in human non-linguistic behaviour, resemble our extremely fluent and flexible combination of constructions closely enough to suggest any obviously plausible pre-existing evolutionary platform. Logically, there had to be a pre-existing platform, somewhere between Australopithecines and modern humans, but it has left no traces. (519)

Here, co-evolution might be taken in a broad sense in line with Godfrey-Smith (2018), referring to the behaviours *within* a species, as well as *between* them, or potentially within the same agent.

Though the main content of this chapter has circled around a negative argument—that compositional syntax in the wrong explanatory target for language origins research—the subsequent chapter provides a positive account for an alternative explanatory target—reflexivity—which is also a unique property of language.

[73]Note that production requires different physical mechanisms in the sense that we need to access, e.g., motor mechanisms to produce sound, whereas perception requires auditory or visual processing. Broca's area is proximal to the motor cortex, whereas Wernicke's area is located between the auditory and visual cortices.

Chapter 3

Communication and Modularity

> *If as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible for them.*
>
> — Genesis, *11:6*

In Chapter 2, I surveyed some key distinctions between communication and language to get clear on the relevant phenomena that we ought to take as our target(s) of inquiry in studying the evolution of language out of simple communication. It was said that one salient feature of human language that is lacking in (almost) all animal communication systems is *compositionality*. Compositionality is supposed to explain the *generative* features of language that allow for arbitrary specificity and that give rise to a potentially infinite number of unique expressions. Because systems of animal communication lack this generative capacity—at least in a nontrivial way—an evolutionary explanation of compositionality must explain how compositional communication possibly evolved from non-compositional communication.

However, using the signalling-game framework (see Chapter 1) to explain the evolution of complex communicative dispositions, I suggested that the models that have been proposed to date—i.e., those focusing on how, what we might call, *linguistic* compositionality evolves—

fall short of this explanatory goal. In particular, the evolution of linguistic compositionality, it was suggested in Chapter 2, is the wrong way to think about how compositional communication in fact evolved, given that there is no empirical evidence for any natural precursor to human language.

However, this does not negate the requirement for an evolutionary explanation: given that compositional communication evolved, there must have been some mechanism by which it evolved. As an alternative explanation, I will suggest in this chapter a variety of ways in which simple communication systems themselves might compose to create more complex systems. Communication is a unique evolutionary process in the following sense: once a group of individuals has learned some set of simple communication conventions, those learned behaviours may be used to influence future communicative behaviours, giving rise to a feedback loop. When faced with a novel context, an individual can always evolve a brand-new disposition. However, the individual may learn to take advantage of previously evolved dispositions. (In Part II, I propose several concrete ways that this might happen.) When the contexts are similar in the relevant ways, the individual may *appropriate*, or transfer knowledge from, her previous disposition for use in the new context. Or, if the individual has learned several independent dispositions, she might learn to combine them in suitable ways for the new setting.

Indeed, individuals may learn to take advantage of pre-evolved *communicative* dispositions to thereby influence the evolution of future communication; this is a conception of *reflexivity*, as an evolutionary mechanism, which I examine. Crucially, after signals have become functionally referential, they may come to refer to other signals or systems of communication. This reflexive ability may in turn lead to complex, compositional *structures*.

Thus, I suggest that it is the *reflexivity* of language that provides an apt explanatory target for an evolutionary account. Reflexivity, I will argue, depends inherently upon a notion of *modular composition*. This is a more general notion of compositionality wherein the units of

composition need not themselves be linguistic: they may refer to the composition of, e.g., cognitive or structural modules. A notion of modularity arises in a variety of fields, including biology, linguistics, and cognitive science.[1] As such, it will be essential to remain clear on what the units of composition are and how they evolve.

This position challenges the existing paradigm (at least in work on the evolution of language that arises in linguistics), which focuses on the evolution of syntax. Jackendoff (2007) refers to this tendency as *syntactocentrism* and argues that it was a 'scientific mistake' (Sec. 2.4). This bias toward syntax arises from Chomsky (1957, 1965), who showed that language requires a generative system that makes an unlimited variety of sentences possible. Even today, this paradigm has deep roots. However, Chomsky *assumes* without explicit argument that the generativity of language arises entirely because of the syntactic component of grammar. On this account, the combinatorial properties of phonology and semantics are strictly derivative of the combinatorial properties of syntax.

We have already seen how semantic properties arise in nature, generally prior to syntax. Thus, the position I take is the reverse of the 20th-century paradigm in linguistics: semantics takes priority over syntax. Given that simple semantics can arise easily in nature, the goal is then to explain how more sophisticated communication systems might arise out of these (syntactically) simple, albeit semantically rich, communication systems. Of course, one might argue that semantics is 'prior' (in a trivial way) to syntax even on Chomsky's view, since the precursors to the single mutation leading to language in humans would have included, for example, simple semantics. That is, the faculty of language in the broad sense includes the conceptual-intentional and sensori-motor systems that underlie, e.g., vocal imitation and invention or referential vocal signals. Nonetheless, on the view that reflexivity plays a crucial role in the 'complexification' of simple signalling systems, the semantic content of signals that come to be reflexive is the driver of this process. Therefore, semantic content is not just

[1] Fitch (2010) suggests that, because each of these disciplines have a different interpretation, the idea of modularity is ambiguous. Perhaps a better notion is that of *encapsulation*, in the sense of Fodor (1987).

a pre-adaptation, but it plays a pivotal role in the evolution of complex communication systems (leading to language).

This chapter will proceed as follows to finish laying the philosophical groundwork for the empirical contribution of this dissertation, offered in Part II. I will begin in Section 3.1 by surveying a recent suggestion (Barrett and Skyrms, 2017) that takes modular composition into account. The purpose here is to make explicit what underlies this model. The process by which pre-evolved dispositions might be used in a novel context vary; I analyse several of these compositional processes. Section 3.2.1, 3.2.2, and 3.2.3 discuss notions of *transfer learning*, *analogical reasoning*, and *modular composition* concerning how they might affect the evolution of complex communication. This shows that reflexivity, via modular composition, has plausible empirical precursors—unlike linguistic compositionality—and this provides the sort of *graded* distinction that is required for a gradualist perspective. Section 3.3.1 further highlights the close relationship between language and modular composition, and Section 3.3.2 highlights the same for hierarchical social structures. This reinforces what has already been suggested: that a satisfactory account of how language might have evolved out of simpler mechanisms will require a treatment of cognitive and social structures that co-evolved along with linguistic capacities. I demonstrate that modular composition is a common link between all of these processes, and thus provides a unifying framework for talking about all of them, at a certain level of generality. Finally, Section 3.4 summarises my position concerning a plausible explanation of how to learn complex signalling dispositions out of simple ones.

This chapter will remain as general as possible, in an attempt to suggest a reorientation in how we think of language (as a complex system of communication) as having evolved from simpler systems. Part II goes into more detail about the specifics of some of these possible mechanisms and applies them to several models for the evolution of complex communication.

3.1 Self-Assembly and Modular Composition

The examples of signalling-game contexts that we have seen thus far have been relatively static. Namely, a signalling context is given, with its underlying structure—including the states of the world, the possible dispositions, etc. Concerning *linguistic* compositionality, these assumptions themselves entail a necessarily 'trivial' notion of compositionality wherein complex expressions are always interpreted by the intersection (or *generalised conjunction*) of the meanings of the parts of the expression (Steinert-Threlkeld, 2018, 2020).[2] However, games themselves may compose in some way to form more complex games. This is the notion of *self-assembly* suggested in Barrett and Skyrms (2017).

They highlight how the very structure of the game might evolve, and the dispositions of the agents might be appropriated to accommodate new tasks in response to the evolution of the game structure. This might give rise to complex dispositions in the same way that modular composition of simple signals might give rise to complex signals, or modular composition of simple core cognitive modules might give rise to more complex cognitive processes.

In the linguistic context of a signalling game, we might ask how pre-evolved *communicative* dispositions might compose to form more complex communication *systems*. One key idea that I want to advance here is that *context* matters for the evolution of complex communication. As with language, social structures—including cooperation, competition, dominance, etc.—must be integrated dynamically into an individual's understanding of a particular situation from moment to moment: the way an agent should respond to a specific situation, when her

[2]For example, 'brown' might evolve to partition nature, conceptually, by picking out all and only brown things; and 'dog' might similarly evolve to pick out all and only dogs. 'Brown dog' is (at least syntactically) compositional. However, it only serves to pick out the intersection of the sets of concepts picked out by each of the terms individually.

opponent is an in-group member versus an out-group member, for example, might change the appropriateness of her response. This is well documented in animal communication.[3]

Goffman (1974) refers to this as *framing*. The basic idea is that the same event might have different significance for different individuals depending upon the *frame* (or context) in which it is interpreted. Searle (1995) and Bratman (1999) point out that though many games employ a significant competitive element (e.g., trying to win), they are contextualised within a cooperative frame—i.e., agreeing to play, abiding by a given set of rules, assuming fairness on the part of the opponent, etc. Similarly, cooperative bargaining might be framed within a context of competition—an individual wants the best deal for herself (Jackendoff, 2007).

The discussion in Barrett and Skyrms (2017) is quite subtle, and so it is worth going over in a bit of detail. We will then see how this general model of self-assembly and modular composition gives rise to a rich set of processes by which agents might learn or evolve complex dispositions. Furthermore, this set of processes is more appropriate than linguistic compositionality as a gap-bridging explanation of how complex communication might arise out of simple communication, to the extent that there is good empirical evidence that animals do take advantage of (at least some of) these processes. Finally, it provides a unifying framework for a range of mutually co-dependent phenomena that likely would have been required to evolve or learn complex communicative, and even linguistic, dispositions.

3.1.1 Cue-Reading and Sensory Manipulation

Barrett and Skyrms (2017) argue that signalling games might evolve in a variety of ways. In the process of ritualisation (Tinbergen, 1952; Lorenz, 1966; Huxley, 1966), individuals

[3]See, for example, Keller and Ross (1998); Kutsukake et al. (2006); Lusseau et al. (2006); Sinervo et al. (2006); Gardner and West (2010); Madden et al. (2011), as well as the discussion in Fu et al. (2012); Masuda and Fu (2015).

might have a fixed disposition. On the one hand, a receiver may have a fixed strategy—for example, some innate response to a particular stimulus. In this case, it is the sender's job to determine what the right action is to stimulate that response. An example of this is the *Physalaemus pustulosus* species group of frogs (Ryan and Rand, 1993). Barrett and Skyrms (2017) point out that the females of this species are attracted to a particular type of sound—this is an innate or fixed disposition to respond to a specific stimulus in a specific way. Thus, the males must learn how to send the appropriate stimulus to elicit the 'desired' response. In this sense, the male is exploiting a pre-existing disposition on the part of the female. Barrett and Skyrms (2017) highlight that a game may be formed over time by the process of ritualisation via a positive feedback mechanism—i.e., some dynamical process.[4] This constitutes a 'sensory-manipulation game', wherein the receiver's dispositional responses to an incoming signal are fixed, and the sender must learn which signal will offer the correct stimulus to elicit the appropriate response.

Conversely, when an agent must evolve or learn to *respond* to some fixed set of dispositions—either those of nature or those of the signaller, for example—we have a 'cue-reading game'. This terminology is consistent with the distinction between signals and cues suggested by Maynard Smith and Harper (2003). In their vernacular, a signal is 'any act or structure which alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved' (3). A cue, on the other hand, can be used to guide an individual's actions—the receiver of the cue—although the cue itself did not evolve *because of* the receiver's receiving the cue. Instead, a cue is a byproduct of some other process, which ends up being beneficial to some individual capable of reading the cue. For example, the presence of CO_2 is a cue to the location of a mammal, though it is not a signal that the mammal sends. The presence of CO_2 near the location of a mammal is the byproduct of some *other* process; however, mosquitoes have evolved to receive or 'interpret' these cues appropriately and perform a beneficial act (at least to itself).

[4]See also Endler (1993); Dawkins and Guilford (1996).

Thus, signals are a *subset* of cues: a cue is any aspect that may be utilised by a receiver, whereas a signal is a cue which is emitted by a signaller *because* it affects the receiver's behaviour. When there is a direct causal link between some important characteristics of an individual and some specific aspects of the signals that encode that information, the signals are referred to as *indices*. Namely, an index is a signal that cannot be deceptive, due precisely to the causal relationship between the form of the signal and the characteristic of which it is a signal. For example, formant frequencies are often honest indicators of body size in a range of species (Fitch, 1997; Riede and Fitch, 1999; Reby and McComb, 2003; Rendall et al., 2005)—this is an index because there is a direct causal relation between, e.g., the roar of a red deer and its size (Reby et al., 2005).

A specific cue-reading game, with the sender's dispositions being fixed, is given in Figure 3.1.

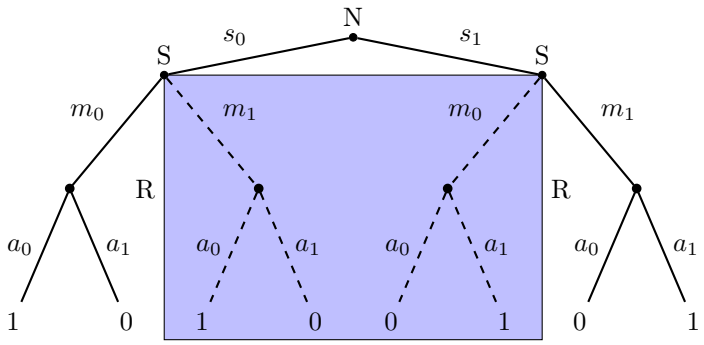


Figure 3.1: The extensive form of a simple cue-reading game, modelled as a (pruned) subgame of the simple 2×2 signalling game.

The game tree has been *pruned* to exclude any choice on the part of the sender. Thus, the sender *always* sends m_0 in s_0 and always sends m_1 in s_1. However, the basic idea of the Lewis signalling game is still present: the sender has some information (which state obtains) which the receiver does not have. Thus, the sender's dispositions are static in some sense, and the receiver must learn to read the appropriate cues given.

Similarly, a specific sensory-manipulation game, with the receiver's dispositions being fixed, is given in Figure 3.2.

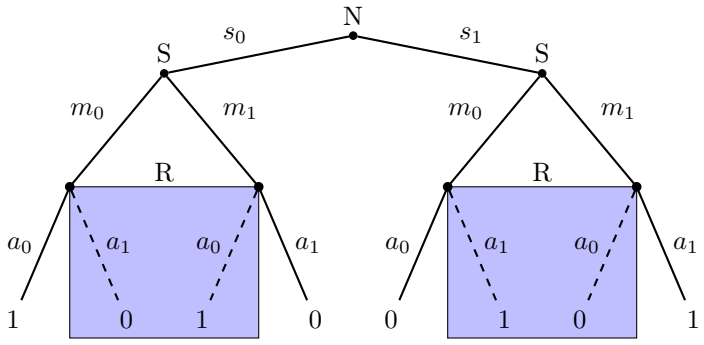


Figure 3.2: The extensive form of a simple sensory-manipulation game, modelled as a (pruned) sub-game of the simple 2×2 signalling game.

Here, the game tree has been *pruned* to exclude any choice on the part of the receiver. Thus, the receiver *always* responds to m_0 with a_0 and always responds to m_1 with a_1. As a result, the receiver's dispositions are static, and the sender must learn which stimulus to provide to elicit an appropriate response.

The sensory manipulation game and the cue-reading game, individually, highlight different facets of the same phenomenon. Namely, two kinds of behaviour are relevant to communication—production of meaningful signs on the one hand, and the interpretation of meaningful signs on the other. When communication is successful, we have seen, the production and interpretation of arbitrary signs 'fit' together in some sense. Similarly, a failure to communicate arises from a failure to fit dispositions together in an appropriate way. When a communication system has become firmly established, the particular signals that a sender uses are the ones that have been conditioned by the receiver's patterns of interpretation occurring downstream. Similarly, patterns of interpretation on the receiver's part are conditioned by the sender's production of particular signals occurring upstream. As such, interpretation and production co-evolve, and the theory of self-assembling games demarcates these two processes—in a way that the full signalling game does not—to show *how* these might co-evolve.

When cue-reading and sensory-manipulation co-evolve, so that neither disposition is fixed, we have a signalling game, as in Figure 1.2 (Chapter 1). Each of these may arise through a process of ritualisation, and this ritualisation process, Barrett and Skyrms (2017) suggest, might serve as 'the glue that binds agents to form simple games from their basic decisions, then increasingly complex games from simple games' (335).

3.1.2 Template Transfer

Though ritualisation might serve as a binding mechanism for building complex games out of simple ones, we might ask how such a process naturally evolves. The idea of template transfer is that a game which develops in a particular context might come to be used successfully in a novel context. This is a form of appropriation of strategy or transfer of learning. In the signalling-game framework, supposing that the sender and receiver have arrived at a signalling system for one particular context, their joint strategies constitute a stable mapping from states to signals to actions. Template transfer might occur when, given a stable disposition for a particular context, the stable strategy is appropriated for use in a context that differs from the original context for which that particular rule was evolved. Barrett and Skyrms (2017) suggest that '[i]n many cases the appropriation of an old rule to a new context may be significantly more efficient than evolving a new rule from scratch' (337).

In transitive inference tasks, an animal must learn an arbitrary association between, e.g., the colour of a container and the relative amount of food in it. They are given pairwise comparisons (e.g., the green container has more food than the red container, and the red container has more food than the blue container), and then given a novel pairing (e.g., green and blue). Spontaneous correct choices in these cases are taken as evidence for transitive inference, and several diverse species—rats, birds, squirrel monkeys, and chimpanzees, for

example—perform successfully on such tasks (McGonigle and Chalmers, 1977; Gillan, 1981; Davis, 1992; Paz-y-Miño et al., 2004).

An example of template transfer, discussed at length in Barrett (2013, 2014, 2018); Barrett and Skyrms (2017), is the transitive inference behaviour of Pinyon Jays (*Gymnorhinus cyanocephalus*) and Scrub Jays (*Aphelocoma californica*), as reported by Bond et al. (2003). On this experimental setup, each bird is given a linear ordering of seven stimulus colours—that is, the linear order is fixed for each particular bird. The birds are initially presented with one of 6 *adjacent* pairs of colours. By analogy, suppose the colours are represented by a linear ordering, $\{1, 2, 3, 4, 5, 6, 7\}$. On one round, a pair might consist of $\{1, 2\}$, $\{5, 6\}$, $\{2, 3\}$, etc., with the position of the higher-ranked stimulus randomised between left and right on each particular trial. If the bird chooses the *higher* of the two colours in the ordering, then it is rewarded. These birds learned to select the higher-ranked colour with greater than 0.85 accuracy.

Next, the birds are presented with nonadjacent pairs of colours, in the same ordering, to determine whether they would use previously acquired knowledge of the adjacent pairing orderings to order the new pairs. Very quickly, the Pinyon Jays were able to choose the correct colour with an accuracy of 0.86 for non-adjacent pairs, and the Scrub Jays were able to do so with an accuracy of 0.77. Bond et al. (2003) conclude that they are performing some transitive inference here—i.e., when presented with non-adjacent pairs, $\{2, 4\}$ or $\{2, 6\}$ for example, the jays can take advantage of the previously acquired knowledge that, e.g., $3 > 2$ and $4 > 3$, which jointly implies the correct choice is '4', in the first case, by transitivity. Barrett and Skyrms (2017) claim that the birds are doing more than just transitive inference since the order on adjacent pairs of colours in no way implies or determines the order on non-adjacent pairs.

> It is only by appropriating a pre-existing linear template that the birds could get from their experience with adjacent colour pairs to judgments that immediately

> agreed with the experimenters predetermined full linear order. Indeed, that the experimenters themselves took the birds' judgments on non-adjacent colour pairs to be correct and simply inferential suggests that the experimenters were also appropriating a pre-existing linear template to their understanding of the birds' experience. (338)

Thus, these experiments seem to strongly suggest that in addition to *using* transitive inference, the jays were also appropriating prior knowledge of a linear ordering of adjacent pairs and imposing it on non-adjacent pairs. This last point is worth highlighting: in fact, when human subjects are required to learn pairs of 'nonsense' items (e.g., Japanese characters, for non-speakers of Japanese) while unaware that the pairs form an (implicit) ordered set, they tend to perform similar to nonhuman animals faced with the same task. As such, when explicit reasoning is not available to the human agent, implicit transitive inference is still achieved (Frank et al., 2005; Lazareva and Wasserman, 2010).

Importantly, for our purposes, this process can be modelled with a signalling game. Suppose there are two senders (which may be functional elements of a single individual) and one receiver. Call these σ_1, σ_2, and ρ, respectively. From a set of seven equiprobable stimuli, Nature chooses 2—these are the states s_{1i} and s_{2j}. The two senders react to the stimuli by sending a message to the receiver. The receiver has three possible actions at her disposal—a_1: $s_{1i} > s_{2j}$, a_2: $s_{1i} < s_{2j}$, and a_3: $s_{1i} = s_{2j}$. The action is a success just in case it matches the pre-determined linear ordering of states: $s_{-i} > s_{-j}$ if and only if $i > j$.

Barrett and Skyrms (2017) run simulations for this linear-ordering game, where the dynamic is simple reinforcement learning with invention, as was discussed in Chapter 1.[5] Thus, there is no fixed set of signals at the outset, but signals are invented as is necessary. The senders begin by signalling randomly, but by 10^7 plays, the communicative success rate is typically (0.99) better than 0.75. Barrett and Skyrms (2017) point out that this composite two-sender system might be thought of as 'having evolved to implement a dispositional rule that takes

[5]See also the discussion in Skyrms (2010a, Ch. 10); Alexander et al. (2012).

naturally ordered stimuli as input, represents the stimuli as signals, then outputs an act that reliably indicates the natural order of stimuli' (339). Further, once such a system is evolved, it might be appropriated to represent an ordering on novel (i.e., non-adjacent) stimuli. Thus, this is an implementation of template transfer:

> Such a template might be fit to a new context by coordinating the new stimuli with the old inputs to the dispositional rule. The association of the new stimuli with the old inputs to the dispositional rule might be thought of as implementing an analogy between the new and old stimuli. When such an analogy evolves, the old dispositional rule evolves to treat the new stimuli similarly to how it treated the old stimuli that were involved in forging the old dispositional rule. This sort of template transfer might be evolutionarily favoured when the process that coordinates the new stimuli to the old inputs is more efficient than evolving a new rule for the new context. (Barrett and Skyrms, 2017, 339)

They assume a pre-evolved, and fixed, ordering-template and show that, under reinforcement learning with punishment, the agents typically (0.995) evolve to successfully match the new stimuli to the old ordering system with a success rate better than 0.8 with 10^5 plays per run—several orders of magnitude faster than learning a disposition from scratch. This is because the receiver's dispositions are already well-tuned to make successful linear-ordering judgements; the senders just need to learn how to represent the analogy between the old stimuli and the new. They refer to this process as *polymerisation* and suggest that this is a special case of the more general process of *modular composition*.

3.1.3 Modular Composition

Modular composition is a process by which complex, composite games are formed—i.e., when one game evolves to accept the play of another game as input. For a particular case-study, Barrett and Skyrms (2017) examine a situation in which a two-sender addition game—where the players evolve to compute the sums of cardinalities presented to each sender—and a two-

sender ordering game—as was previously discussed—compose to form a complex game which compares the value of a sum with a single input.

In the addition game, under reinforcement with invention, they report that 'as the agents update their first-order dispositions by reinforcement, they typically evolve a set of systematically interrelated dispositions where the receiver's act corresponds to the sum of the cardinalities presented to each of the senders' (346). In the modular composition of addition and ordering, for the addition game and the ordering game to work together, the players need to evolve an association between the distinct modules. See Figure 3.3.

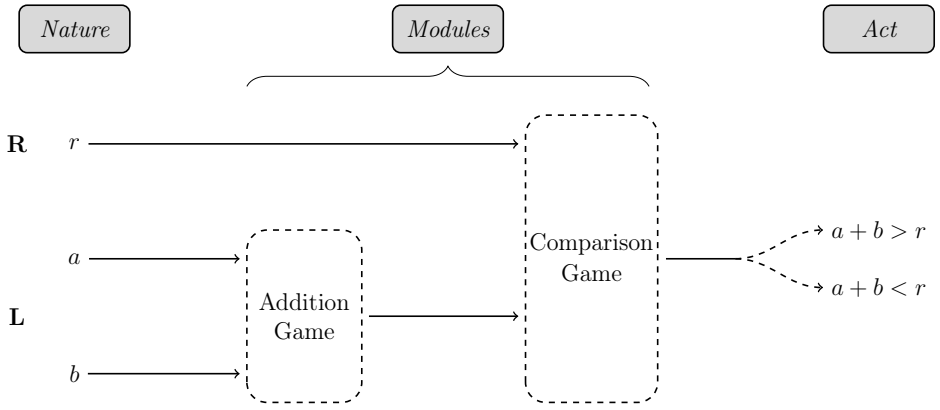


Figure 3.3: Example of modular composition of two separate games

This can be done via template transfer: the ordering game takes a single cardinal, r, as one of its input, and takes the translated output of the addition game ($a+b$) as its other input. The players are successful just in case they learn to judge correctly whether $a+b > r$ or $a+b < r$. Barrett and Skyrms (2017) report that this composite system evolves to successfully order $a+b$ and r with accuracy typically around 0.94 and always better than 0.90. They note that when the system gets the order wrong, the mean difference between $a+b$ and r was 1.52, which is consistent with empirical data from experiments on Rhesus macaques (Livingstone et al., 2014).

Finally, they show, using a simple toy model from logic, that template transfer under simple reinforcement learning of this variety evolves to success an order of magnitude faster than

if the agents must learn a new disposition from scratch under the same dynamic process. These results are replicated and expanded upon in Chapter 6, so I will leave the details of the logic game until then. For now, we want to examine compositional processes in general.

In their discussion of self-assembly, Barrett and Skyrms (2017) point out that the ritualisation of decisions explain how cue-reading, sensory-manipulation, and simple signalling might initially evolve. Further, new games can evolve out of old games via interrelated processes of polymerisation, template transfer, and modular composition. Their key results are that functional composition between *modules* may be more efficient than evolving a new disposition from scratch. They suggest that, though the evolution of *strategies* within a game has received considerable attention, 'the evolution of games themselves is important and deserves to be explored' (351).

Two questions naturally arise at this point. Do these various notions bring anything to bear on the evolution of complex or compositional communication? If so, are these processes empirically well-grounded? Underlying the process of modular composition, in the context of template transfer, is a notion of *analogical reasoning*. In the subsequent sections (3.2.1, 3.2.2, 3.2.3), I make explicit what is implied in the discussion of Barrett and Skyrms (2017), thus answering the second question. In the final section of this chapter, I suggest an answer to the first question.

3.2 Modular Compositional Processes

3.2.1 I: Transfer (of) Learning

What Barrett and Skyrms (2017) refer to as 'appropriation', in its most basic form, is perhaps the simplest way of evolving new strategies from old. A process of appropriation,

minimally, requires the following. First, the agents must have evolved a disposition for a particular context. This can happen in the usual way, via reinforcement learning or some other evolutionary dynamic. The agents are then faced with a novel context, where the prior disposition just happens to be appropriate—though this may not be known at the outset. We may assume that this novel context is relevantly similar, but non-trivially distinct, from the original context. Appropriation then consists in applying the prior strategy to the novel context. While processes of appropriation may indeed take advantage of other capacities, such as analogues or composition, so that the agent actively 'realises' that the old strategy is appropriate in this new context, this is not necessary. It may be the case that the agent happens, by chance, to try something pre-evolved when faced with a novel context. The appropriateness of the pre-evolved strategy may determine a sufficiently beneficial reward such that, when faced with this same context again, the agent learns quickly (even by simple reinforcement) to do the old action.

For example, a selective theory of antibody formation had initially been proposed by Paul Ehrlich, around 1897, according to which our immune systems respond to new viruses with the same antibodies—ones that *just happen* to bind appropriately to those novel antigens (Ehrlich, 1900). This is a case of transfer (in terms of generality) providing sufficient conditions for selection in an evolutionary context, rather than transfer in learning—though see the discussion in Piattelli-Palmarini (1989).

'Appropriation', however, has a technical meaning in the psychological literature on (human) learning processes. It is the process of constructing new knowledge out of various socio-cultural sources. This knowledge construction is generally taken to be mediated by an individual's own knowledge (Rogoff, 1990, 1993, 1995). In this sense, appropriation involves an agent's constructing her own version of knowledge—i.e., the appropriation of knowledge from an external source—which is then combined with pre-existing knowledge (Leontyev, 1981). The idea is that an agent's knowledge is socially constructed and that the agent plays

an active role in its construction; appropriation, in this sense, involves an agent adapting new information in such a way that it can be utilised in the context of her own (prior) knowledge (Cook et al., 2002; Grossman et al., 1999; Johnson et al., 2003). This is based upon constructivist views of developmental theories in cognitive psychology and sociocultural theory—influentially developed by Piaget (1950, 1968, 1976, 1980) and Vygotsky (1978, 1987).[6]

This process involves (1) submitting to a dependency, wherein the learning agent recognises a difference between her knowledge and that of another agent; (2) mirroring, wherein the learning agent adapts her own beliefs to the new knowledge (perhaps questioning certain aspects); and (3) construction, wherein the learning agent incorporates the new knowledge into prior knowledge (Hung, 2013). Alternatively, Grossman et al. (1999) propose five degrees of appropriation: (i) lack of appropriation; (ii) appropriation of label, wherein the learning agent knows the name of the concept, but does not know any of the features of the concept; (iii) appropriation of surface features, wherein the learning agent knows the particular features of a concept, but cannot synthesise them to form a conceptual whole; (iv) appropriation of conceptual underpinnings, wherein the learning agent understands the theoretical basis of a concept, which in turn informs the use of the concept in novel contexts; and (v) mastery of the concept.

Appropriation, in this psychological context, involves an explicit teacher-learner relation. Learning agents that learn actively are more likely to appropriate knowledge (Grossman et al., 1999; Johnson et al., 2003). Further, the social context—i.e., the environment in which the learning occurs—can play a role in the efficacy of student learning—for example, how knowledge is produced in that environment, and how social practices occur in that environment. Similarly, the student's background and motivation, as well as the teacher's motivation to teach, play a significant role here (Poleman, 2006). While appropriation in

[6]See also, Greeno (1989a,b); Glaser (1990); Lave (1990); Gauvain (1993); Pelissier (1991); Bredo (1994); Prawat and Floden (1994); Stevenson (1994); Billett (1998).

this technical sense might be relevant to the type of learning in which we are interested, the notion of appropriating a prior strategy for use in a novel context (at least in the sense that Barrett and Skyrms (2017) have in mind) is more readily described by *transfer of learning*.

This notion, like reinforcement learning, has a long psychological pedigree: it was initially introduced as *transfer of practice* (Thorndike and Woodworth, 1901a,b,c). Transfer of learning explores how individuals might transfer learning from one context to another (relevantly similar) context. Transfer of learning depends upon how similar the learning and transfer tasks are. In more contemporary research, transfer of learning describes the process by which (and the extent to which) past experience—the *transfer source*—affects learning in a novel context—the *transfer target* (Ellis, 1965).[7] Specifically, *positive* transfer occurs when what is learned in one particular context has a positive influence on learning in a novel context; negative transfer occurs when prior learned behaviours are detrimental to learning in a new context (Cree and Macaulay, 2000; Schunk, 2004).

From a computational perspective, in the context of machine learning, transfer of learning is modelled using transfer-learning algorithms. Transfer learning in this situation—which is conceptually related to the psychological notion of transfer of learning, though there is little to no formal relationship between the two—is a method wherein a model that is developed for a particular task or context is utilised as the starting point for a novel task. This was first described in Pratt (1993). As Goodfellow et al. (2016) summarise: '[t]ransfer learning and domain adaptation refer to the situation where what has been learned in one setting . . . is exploited to improve generalization in another setting' (526). This allows for the improvement of learning in a novel context by relating knowledge from a previous context which has already been learned (Torrey and Shavlik, 2009). However, transfer is only going to be useful when the features learned from the source task are relevantly general—i.e., they must be suitable to both the source and the target contexts (Yosinski et al., 2014). Transfer

[7]See also, Pugh and Bergin (2006); Hung (2013).

learning is commonly used in natural language processing problems that have text input or output. In this case, models utilise a word embedding, which is a mapping of words to a high-dimensional vector space where different words with similar meanings have similar vector representations.[8]

Though transfer learning is an optimisation, it is a nontrivial fact whether it will be beneficial in learning. Torrey and Shavlik (2009) describe three possible benefits when using transfer learning: (1) higher start, where the initial success (e.g., before further training on the target domain) is higher than learning from scratch; (2) higher slope, where the rate of improvement of successes during training is steeper than it otherwise would be; and (3) higher asymptote, where the converged successes of the trained model are better than they otherwise would be. They highlight that an ideal model would see benefits in all three of these categories.[9]

Machine learning is a relevant context to consider here since teaching an artificial agent to perform a particular task requires a precise specification of the task at hand. For example, suppose an AI is trained on a specific source context. To perform reasonably well in the test context, the AI needs to 'notice' that the new context is relevantly similar to the prior context. With supervised learning, for example, this is technically built-in by the fact that the operator of the system will generally only feed the AI relevant contexts. For example, suppose an ML programme is trained on reading handwritten letters using supervised learning; when it comes to the testing context, it is fed new *handwritten letters* which it has previously never seen. Thus, the context is relevantly similar by design. If the programme is well constructed, it should be able to determine the letter in question without trouble. This context is trivially novel—formally, it is identical to the training context, but it is distinct from any particular sample upon which the programme was trained.

[8]See, for example, Google's *word2vec* model [code.google.com/archive/p/word2vec/], Stanford's *GloVe* model [nlp.stanford.edu/projects/glove/], the Caffe Model Zoo [github.com/BVLC/caffe/wiki/Model-Zoo], etc.

[9]For more on transfer learning from a computational perspective, see Taylor and Stone (2009); Torrey and Shavlik (2009); Pan and Yang (2010); Goodfellow et al. (2016); Goldberg (2017).

If the AI was trained on handwritten letters and then during the testing phase was given handwritten numbers, the context would be dissimilar, but so much so that the AI would not be able to use its pre-learned disposition to guess at what the numbers are—though it might reasonably guess that '1' is 'l' or 'I', '2' is 'Z', '4' is 'A', '5' is 'S', etc. Hence, the requirement for generality in the learned features.

Consider now a natural context where transfer of learning might be beneficial. Suppose a troop of vervet monkeys has learned a signalling disposition for several alarm calls, as was discussed in Chapter 2. The primary three predators of the vervets are snakes, leopards, and eagles. Suppose a new predator is introduced to the vervets' environment—say wild dogs. The vervets might coordinate upon a new signal to alert others that a dog is nearby—this would involve coordinating upon a signal from scratch, and would doubtless result in several casualties during the learning period when, for example, a new signal is initially sent, and some vervets climb a tree, some scan the ground, etc. in response. However, given the structural similarities of a wild dog to a leopard, and the dissimilarities of a dog to a snake or eagle, a vervet might fruitfully take advantage of the pre-evolved leopard alarm call (assuming that climbing a tree is indeed an appropriate response in this case).

In fact, this is precisely what happens. Feral dogs sometimes attack vervet monkeys in the Cameroon Savannah. When a dog is present, a 'leopard' call is sent, and receivers climb into trees (Seyfarth et al., 1980b; Cheney and Seyfarth, 1990). However, vervets in the forests of Cameroon are hunted by humans with dogs. Here, they send a distinct type of signal, which is soft and pitched to match ambient background noise—thus, it is difficult to detect (Kavanagh, 1978). In this case, the response is also different: upon hearing this alarm call, receivers quietly flee into dense bush.

A philosophical issue arises here concerning the interpretation of alarm calls. As Cheney and Seyfarth (1990) highlight:

> [i]n Amboseli, where leopards hunt vervets but lions and cheetahs do not, leopard alarm could mean 'big spotted cat that isn't cheetah' or 'big spotted cat with shorter legs', or however you want to describe it. In other areas of Africa, where cheetahs *do* hunt vervets, leopard alarm could mean 'leopard *or* cheetah'. . . . [When dogs are included in the alarm call, its] meaning appears to be 'terrestrial predators from which you can escape by running into trees'. (169)[10]

We might speculate upon the following: it is entirely possible that, in the case of vervets in the Cameroon Savannah, the Leopard alarm call was learned and resulted in a stable disposition before the introduction of wild dogs in the vervet habitat. It might also be possible that, via simple reinforcement, selection pressures, noted salient similarities, etc., the vervets transferred their strategy for leopard contexts to dog contexts. Cheney and Seyfarth (1990) highlight that 'as long as new predators [i.e., new contexts] fall within this [i.e., "appropriate"] category the same alarm call will presumably be used' (169).

However, there is nothing *a priori* (or the monkey equivalent of '*a priori*') that determines that this should be the case. The appropriate action still needs to be learned in the novel context—e.g., a hitherto unseen predator. It may well be the case that the salience of similarity in form—terrestrial, four-legged, furry, etc.—helped to evolve such a disposition more quickly, but this is not necessary. Minimally, transfer of learning requires *only* that an agent try prior strategies. Successful strategies may be learned via simple reinforcement, or they may be learned via a more sophisticated trial-and-error. When salience is present—e.g., the physical properties of a new predator being saliently similar to an old predator—the new strategy may be implemented immediately; however, this is a more sophisticated version of transfer of learning, which requires a notion of *analogical similarity*—see Section 3.2.2 below.

Template transfer, in its most basic description, is a form of transfer of learning, mediated solely by reinforcement learning. In the example of Barrett and Skyrms (2017), we suppose that the agents have pre-evolved a NAND disposition, for example, and then are presented with a new context where NAND is appropriate, but it is not necessarily evident at the outset

[10]See also the discussion in Quine (1960); Harms (2004b); Huttegger (2007b); Zollman (2011).

that NAND is indeed appropriate. The agents learn to utilise the NAND disposition. They thus 'transfer' a previously learned 'template' and apply it to the new context. As Barrett and Skyrms (2017) highlight, this would work equally well for learning an OR disposition in a novel context—here the template is transferred and permuted slightly—however, it will not be sufficient for the agents in the template transfer context to learn AND or XOR, for example (though see Chapter 6).

Transfer of learning, as a cognitive psychological process, is often researched anthropocentrically in terms of both content and theory (Zentall et al., 2008). Nonetheless, there is good evidence that corvids, in addition to primates, are capable of transfer. This allows for flexibility of behaviour in problem-solving, via the ability to generalise learned rules to novel contexts—hence the relation to composition, in the linguistic sense. This is the example that Barrett and Skyrms (2017) highlight.[11]

One further example of transfer of learning in the context of nonhuman animals is an extension of classification tasks, which involves so-called 'reversal learning'. In this case, an animal is trained to associate a particular stimulus with a reward—this can be modelled as usual using reinforcement learning. Once the agent exhibits some particular degree of success—say, 0.87—the relation between the stimulus and the reward is reversed. As such, the agent must replace the prior association with the *opposite* association. For example, in a lab setting, an animal might have to pull one of two levers—'left' or 'right'. Suppose the animal has learned to associate 'left' with a reward and 'right' with punishment (or at least no reward). The environment then switches so that 'right' produces a reward, and 'left' produces punishment (or no reward).

The underlying assumption of this experimental set up is that if the animal can quickly reverse its associations, then it does so based on some *concept* of OPPOSITENESS. On the

[11] See also Hunter III and Kamil (1971); Wilson et al. (1985); Bond et al. (2003); Paz-y-Miño et al. (2004); Emery and Clayton (2004).

other hand, it may equally be the case that the new association takes as long or longer to be learned—this might be because the agent does not understand that or why the rewards are different, given the prior learned association. In this case, no such application of conceptual understanding used to facilitate learning may be attributed to the agent. Furthermore, the success threshold for the initial training can be varied in an experimental setting. A high degree of success required in the initial training before shifting the context should have the following effect on training in the new context: if the agents do not make use of a concept of OPPOSITENESS, then a higher success threshold in the initial training should entail more extended training in the new context, whereas if the agents do make use of a conceptual shortcut—where they transfer prior knowledge—then higher success rates on the initial training context should entail expeditious successes in the new context. This experimental discrimination reversal paradigm is referred to as the *transfer index* (TI) (Rumbaugh, 1970, 1971). This method has been used extensively to compare the cognitive performance of nonhuman primates and to derive information on the evolution of intelligence (Bonte et al., 2014).

Rumbaugh and Pate (1984a,b) tested this hypothesis with a variety of species of great apes, old- and new-world monkeys, and prosimians. They showed that great apes perform significantly better than monkeys, and monkeys perform markedly better than prosimians. They examined threshold levels of 0.64 and 0.87 for the source context. Learning in the source context involved a series of two-choice, object-discrimination problems. Each problem consisted of a pair of objects that differed clearly in size, colour, and form. Choosing one object resulted in a reward (food), whereas selecting the other resulted in no reward. Once the agents achieved mastery of at least the threshold level, the rewards for the choices were reversed.

Prosimians tended to exhibit *negative* transfer—where prior training in the source context inhibited learning in the target context. Further, when the threshold levels were increased,

prosimians tended to do worse. The opposite was true for the great apes and several species of monkey.[12]

Hurford (2007) argues that reversal learning experiments do not merely highlight an ability to apply the relation of OPPOSITENESS between a source and a target context; instead, the agent 'seems to be keeping its old mental representation (concept) of the general class of stimuli acquired in the first training regime and relating the new set to that acquired concept' (25). In a similar set of experiments, Deacon (1997) highlights the importance of recognising higher-order regularities between various (lower-order) associations. This is 'a trick that can accomplish the same task without having to hold all the details in mind' (Deacon, 1997, 89). The modelling presupposition of transfer learning, as was highlighted in Barrett and Skyrms (2017), and has been highlighted here, does not require that the agents 'notice' any similarity between the prior context and the novel context. Rather, this can be accomplished by simple reinforcement. However, it is possible that the agents *do* note a similarity between the two contexts and utilise this similarity to (more quickly) reason about what is appropriate in the novel context. This gives rise to the notion of *analogical reasoning*.

3.2.2 II: Analogical Reasoning

An analogy is a comparison of the apparent similarity between two objects or systems of objects. Analogical reasoning is a form of reasoning that takes advantage of such apparent similarity. An *analogical argument* is an explicit representation of analogical reasoning, which depends upon cited similarities between the objects or systems in question, and which supports the (explicit) conclusion that some further similarity exists. Analogical reasoning plays an important role in problem-solving in human and nonhuman animals. In analogical reasoning, analogy serves an *heuristic role*, and it can also serve a justificatory role.

[12]See also the discussion in Rumbaugh (1995).

Analogies further have some *predictive* value, and they might be used for *conceptual unification*. An analogical argument is inductive to the extent that the analogy, upon which the argument depends, makes the argument's conclusion plausible (in the sense of enhancing its probability).

In analogical reasoning, there is a noted similarity between a *source domain* or *context* and a *target domain* or *context*. The analogy comes into play when noting some property, P, in the source context and reasoning that the property, P, or some similar property, P', holds in the target context. This purported property in the target domain is referred to as the *hypothetical analogy* (Keynes, 1921). This sort of analogical reasoning is well modelled by the simple reinforcement learning processes that were discussed in Chapter 1 because the notion of entailment is inductive and concerns *plausibility*—e.g., given that the source and target contexts share some relevant structural properties, it is plausible that the target context also contains some other property which obtains in the source.

Bartha (2016) enumerates the following general guidelines for the strength of analogical reasoning:[13]

1. The more similarities (between two domains), the stronger the analogy; similarly, the more differences, the weaker the analogy.
2. The greater the extent of our ignorance about the two domains, the weaker the analogy.
3. The weaker the conclusion, the more plausible the analogy.
4. Analogies involving causal relations are more plausible than those not involving causal relations.
5. Structural analogies are stronger than those based on superficial similarities.

[13]See also Mill (1843); Keynes (1921); Robinson (1930); Stebbing (1933); Moore and Parker (1998); Woods et al. (2004); Copi and Cohen (2005).

6. The relevance of the similarities and differences to the conclusion (i.e., to the hypothetical analogy) must be considered.

7. Multiple analogies supporting the same conclusion make the argument stronger.

A common way of thinking about the *structural* analogy between two distinct systems is through the concept of a *model-theoretic isomorphism*. In this case, the strength of an argument from analogy depends inherently upon the strength of its associated analogy mapping.[14] In first-order model theory, given two structures—in this case, the source domain, S, and the target domain, T—a *homomorphism* from structure S to structure T is a function f from the domain of S to the domain of T with the property that, for every atomic formula $\phi(v_1, \ldots, v_n)$ and any n-tuple $\mathrm{s} = (s_1, \ldots, s_n)$ of elements of S,

$$S \vDash \phi[\mathrm{s}] \Rightarrow T \vDash \phi[\mathrm{t}],$$

where $\mathrm{t} = (f(s_1), \ldots, f(s_n))$. When the converse of this also holds, f is called an *embedding* of S into T. An embedding of S into T is always *injective*; if it is also *surjective*, then the inverse map, f^{-1}, from the domain of T to the domain of S is also a homomorphism. In this case, the embedding and its inverse are *isomorphisms*. Finally, the two structures S and T are called *isomorphic* when there is an isomorphism from one to the other. Isomorphism is an *equivalence relation*—a binary relation that is symmetric, reflexive, and transitive—on the class of all structures of a fixed signature K—where a signature is a set of individual constants, predicate symbols, and function symbols. If two structures are isomorphic, then they share all model-theoretic properties—specifically, they are *elementarily equivalent*.[15]

Gentner and Gentner (1983) examine the conceptual role of analogy for building mental models of complex systems, and they ask whether analogical reasoning consists in merely borrowing available language from one domain to apply to another, or more profoundly

[14] Though see Schlimm (2008) for a critique of the structure-mapping theory presented here.

[15] See the discussion in Hodges and Scanlon (2018).

thinking *in terms of* analogies to have real conceptual effects. They test, what they call, the *Generative Analogy Hypothesis*, that 'conceptual inferences in the target follow predictably from the use of a given base domain as an analogical model' (100).

The theoretical framework that Gentner and Gentner (1983) employ requires a notion of *structure mapping*. On their view, analogies take advantage of 'certain aspects of existing knowledge, and that selected knowledge can be structurally characterized' (101). A comparison of two complex concepts by analogy takes advantage of relations between the constituent parts of the complex concept but does not require that any two objects in that domain are similar. They use the example (1) *The hydrogen atom is like the solar system*, which is often invoked to explain this model,[16] and compare it to (2) *There's a solar system in the Andromeda nebula that is like our solar system.* The former conveys structural similarities between the structural relations of the electron to the nucleus and the planets to the sun; whereas, the literal comparison relates the similarity of both structural relations and objects. Useful analogies are supposed to be characterised, on this view, by *systematic* relational correspondences: '[a]nalogies are about relations, rather than simple features. No matter what kind of knowledge (causal models, plans, stories, etc.), it is the structural properties (i.e., the interrelationships between the facts) that determine the content of an analogy' (Falkenhainer et al., 1989, 3).

In this sense, we might say that an agent utilises analogical reasoning when she applies the *predicates* of a known base domain to those of a (lesser- or unknown) target domain—the domain of inquiry. Therefore, the relational structure between objects is preserved, but the objects themselves may be different. Thus, in the above example, the structural relations $\textsc{Attracts}(x, y)$, $\textsc{More-Massive-Than}(x, y)$, $\textsc{Revolves-Around}(x, y)$ are relations that

[16]Bartha (2010) notes that though the orbit of planets in the solar system is often invoked, by analogy, to explain the orbit of electrons around a nucleus, this analogy did not appear to have played any role in Rutherford's thinking (4, FN 3).

hold *analogously* to the separate and distinct object pairs, $(x, y) = (Sun, Planet)$ and $(x, y) = (Nucleus, Electron)$.

The most influential proposal for a theory of analogy is the *structure-mapping theory*, which was first proposed in Gentner (1983).[17]

Definition 3.1. *Analogy (Structure Mapping)*

Let S be a set of objects, $s_1, \ldots, s_n$, and predicates, A, R, R', called the *source domain*. Let $T = \{t_1, \ldots, t_m\}$ be a set of objects, called the *target domain*. A *structure-mapping analogy* is a function,

$$M : S \to T,$$

that maps objects from the base domain to objects in the target domain, subject to the following rules:

1. *Preservation of Relation*: If a relationship, R, between objects exists in the source domain, then the same relation holds to the corresponding objects in the target domain.

   $$M : [R(s_i, s_j)] \to [R(t_i, t_j)].$$

   However, no such requirement holds (necessarily) for attributes.

   $$M : [A(s_i)] \not\to [A(t_i)]$$

2. *Systematicity*: Higher-order relations are more strongly predicated than isolated relations—i.e., predicates are more likely to be imported to the target when they belong to a system of coherent, mutually constrained relationships,

[17] See also, Forbus et al. (1994); Forbus (2001).

which map into the target.

$$M : [R'(R_1(s_i, s_j), R_2(s_k, s_l))] \rightarrow [R'(R_1(t_i, t_j), R_2(t_k, t_l))]$$

According to the *systematicity principle*: '[a] predicate that belongs to a mappable system of mutually interconnecting relationships is more likely to be imported into the target than is an isolated predicate' (Gentner and Gentner, 1983, 163). As a result, we have it that predicates which occur in statements that involve higher-order relations are more likely to be imported into the target than those predicates which only occur in lower-order relations; further, properties and functions of objects are unimportant in an analogy except in the extent to which they are part of a relational network. This principle of systematicity that arises in Gentner's structure-mapping theory is taken to be descriptive insofar as it 'fits with evidence that people naturally interpret analogy and metaphor by mapping connected systems of belief, rather than independent features' (Gentner et al., 2001, 208). However, Dunbar (2001) suggests that 'unless subjects are given extensive training, examples, or hints, they will be much more likely to choose superficial features than deep structural features when using analogies' (313).[18]

We have three further possible restrictions on the mapping, M:

1. *Identicality.* Only identical relational predicates can be matched, although non-identical objects, functions, and monadic predicates may be matched.[19]

2. *n-ary restriction.* M must map objects to objects, n-place functions to n-place functions, and n-place predicates to n-place predicates.

3. *Consistency.* Whenever M maps P to P^*, it must map the arguments of P to the corresponding arguments of P^*.

[18] See also Gick and Holyoak (1983); Forbus et al. (1995).
[19] This is relaxed in later work; see Forbus (2001).

This formulation does give rise to several philosophical problems, however—for example, the n-ary restriction entails discounting certain kinds of systematicity, increased systematicity is not sufficient for increased plausibility, and the systematicity principle does not account for the 'direction' of relevance (Bartha, 2010). To try to deal with some of these issues, Holyoak and Thagard (1989) introduced a *Constraint-Satisfaction* model of analogy, which takes account of the pragmatic importance of analogies: '[a]nalogies are virtually always used to serve some known purpose, and the purpose will guide selection [of the aspects of the source relevant to the analogy]' (Holyoak and Thagard, 1989, 297).[20]

Following Hesse (1966), we can distinguish between *horizontal* and *vertical* relations in an analogy. Horizontal relations are similarity (and difference) relations between domains, whereas vertical relations involve correspondence between objects, properties, and relations within a domain.

Following Keynes (1921), we can further distinguish between a *positive* analogy and a *negative* analogy:

> **Definition 3.2.** *Positive Analogy*
>
> Let $P = \{P_1, \ldots, P_n\}$ be a set [or 'list'] of accepted propositions about the source domain, S. Let $P^* = \{P_1^*, \ldots, P_n^*\}$ be a set of corresponding propositions, which are all accepted as holding of the target domain, T. P and P^* represent accepted (or known) similarities. We refer to P as the *positive analogy* .

> **Definition 3.3.** *Negative Analogy*
>
> Let $A = \{A_1, \ldots, A_r\}$ be a set [or 'list'] of propositions that are accepted as holding in S, and let $B^* = \{B_1^*, \ldots, B_s^*\}$ be a set of propositions holding in T. Suppose the analogous propositions $A^* = \{A_1^*, \ldots, A_n^*\}$ fail to hold in T, and the propositions $B = \{B_1, \ldots, B_n\}$ fail to hold in S. We can write $A, \neg A^*$ and $\neg B, B^*$ to represent accepted or known differences and refer to A and B as the *negative analogy.*

[20] This account is updated to the *multiconstraint theory* in Holyoak and Thagard (1995).

Definition 3.4. *Neutral Analogy*

The *neutral analogy* consists of accepted propositions about S for which it is not known whether an analogue, Q^*, holds in T.

Definition 3.5. *Hypothetical Analogy*

The *hypothetical analogy* is the proposition, Q^*, in the neutral analogy that is the focus of our attention.

Note that what I have said about analogy thus far seems to imply a relatively strict cognitive presupposition to the extent that analogy seems to require noting and comparing differences and similarities between source and target domains, and drawing conclusions from this information.

However, Bartha (2010) notes that an analogical *argument* has the form: '[i]t is plausible that Q^* holds in the target because of certain known (or accepted) similarities with the source domain, despite certain known (or accepted) differences' (15). Here, plausibility might be simply interpreted as *prima facie* plausibility. Furthermore, to say an hypothesis is *prima facie* plausible is to say only that (1) it has epistemic support, and (2) it has pragmatic importance. The first of these simply requires an appreciable likelihood of being true (or *successful*); this notion is well modelled by the accumulated rewards of simple reinforcement learning in the signalling-game context. The second requires that the hypothesis is worth investigating, but this is captured by the probabilities of choosing a particular option as one's strategy in a given round—recall the distinction between exploration and exploitation discussed in Chapter 1. Therefore, though sophisticated cognitive capacities may help in analogical reasoning, this is not necessary.

How does this notion of analogy map onto what is going on in modular composition, as we are concerned with here? In this case, our *source domain* is a known context, c_1, in which a pre-established disposition exists. The novel context is our *target domain*, c_2. In this

case, we have a relationship between objects in the base and the target. Namely, the sender constitutes a pre-evolved signalling disposition given by the sender and receiver strategies

$$\sigma_{c_1} : S \to \Delta(M)$$
$$\rho_{c_1} : M \to \Delta(A)$$

Where σ and ρ are supposed to be bijective in the (source) context c_1. That is, there is a relation between the state and the message, on the part of the sender, and a correspondent relation between the message and the act on the part of the receiver. Now, suppose there is a novel (target) context c_2, where the same actions on the part of the sender and receiver might be appropriate. An analogy is a relation between the sender strategy in the base domain—context c_1—and the target domain—context c_2.

In the NAND game that Barrett and Skyrms (2017) discuss, we have the following concrete example. c_1, the base domain, is the signalling game that has previously evolved. This has the following components:

$$S = \{s_i s_j\} = \{00, 01, 10, 11\}$$
$$M = \{m_1, m_2, m_3, m_4\}$$
$$A = \{0, 1\}$$
$$u(s, a) = \begin{cases} 1 & \text{if } a \equiv (\overline{s_i \wedge s_j}) \\ 0 & \text{else} \end{cases}$$
$$\sigma_{c_1} : S \to \Delta(M)$$
$$\rho_{c_1} : M \to \Delta(A)$$

Formally, in this context, an analogy is a structural mapping, M, from the source—the pre-evolved disposition—to the target—the novel disposition. Both preservation of relation and systematicity are satisfied. However, the n-ary restriction entails that this structure-

mapping theory will not capture an analogy between, e.g., NAND with 2 inputs and NAND with 3 inputs—see Chapter 6 for further discussion.[21]

Like transfer of learning, analogy has been well researched in the context of human leaning; however, if analogy is to play any role in the evolution of novel dispositions out of pre-evolved dispositions for an explanation of the evolution of complex communication in nature, we must remain sensitive to empirical evidence concerning the ability (or inability) of nonhuman animals to utilise analogies.

Thompson and Oden (2000) argue that 'There is no evidence that monkeys can perceive, let alone judge, relations-between-relations. This analogical conceptual capacity is found only in chimpanzees and humans' (363). However, Katz et al. (2002) write (with respect to the analogical conceptual capacities of certain species of animals):

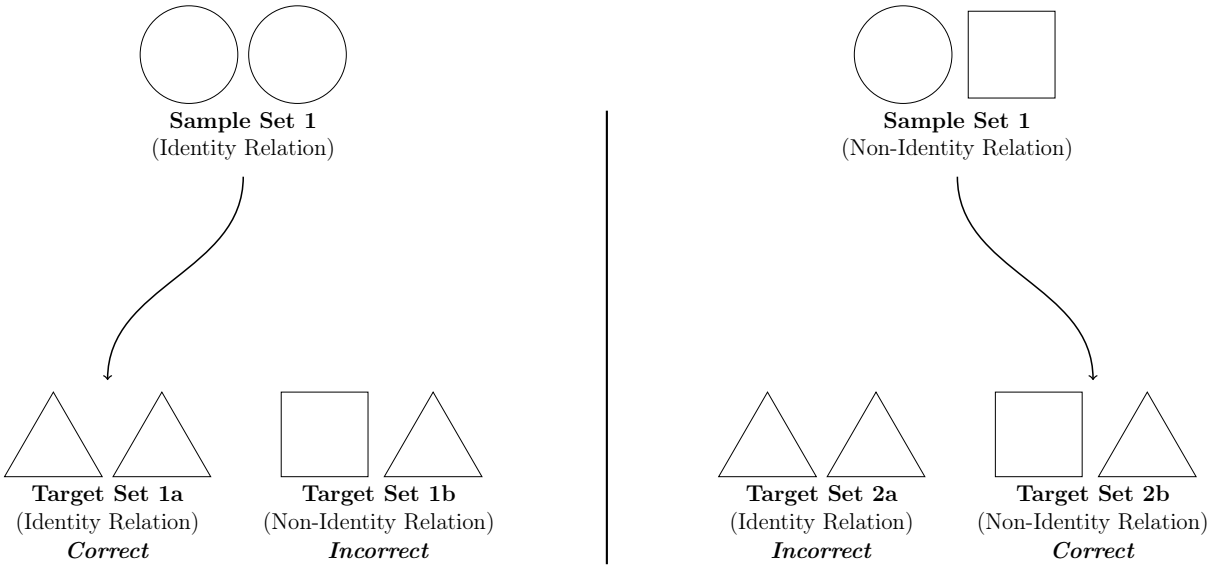
> most species (e.g., most vertebrates) ultimately have a set-size function for abstract-concept learning. Some species (e.g., pigeons) may have a set-size function that is lower than that for rhesus monkeys and would require larger set sizes to achieve full abstract-concept learning. Other species (e.g., chimpanzees) may have a set-size function that is elevated relative to that for rhesus monkeys. If the set-size function is sufficiently elevated, then that species under those conditions might be able to learn an abstract concept with very few items. Humans can demonstrate equivalence relationships after being trained with small stimulus sets. (367)

In this case, the analogy between various stimuli requires a concept of SAME versus DIFFERENT. Thus, as with transfer learning, there is some evidence that nonhuman animals can utilise analogical reasoning.

The most common way of testing this is with a set of analogy problems known as *relational matching-to-sample* (RMTS) tasks (Skinner, 1950; Blough, 1959; Ferster, 1960). This experimental task involves showing the agent a sample set, which consists of two or more objects that are either identical or non-identical. The agent is then shown two comparison sets,

[21] This is, in effect, the same issue that Bartha (2010) takes with the structure-mapping theory.

which contain novel objects—one of which involves identity, and the other of which involves non-identity. To be successful, the agent must choose the comparison set, which matches the sample set. (See Figure 3.4 for an example.)



(a) RMTS task matching identity

(b) RMTS task matching non-identity

Figure 3.4: Two examples of relational matching-to-sample task involving matching identity or non-identity relation between two objects

Note, in particular, that this task does not rely on any linguistic capacities.

In a set of experiments with the RMTS task, Truppa et al. (2011) showed that a tufted Capuchin monkey (*Cebus apella*) was successful with four-item stimuli (a 2×2 grid of 'all same' or 'all different') and subsequently with two-item stimuli; the latter condition is noted as being the most challenging condition to master (because fewer data are available to the subject). The two-item stimuli task was previously thought to be mastered only by apes (Thompson and Oden, 2000). For example, Fagot et al. (2001); Wasserman et al. (2001) show that baboons can perform the RMTS task with 4×4 icons. However, successes decrease as the size of the icon set decreases: by the time the test set gets down to two-item stimuli, the baboons performed no better than chance. However, Flemming et al. (2011) show that, though Rhesus macaque monkeys (*Macaca mulatta*) failed to perform better than chance on the two-item stimuli RMTS task under differential reward-only or punishment-only

conditions, they performed significantly better than chance when they received a combination of reward and punishment.[22]

Thus, noting and taking advantage of analogy for transfer of learning is more cognitively complex than simple transfer. However, there is still good evidence that apes and monkeys can perform such tasks. Increasing complexity again, we arrive at a full notion of *modular composition*.

3.2.3 III: Modular Composition

The notion of modular composition (in terms of communication) is intimately related to *cognitive* differences between human and nonhuman animals. Cognitive capacities, like communicative capacities, vary significantly between and within species. Even so, Darwin (1871) suggested that 'the difference in mind between man and the higher animals ... is certainly one of degree and not of kind' (105). However, this is controversial, and some researchers believe it is a mistake (Penn et al., 2008). Though social animals may be extremely adept at responding to (past and present) behaviour of their companions, this does not necessarily entail a 'theory of mind' (Penn and Povinelli, 2007b; Seed et al., 2012; van der Vaart et al., 2012)—though see Hurford (2007). Nonetheless, so-called 'higher' cognitive processes cited in humans can often be shown to arise out of simple process—ones which may be readily available to nonhuman animals (Shettleworth, 2010).

Modularity provides a clear theoretical advantage over serial processes. Simon (1962) illustrates this with a parable of two watchmakers, Hora and Tempus: Hora builds watches out of independent modules, whereas Tempus builds them piece-by-piece. Simon (1962) posits that if either watchmaker is interrupted (and they are often interrupted), the ongoing work

[22]This is consistent with simulation results in higher-dimensional signalling games, as was discussed in Chapter 1—though reward alone may not be sufficient to avoid partial-pooling equilibria, adding punishment can significantly improve the results of the simulations.

is undone. In this case, Hora has the significant advantage that an interruption only disrupts the assembly of a specific module, whereas when Tempus is interrupted, he must start from scratch.

Though this is merely an illustrative anecdote, modularity is a fundamental principle in evolutionary developmental biology. It is argued that evolution might *only* be possible via modular systems, wherein the parts of an organism can change while other, well-adapted, parts remain (at least relatively) unchanged (West-Eberhard, 2003; Ploeger and Galis, 2011). For example, Sherry (2006) highlights that species that bear high cognitive demands for spatial memory—concerning, e.g., retrieving stored food or defending vast territories—might evolve special spatial memories (and so exceptional hippocampi), while maintaining cognitive similarity in other respects to their close relatives. In the context of cognitive processes, the idea of modular composition is that (adult) human cognition shares simple basic ('core') processes with nonhuman animals, but it also includes one or more slower-developing, slower-acting, and more explicit (consciously accessible) processes (Shettleworth, 2012).

Carruthers (2006) argues that animal minds *must* be 'massively' modular. However, modularity also falls on a spectrum of degree: 'since the need for modular organization increases with increasing complexity, we can predict that the human mind will be the *most* modular amongst animal minds, whereas the minds of insects (say) might hardly be modular at all' (22, FN 12). Shettleworth (2012) argues that nonhuman animals provide excellent evidence for modularity of cognition, since

> [d]istinct operations on computationally distinct kinds of inputs are demanded by spatial, temporal and numerical cognition, recognition of animacy or social relationships and so on. Even associative learning is not a general process of information acquisition but a specialization for tracking contingencies, or temporally predictive relationships, among events. (2796)[23]

[23]See also, Shettleworth (1998).

Furthermore, the output of distinct modules must be integrated in some way, to determine action, or even memory (Shanahan, 2012; Clayton et al., 2001).[24]

Spelke (2003) suggests that humans and other animals are endowed with early developing, core systems of knowledge, called 'modules'. However, these core systems are limited in several ways: (1) they are *domain-specific*, in the sense that these modules represent only a subset of entities in the surroundings of the agent; (2) they are *task-specific*, in the sense that they inform only a subset of the repertoire of the agent's actions and cognitive processes; (3) they are (at least relatively) *encapsulated*, in the sense that there is a restriction on the flow of information into and out of a module; and (4) modules are (at least relatively) isolated from one another, in the sense that they do not readily combine (Spelke, 2003, 291).[25]

This is a slight departure from Fodor (1983), where modules are domain-specific, peripheral, perceptual mechanisms, which are innate, fast-acting, unconscious, obligatory and encapsulated. Modules, in the sense of Fodor (1983), are also usually assumed to be neurally specific, or localisable in the brain.[26] Barrett and Kurzban (2006) have argued that the *essential* property of (cognitive) modularity is functional specificity—that is, unique informational domains require unique processes to operate on them.[27] Coltheart (1999); Shettleworth (2012) highlight that whether such functional modules have other 'Fodorian' properties is an empirical question.

For *cognitive* mechanisms, an encapsulated mechanism has a specific task, which is defined by some particular set of inputs and which gives rise to a particular set of outputs. The encap-

[24]However, Shettleworth (2012) points out that 'A strictly modular view seems to preclude processes like associative learning, attention or working memory that cut across domains, although it does not preclude separate modules having some properties in common. It also seems to have no place for general intelligence, for which there is increasing evidence in non-human species' (2796-2797). See also, Matzel and Kolata (2010); Reader et al. (2011); Herrmann and Call (2012).

[25]See also Fodor (1983, 1984a, 2000); Sperber (1994, 2002); Carruthers (2002) and the discussion in Robbins (2017).

[26]However, Coltheart (1999) highlights that Fodor did not always require that cognitive modules always have all of these properties.

[27]See also, Sherry and Schacter (1987).

sulated mechanism, in this sense, has *internal* computational resources, which are 'hidden' to other mechanisms and so cannot be exported as part of the output of that mechanism.[28] For example, visual perception is often taken as a paradigm example of cognitive encapsulation, since optical illusions persist *even when* an individual is consciously aware it is an illusion (Fodor, 1983; Shettleworth, 2012; Robbins, 2017).[29] This is in contrast to a general-purpose mechanism (such as working memory) which is broad in both its inputs and outputs, as well as open in its computational structure. Such a non-encapsulated mechanism is called *executive*.[30]

Rather than focusing on cognitive differences, a modular view of cognition entails examining cognitive similarities between human and nonhuman animals. Many core cognitive capacities that are available to humans are also available to nonhuman animals—specifically, capabilities that were once thought to be unique to humans such as object mechanics, number sense, natural geometry (Spelke, 2003). In these and other human 'mental powers'—e.g., memory, language, tool use, imitation, etc. (Darwin, 1871)—adult humans tend to exhibit capacities over and above other species (hence the longstanding belief that these capacities were unique to humans); however, young children tend to perform, at best, on a par with other species (Spelke, 2003; Spelke and Kinzler, 2007; Gelman, 2009; Carey, 2011b; Spelke and Lee, 2012).

For example, there is good evidence that human infants have a system for perceiving objects and their motions, for filling in the surfaces and boundaries of a partly hidden object, and for representing the continued existence of an object that moves entirely out of view.[31] These studies control for the agent's representations of the features and spatial locations

[28]This is related to, but distinct from, a notion of *cognitive impenetrability* (Pylyshyn, 1984, 1999). See the discussion in Robbins (2017).

[29]However, see Ogilvie and Carruthers (2016).

[30]By analogy, we might consider an object-oriented programming language, such as Java. An encapsulated mechanism is like a *private* class, which is only accessible to components contained within that class—so it is computationally isolated or encapsulated—whereas a general purpose executive mechanism is like a *public* class, which is accessible to any relevant components.

[31]e.g., Wynn (1992); Simon et al. (1995); Koechlin et al. (1998); de Walle et al. (2001); Feigenson et al. (2002); See Wynn (1998) and Spelke (1998) for reviews of this literature.

of the objects. However, preferential looking, manual search, and locomotor choice tasks have also been presented to free-ranging (adult) rhesus monkeys (Hauser et al., 1996, 2000), whose performance equalled or exceeded that of human infants. Further, such capacities to represent objects have been demonstrated in infant nonhuman animals as well, including 1-day old chicks (Regolin et al., 1995; Regolin and Vallortigara, 1995; Lea et al., 1996). Thus, core systems for representing objects are not unique to humans.

Core systems for a 'number sense' (a sense of approximate numerical values and relationships) are present in human adults (Dehaene, 1997; Gallistel and Gelman, 1992), and performance adheres to Weber's law.[32] Xu and Spelke (2000) test 6-month old infants' abilities to discriminate between large numerosities and find that infants fail to discriminate between arrays of 8 versus 12 dots, implying that their number sense is imprecise. However, many species of nonhuman animals, including fish, pigeons, rats, and primates, are capable of discriminating between numerosities, and their abilities are also in accordance with Weber's law (Dehaene, 1997; Gallistel, 1990).

Finally, a core system of 'natural geometry' is present in young human children (Landau et al., 1984); however, a wide range of nonhuman animals outperform human infants in navigation tasks requiring geometric concepts—notably, desert ants (Wehner and Srinivasan, 1981). Spelke (2003) is worth quoting at length:

> These ants leave their nest in the nearly featureless Tunisian desert in search of animals that may have died and can serve as food, wending a long and tortuous path from the nest until food is unpredictably encountered. At that point, the ants make a straight-line path for home: a path that differs from their outgoing journey and that is guided by no beacons or landmarks. If the ant is displaced to novel territory so that all potential landmarks are removed, its path continues to be highly accurate: within 2 degrees of the correct direction and 10 percent of the correct distance. This path is determined solely by the geometric relationships

[32] As numerosity increases, the variance in subjects' representations of numerosity increases proportionately; thus, discrimination between distinct numerosities depends on their difference ratio.

> between the nest location and the distance and direction traveled during each step of the outgoing journey. (289)

Therefore, humans, but also nonhuman animals, have early developing, core-knowledge systems, which allow for an extensive range of intelligent behaviour and cognitive capacities; however, in many cases, these same core systems enable nonhuman animals to outperform human infants in similar tasks. Thus, core systems alone do not account for uniquely human cognitive capacities.

Historically, one of the hallmarks of human-level cognitive capacities involved the ability of humans to use tools. However, it has now been known for half a century that many nonhuman animals make use of tools.[33] More recently, crows have been shown in their natural environment to use tools with a sophistication rivalling that of chimpanzees (Orenstein, 1972; Chappell and Kacelnik, 2002; Hunt and Gray, 2003, 2004a,b; Weir et al., 2004; Kenward et al., 2005). Further, crows that are raised in the absence of external models (i.e., humans or other birds) still learn to use sticks to probe for food. In this sense, some aspect of the *capacity* for tool use is apparently innate in this species (Kenward et al., 2005).

Tool-use is an important benchmark for cognition for several reasons: to make use of a tool in the first place, an agent must have a *goal* in mind. Further, it must be able to keep this goal in mind long enough to find an appropriate tool to apply to the task in mind. This variously involves conceptual, abstract, and future-directed cognition. Animal *toolmaking* requires even more cognitive resource than this since the agent must keep in mind various sub-goals needed for the construction of a tool while maintaining the primary goal for which

[33]See, for example, Hall and Schaller (1964); Goodall (1968, 1986); Sugiyama and Koman (1979); Beck (1980); McGrew (1992); Boesch and Boesch (1983); Boesch (1991); Boesch and Boesch-Achermann (2000); Seed and Byrne (2010); Beck et al. (2011). Note that the use of tools in chimpanzees was already mentioned by Darwin (1871).

she required the tool in the first place. This seems to imply a hierarchically structured complex goal system.[34]

The general idea of the utility of hierarchical structure in building complex systems finds application in describing social structures, such as business firms, governments, and universities, as well as familial or tribal units; biological systems, in the sense of cells (which are themselves composed of well-defined subsystems such as nucleus, cell membrane, microsomes, and mitochondria, etc.) composing tissues, which in turn compose organs and further systems; physical systems, which are composed of elementary particles, atoms, molecules, etc. at the microscopic level, and satellite systems, planetary systems, and galaxies at the macroscopic scale; and symbolic systems, wherein words combine to form clauses and phrases, which combine to form sentences, which combine to form paragraphs, etc. (Simon, 1962, 1972).

The answer which Spelke (2003) offers to her titular question—*what makes us smart?*—is that human cognitive capacities depend on core knowledge systems, which are shared by other animals, *and* on a uniquely human combinatorial capacity for conjoining these representations to create new systems of knowledge. Furthermore, she suggests that the latter capacity is made possible by natural language, which provides the medium for combining the representations delivered by core knowledge systems (305). Specifically, it is the *compositional* nature of natural language, which gives rise to uniquely flexible human cognition, on her account.

Donald (1991) argues from a neuroscientific point of view that language is executive, rather than encapsulated, in the sense that it can 'reach into' any aspect of cognition. Even so, Fitch (2010) points out that '[t]his distinction between encapsulated and executive function defines a continuum, and from a multi-component perspective there is no reason to think

[34]However, note that Penn and Povinelli (2007a) argue that the behaviour of animals that use tools reflects little to no understanding of physical principles, based on experiments reported in Povinelli (2000). See also Seed and Byrne (2010).

that a complex function like language occupies a single point on this continuum' (82). This position helps to dissolve inevitable tensions between apparently inconsistent views that, for example, speech is encapsulated (Liberman, 1996), whereas semantics and pragmatics are executive (Fodor, 1983).

Language itself can (indeed, should) be understood in terms of modular composition. This view is consistent with that of Fitch (2010) that, instead of 'viewing language as a monolithic whole, I treat it as a complex system made up of several independent subsystems, each of which has a different function and may have a different neural and genetic substrate and, potentially, a different evolutionary history from the others' (17-18). There is some experimental evidence that such a network of components gives rise to language; for example, studies in neural lesions that show brain damage can affect one component of language, such as speech production, while leaving another component, such as comprehension, untouched.

The necessary communicative abilities that give rise to human linguistic abilities are shared with many other species; however, the ability to produce and interpret recursive structures is uniquely human (Hauser et al., 2002). If we are to take the idea seriously that there is no crucial component to human language or linguistic capacity, and that human language is composed of several different subcomponents, all of which are individually necessary and none of which are sufficient, then the natural question that arises is *how* these components might compose. While much of the study of language, from this perspective, is speculative, there are well-defined questions that can be asked, and individuals questions of this sort may have distinct ways of being answered. If we assume that the human capacity for language can be decomposed into a set of well-defined mechanisms that interact via interfaces, then we can begin to examine how such interfaces between individual components may 'hook up' in the first place—this is the notion of modular composition as it is described in Barrett and Skyrms (2017).

We have seen in this section that modular composition is a graded notion: transfer of learning is simpler than analogical reasoning, which in turn is simpler than modular composition more generally. Further, the cognitive requirements of each of these are also graded—less is required for transfer of learning than for analogical reasoning, etc. Finally, in contrast to linguistic compositionality discussed in Chapter 2, we have seen that each of these notions has plausible empirical precursors, and these too are graded: the more straightforward the modular-compositional process, the more prevalent it appears to be in nonhuman animals. Therefore, modular composition is sensitive to empirical data in a way that explanations from linguistic compositionality are not. Finally, I suggested in Chapter 2 that linguistic compositionality takes account only of the internal properties of language but ignores external restrictions or requirements such as cognition and social structure. In the next section, I suggest that modular composition is sensitive to these external constraints in the sense that a notion of modular composition ties together explanations of complex structure in communicative, cognitive, and social structures.

3.3 Two Asides

3.3.1 Language and Cognition

Lacking a sophisticated language is not necessarily an indication of a lack of sophisticated cognitive ability. As we have seen in this chapter and the last, many abilities that were previously thought to be unique to humans have been shown to exist in a wide range of species. This includes cross-modal association, episodic memory, anticipatory cognition, gaze-following, basic theory of mind, and tool use and tool construction, among other things (Fitch, 2010, 171-172).

There is good evidence that animals *think*.[35] Nonetheless, given that nonhuman animals do not have human-level linguistic capacities, as was discussed in Chapter 2, it seems strongly implied that the thought or consciousness of, e.g., nonhuman primates, is significantly different from that of humans.[36] Jackendoff (2007) argues that conscious thought in humans is revealed *mainly* in terms of *linguistic* imagery, which he takes to be correlated with *phonological* structure (as opposed to, e.g., semantics). This view is consistent with visual and proprioceptive imagery in thoughts (where 'proprioception' concerns the sense by which an individual perceives the position and movement of her body, including a sense of equilibrium and balance, and senses that depend on the notion of force (Jones, 2000; Wolff and Shepard, 2013).); however, 'image' need not (and should not) be understood in the sense of a visual image or a picture.

Conscious experience, or thought, in primates, then, will consist primarily in non-linguistic imagery—namely, thought in this sense will be manifested primarily in terms of visual, auditory, or proprioceptive imagery. In this sense, linguistic imagery constitutes an extra modality, in the same way that, e.g., echolocation in bats and dolphins, or olfactory awareness in dogs constitute an additional modality of awareness above and beyond visual, auditory, or proprioceptive awareness. It has been suggested that the critical difference in cognitive capacities between human and nonhuman animals arises from the modularity of core cognitive capacities in humans, which is lacking in other species. This is affected significantly

[35]See, for example, Köhler (1927); Byrne and Whiten (1988); Cheney and Seyfarth (1990, 2007); Hauser (2000); Povinelli (2000); Tomasello (2000), etc.

[36]The consensus in the scientific study of mind/brain is to find an explanation of conscious experience solely in terms of the physical activities of the brain—what Chalmers (1995) dubbed 'the hard problem' of consciousness. The so-called 'easy' problems concern things like explaining the reportability of mental states, the focus of attention, the integration of information by a cognitive system, the difference between wakefulness and sleep, etc. The *hard* problems, on the other hand, concern questions such as, why should physical mechanisms give rise to such a rich inner life in the first place (i.e., conscious *experience*). Majeed (2016) argues that there are in fact two distinct explanatory targets of the hard problem: (1) how physical processing gives rise to experience with a phenomenal character, and (2) how (why) phenomenal qualities are the way that they are. For the materialist, every aspect of conscious experience *must* have a physical correlate in the brain—the 'neural correlates of consciousness' (Crick and Koch, 1990, 1995). Our concern here, of course, is not to examine any physical, neuro-biological, or philosophical theory of mind in great detail; rather, we are interested in the notion that language enhances thought.

by language, and in particular, compositionality (Spelke, 2003). Further, humans have the ability to think in *linguistic* forms, driven by phonology and syntax.

Thus, regardless of what view one takes on precisely *how* language and cognition are related, it seems undeniable that they are. The position argued in the previous section implies that language affects cognition. This idea is by no means new. As Sapir (1921) argued:

> We must not imagine that a highly developed system of speech symbols worked itself out before the genesis of distinct concepts and thinking, the handling of concepts. We must rather imagine that thought processes set in, as a kind of psychic overflow, almost at the beginning of linguistic expression; further, that the concept, once defined, necessarily reacted on the life of its linguistic symbol, encouraging further linguistic growth. ... The instrument makes possible the product, the product refines the instrument. The birth of a new concept is invariably foreshadowed by a more or less strained or extended use of old linguistic material; the concept does not attain to individual and independent life until it has found a distinctive linguistic embodiment. (17)

This highlights the interplay between language and cognition, which was discussed previously; certain linguistic concepts depend inherently upon metacognition. For instance, propositional attitudes expressed by verbs like *believe*, *know*, *want*, *expect*, etc. involve relations to linguistic entities themselves.[37]

Clark (1998) lists the following six ways in which public tokens of communication might increase the possibility for complex cognition:[38] *memory augmentation*, as in written notes as a proxy for cognitive memory; *environmental simplification*, as in labelling for structuring or partitioning the environment in a computationally more straightforward way; *coordination and the reduction of online deliberation*, as in using language in complex collaborative

[37]Carruthers (2003) argues that 'the animal needs to have some way of telling when it is in a state of the required sort ... But this doesn't mean that the animal has to conceptualize the state *as* a state of uncertainty' (243). However, Hurford (2007) points out that this seems to force Carruthers to be committed to something like *telling* as opposed to *knowing*, which is even more linguistically entrenched.

[38]Bermúdez (2003) lists the same 6 items, in his own language. He points out that the function of the first four could be achieved by non-linguistic means: 'all [Clark] really offers is an account of how, given that we have language, we are able to engage in second-order cognitive dynamics—whereas what we need is an argument that second-order cognitive dynamics can only be undertaken by language-using creatures' (158)

problem solving; *taming path-dependent learning*, as in being able to communicate information about extraordinarily complex and abstract concepts (like those of quantum physics); *attention and resource allocation*, as in the way that linguistic phrases allow us to avoid the requirement for medium-term memory, thus freeing up resources for other tasks; and *data manipulation and representation*, as in extended intellectual arguments.

Many of these appear to be unique to humans; though, note that environmental simplification seems to be quite common in nature. Given the dependence of these *cognitive* processes upon linguistic capacity, it seems reasonable to say that the latter affects the former. However, increases in cognitive capacity may well further affect linguistic ability—hence the *co-evolution* of language and cognition. Hurford (2012, 499) suggests that many complex hierarchically structured non-linguistic activities, such as learning or cognition, are fruitfully *mediated* by language. Over time, however, they might get become *routine*—similar to processes of ritualisation—they may become automatic. In this sense, language precedes complex activity.

On the subject of (non-linguistic) action, Sellen and Norman (1992) write that

> There are two main modes of control: an unconscious, automatic mode best modeled as a network of distributed processors acting locally and in parallel; and a conscious control mode acting globally to oversee and override automatic control. Automatic and conscious control are complementary: the unconscious mode is fast, parallel, and context-dependent, responding to regularities in the environment in routine ways, whereas the conscious mode is effortful, limited, and flexible, stepping in to handle novel situations. (318)

Hurford (2012) points out that this is strikingly similar to the way that language works in humans. Namely, most of our speech is relatively fast and automatic (automatic control mode), whereas sometimes we need to express ourselves carefully, and we are more deliberate in our word-choice (conscious control mode).[39]

[39]Though, an important difference between routine physical activities—e.g., shaking hands, making coffee, eating with a spoon—and producing a sentence, also noted by Hurford (2012), is that 'the components of a

This view is re-iterated in Jackendoff (2007): 'inner speech and its capability for enhancing thought would have been automatic consequences of the emergence of language as a communicative system. In contrast, the reverse would not have been the case: enhancement of thought would not automatically lead to a communication system. In other words, if anything was a "spandrel" here, it was the enhancement of thought, built on the pillars of an overt communication system' (108). Further, the advent of language can help support analogical reasoning: '[i]f indeed relational language generally invites noticing and using relations, then the acquisition of relational language is instrumental in the development of abstract thought' Loewenstein and Gentner (2005, 348).

Jackendoff (2007, 105-106) lists several ways in which language enhances cognition. This is couched, again, in terms of his notion of linguistic imagery. His point is that this type of imagery allows for conceptualisations that are unavailable in other modalities. For example, though all sorts of imagery will enable one to attend to *tokens* in the environment, linguistic imagery allows one to attend to *types* as well. Thus, words can be used to pick out conceptual categories. Linguistic imagery allows tokens and types to be explicitly related via predication. Linguistic imagery allows one to attend to lack of information, what is not the case, other modalities (such as necessity and possibility), and temporal indices. Linguistic imagery allows one to make clear inferential, causal, and other such relations between situations. Finally, linguistic imagery enables one to distinguish between, or attend to, valuation features of percepts—e.g., familiarity versus novelty, expression of emotional or affective attributes, considering beliefs, intentions, desires, etc.

As such, what language affords is for one to attend to one's own consciousness explicitly. Thus, it allows one to be aware that one is thinking. Therefore, it is not by dint of the fact that humans have language that they can think, or that they are conscious. Instead, by dint

sentence (words and phrases) are arbitrary symbols for other things. Physical spoons or jars of coffee beans or handfuls of nettles have certain affordances, prompting further action. But the words jar, coffee, and beans do not have the same affordances. The word "jar" is not something you can put beans into; the word "beans" cannot be put into a jar' (504). See also Jackendoff (2007); Searle (1995).

of the fact that humans have language, they are 'better' at thinking (i.e., they can think in ways that would be impossible without language).

Note that, on the evolutionary account, this does not entail that nonhuman animals are inferior in any way to humans simply because they lack language. The idea is that language served some adaptive function in a social community, and it provided humans with an extra modality for thought, which allowed a new locus of attention. Thus, the existence of language is sufficient for the enhancement of thought. However, the converse implication does not necessarily hold: it is entirely possible that enhancement of thought could evolve by some other means, and it would not follow that a complex communication system would fall out of that.

Nonetheless, I note the following theme of this chapter, as a consequence of this view: the complex co-evolution of language ability and communication systems led to more sophisticated thought processes. Once this groundwork is laid, the addition of more advanced cognitive processes allows for further sophistication of linguistic processes. Thus, language emerged in the service of enhancing communication (Pinker and Bloom, 1990), not in the service of enhancing thought.[40]

3.3.2 Language and Social Structure

Jackendoff (2007) further points out that there are several parallels between language acquisition (or capacity for language) and social cognition in humans. These are reproduced in Table 3.1.[41]

[40] Pinker and Jackendoff (2005) argue against this latter position.
[41] See also Cavalli-Sforza (2001).

Language	Social Cognition
Unlimited number of understandable sentences	Unlimited number of understandable social situations
Requires combinatorial rule system in mind of language user	Requires combinatorial rule system in mind of social participant
Rule system not available to consciousness	Rule system only partly available to consciousness
Rule system must be acquired by child with only imperfect evidence in environment, virtually no teaching	Rule system must be acquired by child with only imperfect evidence, only partially taught
Learning thus requires inner unlearned resources, perhaps partly specific to language	Learning thus requires inner unlearned resources, perhaps partly specific to social cognition
Inner resources must be determined by genome interacting with processes of biological development	Inner resources must be determined by genome interacting with processes of biological development

Table 3.1: Parallels between language and social cognition

As has been repeatedly pointed out, both language and culture depend inherently upon the existence of a community for both functioning within a generation and transmission across generations.

The approach to studying language as a result of social structures and cognitive abilities from a biological and evolutionary perspective is not the predominant approach to the study of culture. Jackendoff (2007) points out that the prevailing attitude in (American) anthropology and sociology is to assume that humans are entirely a product of their culture, and it is meaningless to think that cognitive abilities influence culture.[42] Such a view is not only scientifically problematic, but it is socially problematic as well. Jackendoff (2007) points out

[42]See also Tooby and Cosmides (1992); Ehrenreich and McIntosh (1997); Pinker (2002) for a summary of how widespread and deeply entrenched such views are.

that this approach 'only mirrors the colonialist and imperialist attitudes of a century ago' (156).

Thus, social cognition is taken to be a 'core domain structure' in the sense of Spelke (2003). Indeed, Jackendoff (2007) argues that it is one of the *central* systems of cognition. Further, Fitch (2010) examines evolutionary developments leading from the LCA (of humans and chimpanzees) to humans, using a comparative approach comparing reproductive strategies in a variety of primates. He points out that hominids, including humans, diverged from other great apes with respect to reproductive behaviour—namely, paternal care and *alloparenting* (where additional kin other than the mother and father provide some parental responsibility to the young)—as a means for ensuring offspring survival when infant dependency, gestation time, and sexual maturity are comparatively long. By increasing the amount of care-giving, individuals can decrease birth spacing (or, put another way, decreasing birth spacing puts pressure on the father or other kin to provide care to ensure the survival of the infants)—typically, human females have babies every 2 to 3 years, whereas chimpanzee females have babies every 5 to 6 years (Lovejoy, 1981; Locke and Bogin, 2006). Fitch (2010) suggests that this created a novel social environment which was a crucial context for language evolution (since language facilitates coordination).

Fitch (2010) points out that, given the centrality of reproductive success to evolution, the rich social environment that existed in early hominids had the following three crucial impacts on subsequent human evolution: it selected strongly for coordination and cooperation among adults, both mother and father (Deacon, 1997) and other related individuals (Hrdy, 2005); it selected for infants and children able to engage with, and learn from, multiple members of this extended social group; and this enhanced sociality further selected for sophisticated social intelligence, both in terms of pragmatic inference in receivers and intentional information sharing by signallers (Fitch, 2010, 248). He further points out that it is unclear whether or not sexual selection amongst mates (as suggested by Deacon (1997); Miller (2001)), or

kin and natural selection amongst offspring (as indicated by Falk (2004); Fitch (2008)), or a combination thereof was primarily affected by this social environment.

Given the point that communication is an inherently *social* phenomenon which requires an underlying framework of cooperation, the social aspects of communication cannot be understated. Seyfarth et al. (2005); Cheney and Seyfarth (2007) suggest that social intelligence is a necessary (cognitive) precursor for human language. Jolly (1966) initially pointed out that many nonhuman primates live in complex social groups, and that this poses particular cognitive challenges. For example, social animals need to remember the identities of the other individuals in their groups in addition to the outcome of previous interactions with those particular individuals (Fitch, 2010). In terms of interactions with other group members that involve aggression, indirect observations can make for useful additions into an individual's model of its social group's dominance hierarchy (Bergman et al., 2003). Additionally, reconciliation has been observed after fighting (de Waal, 1989). Further, there exist subgroups within a social group, based on kin-relations, which allows for possible coalitionary behaviour (Bercovitch, 1988). Since individuals who live in groups compete with conspecifics (both in terms of in- and out-group members), Fitch (2010) points out that minor differences in cognitive abilities can lead to significant reproductive advantages.

As a result, social structures can be beneficial to fitness in a species. The advent of social structures puts pressure on communicative ability, but also cognitive capacity. Once communication is up and running, it can further affect cognition, which in turn affects communication. This view is mostly in line with that of Clark (1996), reiterated in Seyfarth and Cheney (2018): language is used for social purposes and consists in a type of joint action, and the study of language use is both a cognitive and a social science (14).

3.4 From Simple to Complex Communication

The complexity of a simple or complex communication system or language can be fruitfully thought of using the *Formal Language Hierarchy.*[43] Of course, we are concerned here with natural languages and natural communication systems rather than formal languages; thus, it is beyond the scope of our concerns here to delve too deeply into the annals of the formal language hierarchy. However, Hurford (2012) shows how the hierarchy is useful in categorising the distinctions in the complexity of various animal communication systems, especially as compared with human languages—for example, birdsong and whalesong can be accurately described as syntactically complex, but this raises the question of how complex they are. The formal language hierarchy, at least on its surface, gives a scale by which to measure the complexity of communication systems without relying solely on impression.

The core of the theory is the postulation of a hierarchy of possible language types, ordered by complexity. The ordering is given by containment, or subset, relations—e.g., everything that can be expressed by a language that is generated by a 'type-3 grammar' can be expressed by a language that is generated by a 'type-2 grammar', but the converse is not true. Therefore, languages generated by type-2 grammars are higher up on the hierarchy in terms of complexity than languages generated by a type-3 grammar.

This is particularly relevant to computer science, since the formal language hierarchy gives a way of classifying computer languages in terms of capacity for expression and distinguishes the types of automata capable of implementing such languages. The formal language hierarchy is also important to learnability theory in linguistics—this branch is likewise extremely formal and highly idealised. Thus, it would appear that there is no application for such technical machinery in simple communication systems; however, Hurford (2012) points out that for researchers who are interested in the evolution of language, 'the Formal Language

[43] See Post (1943); Chomsky (1956a,b,c, 1958, 1959a,b, 1962, 1963); Chomsky and Miller (1958); Chomsky and Schutzenberger (1963).

Hierarchy holds out the promise of a kind of easily definable *scala naturae* in terms of which it might be possible to classify the communication systems of various animals' (25), and further that this analogy is 'not totally crazy' (26)—though it is still too idealised to shed light on all the biological factors that might affect real biological systems.

In the context of the formal language hierarchy, a 'language' is a set of strings of elements (the vocabulary or lexicon)—for example, the set of well-formed sentences of French. Such sets are generally assumed to be infinite. A formal grammar is a set of statements (algorithms, rules) which generates all and only the grammatical sentences in a language. The *weak generative capacity* of a grammar simply refers to the language generated by that grammar; the *strong generative capacity* of a grammar refers to the language (set of strings) generated by that grammar in addition to the structural properties, or a structural description, of those strings.

Linguists and computer scientists are often interested in languages further up the hierarchy; however, in the context of the evolution of simple communication systems, we will have occasion to discuss simpler (less expressive) 'languages'.

> **Definition 3.6.** *First-Order Markov Language*
>
> A language is *First-Order Markov* if it can be completely described by a list of pairwise transitions between the elements of the language. The description of a first-order Markov language includes the meta-linguistic items START and END. At least one of the pairwise transitions must contain START as its first term, and at least one of the pairwise transitions must contain END as its second term. The set of transitions must include at least one path connecting *start* to *end*. There is no further restriction on pairwise transitions between elements. This is also referred to as a *strictly 2-local* language or a *linear* language. (Hurford, 2012)

Note that a First-Order Markov Language need not be finite. A First-Order Markov process cannot describe human languages since the transition table only describes transitions between one word and the next—i.e., it generates syntactically ill-formed strings, and there exist strings which depend, structurally, upon something other than the word immediately previous. However, certain complex systems of animal communication can accurately be modelled as a First-Order Markov Language. For example, the song of the white-crowned sparrow (*Zonotrichia leucophrys)* consists of up to five 'phrases' in a stereotyped order—call these **ABCDE**.

Rose et al. (2004) isolated white-crowned sparrow nestlings and tutored them with only pairs of phrases, such as **AB**, **BC**, and **DE**. They never heard an entire song. Nevertheless, when the birds' songs crystallised, several months later, they had learned to produce the whole intact song **ABCDE**. By contrast, birds who only ever heard single phrases in isolation did not eventually produce a typical white-crowned sparrow song. These researchers also gave other birds just pairs of phrases in reverse of normal order—for example, **ED**, **DC**, and **BA**. In this case, the birds eventually sang a typical white-crowned sparrow song backwards.

Definition 3.7. *State Chain Language*

A *State Chain Language* is one which can be fully described by a State Chain diagram. A State Chain diagram represents a set of 'states' (e.g., small circles in the diagram), with transitions between them represented as one-directional arrows. On each arrow is a single element (e.g. word, note, signal, etc.) of the language described. One particular state is designated as START, and one is designated as END. A sentence or song generated by such a diagram is any string of elements passed through while following the transition arrows, beginning at the START state and finishing at the END state. The transition arrows must provide at least one path from START to END. For example, in Figure 3.5, **iabcd** and **iabeafabcd** are valid sequences, whereas **iabcde** is not. There is no other restriction on the transitions

between states. This is also referred to as a *finite-state* language, a *regular* language, or a type-3 language. (Hurford, 2012)

Note that a state chain language also need not be finite, since it may contain cycles. This grammar can be expressed using regular expressions. Katahira et al. (2007) point out that 'Bengalese finch songs consist of discrete sound elements, called notes, particular combinations of which are sung sequentially. These combinations are called chunks. The same notes are included in different chunks; therefore, which note comes next depends on not only the immediately previous note but also the previous few notes' (441); See Figure 3.5.

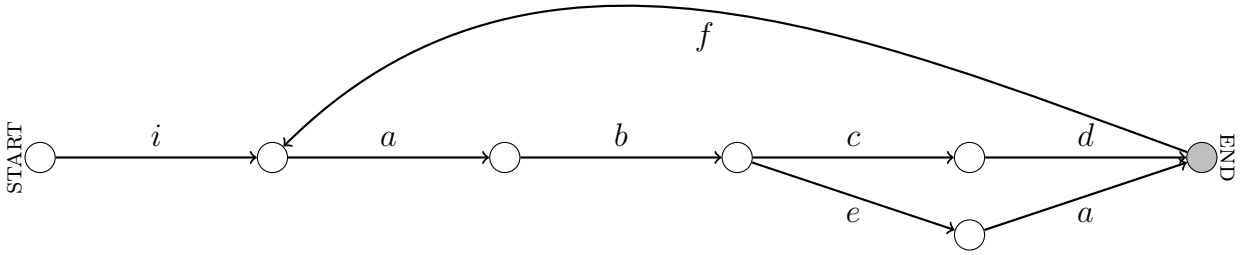


Figure 3.5: State chain diagram of a simple Bengalese finch song. The START state is labelled *start*. The filled circle is the END state, where it is *possible* (though not necessary) to finish the song. Since the note '*a*' appears in two different places, this song pattern could not be described by a first-order Markov transition table. See Katahira et al. (2007).

However, this characterisation crucially depends upon the representation of the note '**a**' actually being the same in both instantiations. If the second occurrence of '**a**' is actually a distinct note (though the distinction is perhaps imperceptible to humans), then this too can be well described by a First-Order Markov Language.

Definition 3.8. *Phrase Structure Grammar*

A *Phrase Structure Grammar* consists of a finite set of 'rewrite rules', each with one abstract, or nonterminal, element on the left-hand side of the arrow, and with any sequence of symbols on the right of the arrow. These latter symbols may be either actual 'terminal' elements of the language described (e.g. notes or words), or

abstract 'nonterminal' symbols labelling phrasal constituents of the language. Such nonterminal symbols are further defined by other rules of the grammar, in which they appear on the left-hand side of the arrow. One nonterminal symbol is designated as the START symbol. A string of words (notes, signals) is well-formed according to such a grammar if it can be produced by strictly following the rewrite rules to a string of terminal elements. Here, following the rewrite rules involves starting with the designated start symbol, then rewriting that as whatever string of symbols can be found in a rewrite rule with that symbol on its left-hand side, and then rewriting that string in turn as another string, replacing each nonterminal symbol in it by a string of symbols found on the left-hand side of some rule defining it. The process stops when a string containing only terminal symbols (actual words of the language or notes of the song) is reached. This is also referred to as a *context-free* language. (Hurford, 2012)

Definition 3.9. *Phrase Structure Language*

A *Phrase Structure Language* is one that can be fully described by a Phrase Structure grammar. (Hurford, 2012)

These definitions from the formal language hierarchy give rise to a continuum of complexity, illustrated by the following containment relations between these various definitions.

		Linear		Finite State		Context-Free		
Finite	$\subset$	**First-Order Markov**	$\subset$	**State Chain**	$\subset$	**Phrase Structure**	$\subset$	$\cdots$
		Strictly 2-Local		Regular		Type-2		

Table 3.2: Containment relations between varying levels of complexity in communication

Note, in particular, that even syntactically complex communication systems, such as bird-song and whalesong, are well modelled by First-Order Markov Languages, or State Chain Languages, whereas even Phrase Structure Languages are insufficient to describe natural language; see, for example, the discussion in Pullum and Gazdar (1982). Though it can be shown analytically that natural languages are neither regular nor context-free, Chomsky

(1957) doubted whether they could be adequately described by either type-1 or type-0 grammars either. Note that the emphasis of such an analysis necessarily depends on *syntactic structure* as primary. If these underlying assumptions were true, then one could see the sway of the saltationist view: the difference between a first-order Markov language and a state chain language is all-or-nothing. Either the current state depends solely upon the previous state, or it is path dependent.

I have suggested throughout these initial chapters that the emphasis on syntax is a mistake. However, even those who are sympathetic to the gradualist view have unduly placed far too much emphasis on the evolution of syntax as the primary gap-bridging step between simple and complex communication and language. This is not to downplay the importance of compositionality for natural languages; indeed, the commonly-held view that compositionality is a key characteristic of human language that is lacking in most if not all animal communication systems. However, we have seen that evolutionary precursors to linguistic compositionality are absent in nonhuman animals. Thus, it is unclear how focusing on the generative capacity of natural languages as the primary (or sole) target of explanation can lead to a solution.

We have seen that communication in the first place depends upon a context of cooperation. This was the central insight of Lewis (1969). Therefore, social structure is necessarily antecedent to communication. Once the appropriate (cooperative) social structures are in place, we have a nice explanation of how communication gets off the ground. This is the explanation given by the Lewis-Skyrms signalling game. However, it is a mistake to overemphasise the importance of linguistic compositionality, and therefore syntax, in further explaining how complex communication systems and language might evolve out of these simple communicative capacities. Instead, highlighting the importance of the relationship between communication and cognition, an alternative is to suggest that simple communicative modules might evolve independently

Therefore, rather than the *generative capacity* of languages constituting a key difference between language and communication systems, I suggest that it is the *reflexivity* of such communication systems that allowed a stark increase in complexity. This is more consistent with a gradualist perspective to the extent that a communication system can vary in terms of degrees of complexity—syntactic structures, as we have seen, tend to be all-or-nothing (this will be discussed in more detail in Chapter 4). For example, a 'yes/no' command might *refer* to a constituent action within the signalling game. However, a 'yes/no' command might evolve independently of any other signalling-game context. Thus, it can serve as a core module which later comes to hook up with different, independent signalling contexts via modular composition. Furthermore, this 'hooking-up' process, itself, is the result of evolutionary or learning dynamics on this story.

Once such complexity is exhibited, at a small scale, it may lead to a feedback loop between communication and cognition that, over time, gives rise to the complexity that we see in natural languages. Thus, this evolutionary story depends inherently on a concept of *bootstrapping*.

A full-blooded notion of so-called *Quinean bootstrapping* is developed in Carey (2009a), and associated articles.[44] The idea, which is the same one which I have been exploring, is that a child uses concepts and core knowledge at her disposal to learn or grasp a concept which she does not yet have.[45] For example, the appreciation of the infinity of integers requires (at least implicit) knowledge that when any set has one more added to it, the numerosity is always the next number in the pre-obtained list of counting words. Note that chimpanzees take as long to learn each new association of a symbol with set size as they take to learn the previous associations—implying that they learn each from scratch (Matsuzawa, 2009). Therefore, this sort of bootstrapping process may also be unique to humans. However, the

[44]See, for example, Carey (2004, 2009b, 2011b,a, 2014).

[45]This view has been both heavily praised—e.g., Shea (2009); Margolis and Laurence (2008, 2011); Piantadosi et al. (2012); Beck (2017)—and heavily criticised—e.g. Fodor (2010); Rey (2014); Rips et al. (2006, 2008, 2013); Rips and Hespos (2011).

empirical precursors are evident in nonhuman animals in a way that, e.g., the empirical pre-cursors of linguistic compositionality are not.

We have now seen how the modularity of signals might serve to create more complex signals. We have further seen a notion of modularity that is intimately tied to cognitive processes and social structures. In both cases, there is an apparent difference between animals and humans—both in signalling capacity, concerning the complexity of information-encoding signals, and in cognitive function concerning higher-order reasoning about the world. Given the apparent importance of modularity for cognitive and linguistic structures, it seems like a notion of modularity for the modular composition of the games themselves might provide a more fruitful, empirically grounded basis for future research.

In the next three chapters, I provide novel models that put this theory into practice.

Part II

Self-Assembly and Complex Communication

In order to say what a meaning IS*, we may first ask what a meaning* DOES*, and then find something that does that.*

– David Lewis, *General Semantics*

In Part I of this dissertation, I argued that reflexivity constitutes a salient distinction between language and communication.

In the subsequent three chapters, I provide several models that show how, and under what circumstances, reflexivity might give rise to more complex communicative dispositions.

In Chapter 4, I discuss further the possibility of modelling compositionality within the signalling-game framework. In Chapter 2, I suggested that such models are not sensitive to empirical data; however, it is still possible that an evolutionary explanation of compositionality might be salvaged. In order for such an explanation to be genuinely gradualist, it is necessary that there be a notion of *degrees* of compositionality. If this is true, then it should be possible to provide a model for measuring how compositional a communication system is. This gives rise to further pressing problems for any gradualist account, specifically within the signalling-game framework. I show that if it is possible to give a model of degrees of compositionality of complex signals, then this depends upon the reflexivity of the structural components of the signalling game; thus, compositionality is strictly secondary to reflexivity.

In Chapter 5, I present a model of learning that varies the reward for coordination in the signalling game as a function of the agents' actions. The model takes advantage of the type of communicative bootstrapping processes that were suggested in Part I—namely,how previously evolved capacities might help to more efficiently evolve new capacities, via reflexivity. This works by means of a pre-evolved sub-game for *correcting* behaviour that resulted in the wrong action being chosen; I refer to this model as the *correction game*.

In Chapter 6, I use basic logical operators as a test bed for the notion of modular composition previously discussed—in particular, with an emphasis on reflexivity. Specifically, I show how modular composition can help agents to evolve complex signalling more efficiently than the simple (atomic) signalling game framework, and I discuss the circumstances under which

these results hold. This chapter builds on the previous work in the theory of self-assembling games Barrett and Skyrms (2017).

Chapter 4

Less Is More: Degrees of Compositionality

We used to think that if we knew one, we knew two, because one and one are two. We are finding that we must learn a great deal more about 'and'.
— Sir Arthur Eddington

In Chapter 2, we saw several examples of how communication is ubiquitous in nature. I further surveyed suggestions for the salient difference(s) between communication, as it occurs in nature and language. Many simple systems of communication in non-human animals are well-modelled by the signalling game. In particular, the signalling-game framework gives us a plausible picture of how meaningful communication can initially emerge—this is especially true in cooperative social groups. While explaining how communication can get off the ground was a significant achievement, there is a chasm to be filled between these various phenomena to explain how distinctly human linguistic capacities could have evolved over and above simple communication.

In Chapter 2, we further saw that a critical difference between language and communication consists apparently in the generative capacities of the former. Some researchers, taking account of the idea that ontogeny recapitulates phylogeny,[1] look toward language acquisition in human children as empirical evidence of evolutionary stages moving from simple holophrastic signalling to a two-word phase, and eventually to fully syntactic, compositional language.[2] Shettleworth (2012) highlights that (cognitive, communicative, etc.) differences between species are significantly less pronounced when comparing young animals (either human or non-human), than when comparing abilities between non-human animals and adult humans. Therefore, any comparison of species' abilities requires a comparison of developmental *trajectories* in addition to the species-specific mechanism (or mechanisms) involved in the development of said capacities.

It is a fact that every speaker of a natural language needs to master an unlimited number of novel expressions in a relatively short period. Human children generally learn to speak in grammatically correct sentences by three years of age; however, it is argued that they are not exposed to rich enough data in their linguistic environment to acquire every feature of their language—this is referred to as the *poverty of the stimulus* (Chomsky, 1980a).[3]

As we have seen, most researchers hold that the openness, (productivity, generative capacity, hierarchical structure) of natural languages is a *key* distinguishing feature. For example, arbitrary, meaningless phonemes can be combined in a potentially infinite number of ways to create meaningful morphemes;[4] similarly, sounds combine to form words, and words combine

[1]This (controversial) suggestion, which was introduced in the 19th century by Ernst Heinrich Haeckel, is largely rejected by biologists in the current day—at least as being a fundamental principle of evolution. Nonetheless, Richardson and Keuck (2002) discuss whether this can be applied up to a degree in biological evolution. In the evolution of language literature, there has been a renewed interest in this principle (Bickerton, 1990; Givón, 2002a).

[2]See Progovac (2015).

[3]This is often used as evidence for the *Universal Grammar*. The poverty of the stimulus is related, conceptually, to the 'new riddle of induction' (Goodman, 1965), which is a successor to Hume's problem of induction (Hume, 1739, 1748).

[4]This phenomenon is referred to as *duality of patterning*, or sometimes *double articulation*; see, Hockett (1958, 1960a,b); Hockett and Ascher (1964); Ladd (2012).

to form phrasal expressions and sentences. Thus, with a finite lexicon and a finite set of grammatical rules, natural languages 'contain' a potentially infinite number of unique, semantically meaningful, and syntactically well-formed expressions.

To account for these sorts of phenomena, researchers frequently point to a principle of compositionality, which is typically formulated as follows (Partee, 1984; Kamp and Partee, 1995; Szabó, 2012):

> **Definition 4.1.** *Principle of (Linguistic) Compositionality*
> The meaning of a compound [complex] expression is a function of the meaning of its parts [constituents] and the ways in which they are combined [composed].

Note that this formulation is 'theory-neutral' in the sense that it requires and entails no specific commitments about, e.g., what 'meanings' or 'ways of combining' might actually be. This principle arises in virtually any field of study concerned with language and meaning—notably, philosophy, logic, computer science, psychology, and semantics of natural language (Janssen, 2012).

This principle serves to explain many observable facts about human language—including its productive and interpretative flexibility, and its systematicity and learnability, among others.[5] The explanatory power of the assumption that languages are indeed compositional is apparent:

> It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to parts of a sentence, so that the structure of the sentence serves as the image of the structure of the thought. (Frege, 1923, 1)

[5]See Pagin and Westerståhl (2010a,b). For an historical overview of this principle in the context of natural languages, see Janssen (2012); Hodges (2012).

In Chapter 2, I further surveyed several models that have been suggested in recent years which try to grapple with these problems using the signalling-game framework (Nowak and Krakauer, 1999; Barrett, 2006, 2007, 2009; Franke, 2014, 2016; Steinert-Threlkeld, 2016). It was suggested there that such models are not empirically well-grounded, as there is scant evidence that compositionality occurs in nature—at least in a communicative context. However, there are other, more philosophical, reasons why such models are not adequate.

In chapter 3, I suggested that *reflexivity*, rather than *compositionality*, and the role that it plays in connection with modular composition is a better target of an evolutionary explanation bridging the gap between simple communication systems and language.

In this chapter, I outline two further problems that arise in modelling compositionality using the signalling-game framework. On the one hand, these models often (if implicitly) take compositionality qua linguistic compositionality (Definition 4.1) as their target for an evolutionary explanation. This gives rise to significant complications to the extent that linguistic compositionality is rife with conceptual difficulties. Thus, by presupposing that the theoretical target of our evolutionary explanation is equivalent to this robust notion of compositionality, these models inherit all the philosophical baggage associated with such a concept. On the other hand, these models fail to consider the role-asymmetry of the sender and receiver in the signalling game, and thus fail to capture how compositionality might be beneficial for communication. To surmount these problems, I suggest that it is more fruitful to build a notion of compositional *signalling* bottom-up, as it were. This requires, first, demarcating atomic and complex signals, and, second, providing a precise specification of what it would mean for complex signals to be compositional—as opposed to, e.g., merely combinatorial—in the first place.

Here, I highlight why the models discussed in Chapter 2 are neither sufficient nor conceptually adequate for explaining the evolution of compositionality in this sense. Specifically, Section 4.1 details the first point concerning linguistic compositionality as the target of ex-

planation, and Section 4.2 highlights the role-asymmetry inherent in the signalling-game framework, and why it is essential for an understanding the possible benefit of compositional structures in simple communication systems.

In Section 4.3, I turn to a notion of information transfer in the signalling game and highlight the usefulness of understanding meaning in an information-theoretic context. In particular, because the information conveyed by a signal about the states and the actions are demarcated, this allows us to maintain sensitivity to the role asymmetries of the sender and receiver; further, since this measure builds a concept of compositionality from the bottom up, as it were, it avoids inheriting the conceptual problems of a pre-theoretic idea of compositionality. In Section 4.5, we examine information in nature and discuss the relationship between this measure of compositionality and the notion of modular composition that was outlined in Chapter 3.

Steinert-Threlkeld (2017) points out that the 'status' questions that often surround philosophical discussions of compositionality—e.g., How can we make precise the meaning of compositionality? Is the principle of compositionality true?—are philosophically interesting, but we can further ask procedural questions: *Why are natural languages compositional? What role does compositionality play in the theory of communication? Does composition itself increase semantic complexity?*[6] Any sort of analysis of a complex language inherits the complexity of the language itself. As a result, it is apt to abstract away these complexities and look at a simple model for communication. Under these circumstances, what does compositionality look like?

[6]For first-blush answers to these questions, see Steinert-Threlkeld (2017).

4.1 Linguistic Versus Communicative Compositionality

The first problem in current evolutionary explanations of syntactic compositionality arises from an equivocal use of the word 'compositionality'. In each case, what it means for a signal to be compositional is presupposed and often left undefined. It appears that the pre-theoretic assumption consists in 'compositionality' just being equivalent to the notion of *linguistic* compositionality, as given in Definition 4.1. This is problematic for at least two reasons.

On the one hand, Szabó (2012, 2017) points out that this formulation gives rise to several pressing questions. For example, does 'is a function of' mean 'is determined by'? Or, does it mean that there is a function to the meaning of a complex expression from the meanings of its constituents and the way they are combined? The first of these is entailed by the second, but not vice-versa; thus, there is a real distinction to be made here. Further, are we concerned with the meanings that the constituents have individually or the meaning that they have when taken together? Szabó (2012) suggests that the various ambiguities inherent in this formulation combine to give eight *distinct* readings of what compositionality is. As a result, if we implicitly take *linguistic* compositionality as the target of an evolutionary explanation of compositional *communication*, this explanation necessarily inherits all the complexity and ambiguity that surrounds this concept.

On the other hand, implicitly taking linguistic compositionality as the target of one's evolutionary explanation runs afoul of the *gradualist* perspective necessary for an adequate evolutionary account, which posits an intermediate step (or intermediate steps) between the 'one-word stage' of language development (which is what the basic signalling game models in a variety of contexts), and full-blown compositional syntax. The gradualist view, as we have seen, posits a *protolanguage* between these evolutionary stages in linguistic develop-

ment. In almost every case, the explanatory target of protolanguage is proto-*syntax*.[7] To explain the emergence of linguistic compositionality, we would need first to explain how some proto-compositional precursor might arise.

This sentiment is present in the preceding accounts of compositional signals; however, the actual *proto-compositional* target is never made explicit. Franke (2016), for example, does spend some time discussing compositionality versus proto-compositionality, and the need for a gradualist perspective; however, when he outlines his desiderata, he refers to the agents' abilities to react to novel stimuli in a 'compositional-like' way but does not make explicit in what this consists.

The saltationist view, as we have seen, posits that some 'catastrophic change' led to a leap from non-language to language. On this view, the language faculty emerged relatively late in human development. There is no protolanguage preceding language; instead, (in the minimalist programme of Chomsky (1995)) the main operation—*Merge*—materialised spontaneously and all at once, independent of any simpler precursor (Berwick et al., 2011; Di Sciullo, 2013).[8]

On a gradualist view, Jackendoff (1999, 2002) argues for *proto-Merge*, which derives flat concatenation/adjunction structures rather than genuine hierarchical structures like *Merge*. Some theories of protolanguage focus on the development of subject-predicate relations (Gil, 2012), whereas others take protolanguage to be limited to the concatenation of predicates only.

Thus, on the gradualist view, the order of development is given as follows:

$$\text{Pre-Syntactic Stage} \rightarrow \text{Proto-Syntactic Stage} \rightarrow \text{Modern Syntax.}$$

[7] See the discussion in Progovac (2019).

[8] Recall that *Merge* is a recursive binary operation that derives hierarchical binary branching structures. It takes two syntactic objects, α, β, and forms a new object $\textsc{merge}(\alpha, \beta) = \Gamma = \{\alpha, \beta\}$

The pre-syntactic stage is sometimes characterised as a 'one-word stage', wherein signals are holophrastic; the proto-syntax stage is sometimes described as a 'two-word stage'. Note, however, that protolanguage is usually *defined* as a communication system which *lacks syntax* (Bickerton, 1990). Therefore, even if the gradualist posits an intermediate stage between pre-syntax and modern syntax, this still leaves a significant gap between protolanguage and language.

The problem is that compositionality appears to be a binary property of language: a communication system either is compositional, or it is not. By analogy, a syntax either is hierarchical or not—the leap from proto-Merge (which is a flat structure) to Merge (which is hierarchical) is still a leap. Similarly, the move from a finite set to an infinite set is still a leap: positing an intermediate stage does not help to bridge the gap between these cardinalities. On this last picture, we have a structure that is analogous to the posited development of language (specifically syntax) via (non-syntactic) protolanguage. Suppose we posit the following explanation of the 'gradual' development of an infinite set out of finite sets:

$$\text{Singleton Stage} \rightarrow \text{Binary Stage} \rightarrow \text{Infinite Stage.}$$

It seems obvious, in this case, that there is still a significant explanatory gap to be filled. I suggest that the same criticism holds of protolanguage. Berwick and Chomsky (2011) highlight that 'there is no rationale for postulation of such a system: to go from seven-word sentences to the discrete infinity of human language requires the same recursive procedure as to go from zero to infinity' (31). If the focus of an evolutionary account is syntax, then the gradualist implicitly posits a significant leap from non-compositional, pre-syntactic protolanguage to full-blown compositionality.

In this chapter, I want to examine the possibility of salvaging compositionality from a gradualist perspective. However, given the problems that prior analyses give rise to, highlighted above, it is apt to abstract away these complexities and look at a simple model of com-

positional *signals*. Under these circumstances, what does 'compositionality' look like? Is it possible to fill in some grey area between non-compositional communication and compositional language? Answering these questions is a requirement for clearly stipulating the conditions under which a complex *signal* or a simple system of communication might be taken to be compositional. In addition to helping to specify what it *means* for a system of communication to be compositional, this mode of analysis allows us to examine the evolutionary contexts under which we might expect something like compositionality to arise, thus helping to bridge the explanatory gap between the evolution of simple systems of communication and human-level linguistic abilities.

When we understand the problem in this way, it becomes clear that any talk of whether or not animal communication systems are compositional is misdirected: such talk already presupposes that we understand what it means for a complex signal to be compositional. By taking a 'bottom-up' approach to compositionality, we might be able to come to some clear understanding of this sort of phenomenon to move forward with explaining how such dispositions might evolve in the first place, and how they might further evolve to a richer degree of complexity.

It is undeniable that examples of complex signals exist in nature. However, there is disagreement as to whether these complex signals are compositional or not. In each of the cases surveyed in Chapter 2, a presupposition of what it means for a signal to be compositional seems to be inherited from a pre-existing conceptual understanding of linguistic compositionality. As a result, prior theoretical biases seep into the discussion of what counts as a compositional signal in the first place. As such, it appears that, at best, 'compositionality', as it is discussed in the literature on evolutionary compositionality, succumbs to a covert polysemy; at worst, it might be an 'essentially contested concept' (Gallie, 1955). Thus, providing clear and coherent necessary and sufficient conditions for a complex signal to be compositional should be the preeminent target for future work in the evolution of composi-

tional communication. Conceptual clarity in this definition will have downstream benefits in building models that explain the evolutionary emergence of this sort of target phenomenon.

As of yet, we lack a coherent and concrete way of saying why a complex signal ought to be considered compositional, as opposed to atomic or merely combinatorial. Furthermore, this problem directly mirrors significant contention within the biological and linguistic literature on whether certain species' communication systems are indeed compositional—certain biologists might suggest that a communication system in nature is compositional, and then certain linguists might suggest that it is not. For example, Zuberbühler (2002) suggests that Campbell's monkeys have syntactically complex communication systems. This and related papers are often cited as evidence of compositionality in nature; however, Hurford (2012) univocally holds that no communication system outside of human language is compositional. Might it not be the case that the latter implicitly defines compositionality as linguistic compositionality, whereas the former has in mind a more simplified notion of compositionality?

It appears that much of the debate is, in essence, a matter of talking past one another due to a lack of clear and coherent definition of the constitution of compositional signals. This nebulousness is only compounded by complexities arising from understanding compositionality, conceptually, in the setting of natural language—i.e., by presupposing an understanding of compositionality for natural languages and attempting to appropriate this concept for simple communication, the discussion inherits all of the complications that arise from considerations of compositionality in natural languages.

4.2 Desiderata for Compositional Signals

Let us try to set aside any pre-theoretic notion of what compositionality is with respect to language, and what it entails or requires to *be* compositionality. Instead, in this section, I will

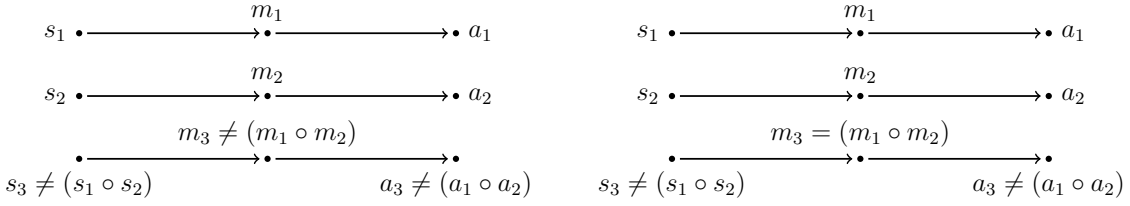
try to build a notion of proto-compositionality from the bottom up. In this way, we can avoid the theoretical complexity that is associated with a full-blown notion of compositionality, while simultaneously making explicit what proto-compositionality is supposed to be. I will suggest two main desiderata. These happen to be consistent with an intuition about the properties of linguistic compositionality that make it desirable; however, I do not presuppose these properties, but show *why* they might be beneficial from an evolutionary point of view—namely, in terms of efficiency.

The first of these is lexical composition. As we will see, this is the notion that is usually targeted in evolutionary accounts of compositionality. However, we will also see that a concept of systematicity is desirable for a proto-compositionality to be genuinely effective—this is the notion of compositionality that is usually targeted by researchers in machine learning who focus on *emergent communication*. This analysis gives rise to a further problem in evolutionary accounts of compositionality: they ignore the role-asymmetry that is inherent in the signalling game, focusing solely on syntactic combination, which provides benefit only to the sender. Thus, in the very least, any account of proto-compositionality is going to require figuring this role asymmetry.

4.2.1 Lexical Composition / Combination

There is an apparent adaptive advantage for combinatorial capacities in a communication system: namely, fewer elements need to be stored in memory to produce the same possible number of messages, thus allowing for more efficient communication; see Nowak and Krakauer (1999); Nowak et al. (2000). To avoid conflating this notion of syntactic composition with the type of syntactic composition required in *linguistic* compositionality (Definition 4.1), I will refer to this as *lexical combination*. How can we demarcate (lexical) combinatorial signals from atomic signals?

Scott-Phillips and Blythe (2013), try to differentiate 'combinatorial' or 'composite' communication systems from 'non-combinatorial' or 'non-composite' (i.e., atomic) communication systems. A signalling system, on their account, is *composite* if it contains at least one pair of composite signals—where the combination of two signals, $m_k = (m_i \circ m_j) \in M$, is produced in at least one non-composite state, $s_k \neq (s_i \circ s_j)$; see Figure 4.1.



(a) Non-Composite Communication System (b) Composite Communication System

Figure 4.1: Composite versus non-composite communication systems. m_1 and m_2 are atomic signals. In (a), m_3 is atomic, and unique from m_1 or m_2, and so is not a combination of these. In (b), m_3 is a combination of m_1 and m_2 (e.g., concatenation). However, it produces a unique action from either of its parts, so it is 'composite' (on their definition).

More specifically, Scott-Phillips and Blythe (2013) suggest the following differentiating features for, what they call, combinatorial and non-combinatorial communication systems. On their model, there are '*default* settings', $s_\varnothing \in S, m_\varnothing \in M$, and $a_\varnothing \in A$, which are orthogonal to all other members of their respective sets. Except for $s_\varnothing$ and $m_\varnothing$, states and messages can be combined with other states and messages. These combinations are denoted by $(s_i \circ s_j) \in S$ and $(m_i \circ m_j) \in M$, respectively. Note that combination is *commutative* on their model, so that, e.g., $(m_i \circ m_j) = (m_j \circ m_i)$. They define non-composite, pseudo-composite, and fully-composite pairs of signals as follows.

1. *Non-Composite.* A pair is *non-composite* if the composite of the two signals is produced only in composite states, and it, in turn, yields the default action—i.e. there is no $s_k \neq (s_i \circ s_j)$ such that $m(s_k) = (m_i \circ m_j)$, and $a(m_i \circ m_j) = a_\varnothing$

2. *Pseudo-Composite.* A pair is *pseudo-composite* if the composite of the two signals is produced only in composite states, and it, in turn, yields a non-default action—

i.e. there is no $s_k \neq (s_i \circ s_j)$ such that $m(s_k) = (m_i \circ m_j)$, while at the same time $a(m_i \circ m_j) \neq a_\varnothing$.

3. *Fully-Composite.* A pair is *fully-composite* if the composite of the two signals is produced in at least one non-composite state, and it, in turn, yields a non-default action—i.e. there exists some $s_k \neq (s_i \circ s_j)$ such that $m(s_k) = (m_i \circ m_j)$ and $a(m_i \circ m_j) \neq a_\varnothing$.

A combinatorial communication system, then, is a system that includes at least one pair of fully-composite signals.

In terms of signalling, the idea is that two or more atomic signals might be combined to form a third composite signal. This composite signal has a different effect than just the sum of the individual signals. Though Scott-Phillips and Blythe (2013) use the words 'composite' and 'combinatorial' and do not mention compositionality *per se*, they clearly have in mind something like compositionality: they claim that there exists 'one extreme exception to the norm of non-combinatorial communication: human linguistic communication' (5). This is despite the fact that they cite the putty-nosed monkey signalling system as an example of combinatorial signalling.

As such, it is not apparent how their differentiation of combinatorial and non-combinatorial communication can be used to clarify what we mean by compositional communication.

Recall from Chapter 2, the signalling system of putty-nose monkeys is composite in this very sense. The presence of eagles elicits a 'pyow' signal, which in turn elicits the action *climb down a tree*; the presence of leopards elicits a 'hack' signal, which in turn elicits the action *climb up a tree*. However, a third, unknown, context (though one which is different from both the presence of leopards and the presence of eagles) elicits the combinatorial 'pyow-hack' signal, which in turn elicits the action *move to a new location* (Arnold and Zuberbühler, 2006a,b, 2008).

This model captures a similar notion of *syntactic* combination that was apparent in the syntactic signalling game (Barrett, 2006, 2007, 2009). (Except, since Scott-Phillips and Blythe (2013) stipulate that (atomic) signal order does not matter in their model, the meaning of $(m_1 \circ m_2)$ is equivalent to the meaning of $(m_2 \circ m_1)$.) Thus, their model fails to capture sensitivity to the syntactic structure that is apparent in complex signals in, e.g., bird song and whale song. Barrett (2006, 2007, 2009) is sensitive to signal order, but we saw in Chapter 2 that the complex signals get *interpreted* atomically. The model suggested by Scott-Phillips and Blythe (2013), for the same reason, can be construed atomically—the meaning of a fully composite signal pair need not have anything to do with the meaning of its parts. A separate notion of systematicity captures this.

4.2.2 Systematicity / Generalisation

Although signal combination is an obvious target for an evolutionary explanation of compositional signals, this cannot, itself, give rise to any form of proto-compositionality. The reason for this, as has been highlighted by Franke (2014, 2016); Steinert-Threlkeld (2016) for Barrett's syntactic signalling game, is that it does not capture a notion of *generalisation* that is required for compositionality.[9] For a receiver to *interpret* a complex signal compositionally, she must be able to *decompose* the meaning of the signal based upon the meaning of the parts. By example, if the receiver knows the meaning of 'pick up x' and the meaning of 'the book', but not the meaning of 'put down x', then she might understand the command 'pick up the book', though she does not understand the meaning of 'put down the book'. Even so, she may still understand that the latter expression has *something* to do with the book.

Syntactic signalling, which accounts for lexical combination alone, only offers a benefit to the sender, insofar as the sender can communicate more with a smaller lexicon (and a small

[9]This is highlighted in Brochhagen (2015).

set of rules for combining lexical items). However, the receiver must still learn to interpret each complex signal atomically.

Recent work in machine learning highlights a problem for learning compositional linguistic structures. Neural networks are the 'workhorse' of natural language comprehension and generation—Bahdanau et al. (2018) highlight that neural networks play a significant role in machine translation Wu et al. (2016) and text generation (Kannan et al., 2016) in addition to exhibiting state-of-the-art performance on several benchmarks, including *Recognising Textual Entailment* (Gong et al., 2017), *Visual Question Answering* (Jiang et al., 2018), and *Reading Comprehension* Wang et al. (2018). However, training an AI to emerge compositional communication in an artificial context runs into parallel problems as giving an evolutionary account of emergent compositionality in a natural setting. Whereas evolutionary explanations tend to focus on the syntactic side of the problem—and thus hit upon the roadblocks described in Steinert-Threlkeld (2020)—computer scientists working in machine learning tend to focus on the generalisation aspect of compositionality.

The idea of systematicity, introduced by Fodor and Pylyshyn (1988), is that 'the ability to entertain a given thought implies the ability to entertain thoughts with semantically related contents'.[10] The problem with neural networks is that they latch on to statistical regularities in datasets. In a synthetic instruction-following task (Lake and Baroni, 2017), the agent does not learn a generalisation for composing words. Thus, when the AI is trained on the commands 'jump', 'run twice', and 'walk twice', it subsequently fails when asked to interpret 'jump twice' (Bahdanau et al., 2018).

This is precisely the problem that we run into for the evolution of compositional syntax. It turns out that machine learning recommends a similar process as that which I have suggested in this dissertation: To improve performance on generalisation, researchers are adding

[10]Whether or not, e.g., a connectionist model of cognition can account for systematicity has been the subject of a long debate in cognitive science; see, for example Fodor and Pylyshyn (1988); Smolensky (1987); Marcus (1998, 2003); Calvo and Colunga (2003).

modularity and structure to their designs (Andreas et al., 2016; Gaunt et al., 2016). In the case of the Neural Module Network paradigm, a neural network is assembled from several *modules*, each of which is supposed to perform a particular subtask of the input processing.[11]

Compositionality in communication will require some notion of combination, but this must account for both the production *and* interpretation of complex signs. For a system to be fully compositional, the sender needs to be able to construct a sign with some internal structure, and the receiver must be sensitive to that structure:

> A communication system that is genuinely complex and combinatorial is one in which rich combinatorial structure figures into the rules on both sides of the signs, rather than a system in which simple nominal signs are produced but complex interpretations are possible given the social context, and rather than a system with very complex production but where most of the complexity is insignificant to interpreters. (Godfrey-Smith, 2018, 120)

Recall from Chapter 3 that systematicity requires a form of analogical reasoning (Gentner and Toupin, 1986).

4.2.3 Moving Forward

I have suggested that the evolutionary explanations offered thus far fail to give a plausible account of how compositionality might arise. In Chapter 2. I suggested that these models are not sensitive to empirical data. In the present chapter, we have seen two substantive theoretical arguments for this claim. On the one hand, there is an inherent complexity in the meaning of *linguistic* compositionality, which is inherited by these models to the extent that they (at least implicitly) take *this* as their target, as opposed to a simpler proto-

[11] Bahdanau et al. (2018) note that although this modular approach is intuitively appealing, 'widespread adoption has been hindered by the large amount of domain knowledge that is required to decide or predict how the modules should be created (parametrisation) and how they should be connected (layout) based on a natural language utterance'. See also, Andreas et al. (2016); Johnson et al. (2016, 2017); Hu et al. (2017).

compositionality. The latter, to the best of my knowledge, is not explicitly defined anywhere. This gives us a target for a model: to provide a 'bottom-up' definition of what it means for a complex signal to be compositional in the first place. This requires explicitly defining a notion of compositional signalling (a sort of proto-compositionality) which is distinct from, and significantly more straightforward than, *bona fide* linguistic compositionality.

On the other hand, these evolutionary explanations are not sensitive to the asymmetric roles of the sender and receiver in the simple signalling-game framework. This provides a restriction for our target definition, which must account for role-asymmetries between the sender and the receiver. Compositionality is only fully effective to the extent that it is possible to *productively* compose simple signals systematically, on the part of the sender, and also to effectively *decompose* those complex signals to understand the meaning systematically on the part of the receiver. I will here refer to this former notion as *syntactic* compositionality, and I will refer to this latter notion as *semantic compositionality*.

We might begin by noting a distinction between atomic and complex signals, as follows.

> **Definition 4.2.** *Atomic Signal.*
> A signal is atomic if it is a holistic unit. i.e., it cannot be decomposed into simpler parts.

> **Definition 4.3.** *Complex Signal.*
> A signal is complex if it is not atomic.

We further note that complex signals may be compositional or not. This is the key distinction that needs to be fleshed out, moving forward. On the face of it, we might suggest the following definition:

> **Definition 4.4.** *Compositional Signal.*
> A complex signal is *compositional* if it is both lexically and semantically compositional.

To be clear, let us refer to the compositionality that is given by a compositional signal as *proto-compositionality*; this is compared to the full-blooded *linguistic compositionality* of Definition 4.1. A signal is thus compositional only to the extent that it is beneficial to both the sender and the receiver. The notion of what it means for a *signal* to be (proto-)compositional, as given in Definition 4.4, take account of both *lexical* composition, in the sense of syntactic combination outlined in Section 4.2.1, and semantic composition, in the sense of systematic generalisation given in Section 4.2.2.

Therefore, all we require is a clear definition for each of these notions. Note that defining compositional signals in this way already takes account of the role-asymmetries of the sender and receiver. Further, this definition of compositional signalling will capture the desired pre-theoretic properties that were argued for in Section 4.2. If we are successful, we should be able to say that the models of Barrett (2006, 2007, 2009); Scott-Phillips and Blythe (2013); Franke (2014, 2016); Steinert-Threlkeld (2016, 2017) are syntactically compositional, but not semantically compositional and thus not (proto-)compositional.

How might we obtain definitions for lexical and semantic compositionality? I suggest the answer lies in the *informational content* of the signals. Before offering a clear definition of syntactic and semantic composition, in the subsequent section, I survey information theory and the role that it is taken to have in the meaning of signals.

4.3 Information and Meaning

The information-based approach to communication is reasonably widely held;[12] However, this view suggests that the selection of signals is driven by the information that the signals carry rather than the fitness benefits that the sender and receiver earn from coordination.

[12] see, e.g., Otte (1974); Zahavi (1987); Bradbury and Vehrencamp (2011); Hauser (1996); Seyfarth et al. (1980a).

Opponents to the information-based approach generally hold that communication should be defined in terms of the influence of manipulation on a receiver by a sender;[13] but, this view fails to distinguish communication from any other form of influence or manipulation in nature.[14]

Part of the problem is that there are several different, possibly unique and possibly inconsistent, uses of 'information' in the literature. These might include, for example, Shannon information, Shannon entropy, quantitative information, colloquial information, semantic information, relative entropy (Kullback-Leibler divergence), mutual information, etc., in addition to several concepts related to the exchange or movement of information. As a result, Scarantino and Piccinini (1993); Piccinini and Scarantino (2011); Scarantino (2013) argue that 'information is a mongrel concept comprising a variety of different phenomena under the same heading' (Scarantino, 2013, 64). We might further highlight the apparent distinctions between, e.g., information transfer, information gathering, information flow, an information channel, information encoding/decoding, etc. Information is often explained by way of metaphor, and Horn and McGregor (2013) suggest that a large part of the confusion is caused by taking the metaphor too seriously. In the 'conduit' metaphor, communication is a *flow* of information from the sender to receiver, as in water flowing through a pipe. In this case, communication begins with encoding, whereby the sender transfers information into a signal; the signal *carries* the message to the receiver; the receiver decodes the meaning of the signal.

This problem compounds because the technical, mathematical definition of information (i.e., Shannon information), is often conflated with an intuitive notion of information. As such, a vague intuitive concept is used as if dressed up with the rigour and clarity of mathematics.

[13] See, e.g., Dawkins and Krebs (1978); Owings and Morton (1998); Maynard Smith and Harper (2003); Owren et al. (2010).

[14] See the discussion in Scarantino (2013).

Further, the actual relation between the mathematical notion of information and a more general intuitive or colloquial sense of information is unclear.

Shannon information is often described as a reduction of uncertainty (Halliday, 1983; Krebs and Davies, 1993; Seyfarth et al., 2010). This is also how the acquisition of knowledge by receivers is described (Quastler, 1956; Wiley, 1983; Seyfarth and Cheney, 2003; Bergstrom and Rosvall, 2011; Wheeler et al., 2011). Slightly different still is the quantitative formalisation of the change in the probability of a predicted event upon perceiving a signal—i.e., the conditional probabilities that play into statistical decision theory (McNamara and Dall, 2010; Skyrms, 2010a; Bradbury and Vehrencamp, 2011).

Shannon entropy is not equivalent to, or a measure of, information in the colloquial sense; e.g., the content of a signal or message (Shannon, 1948; Quastler, 1956; Marler, 1961; Smith, 1965; Markl, 1985; Stegmann, 2013). Since Shannon entropy (H) is an average, every message in a repertoire 'has' the same value of Shannon entropy. However, each of the messages in the repertoire may be about different things—i.e., they may have different meanings or contents. Thus, the Shannon entropy is the same, but the 'information', in the colloquial sense, is different. Therefore, these two concepts are not logically equivalent.

4.3.1 Shannon Entropy and Relative Entropy

According to Shannon (1948) and Shannon and Weaver (1949), a *communication system* is composed of five different parts. (1) The *source* provides a message (or series of messages) to be communicated to the receiving terminal; (2) The *transmitter* takes in the message from the source and produces a signal suitable for transmission over the channel; (3) The *channel* is the medium through which the signal is sent from the transmitter to the receiver; (4) The *receiver* takes in the signal and reconstructs the message from it; (5) The destination

is the intended recipient of the message. Additionally, noise may enter the channel, thus potentially obfuscating the message. See Figure 4.2.

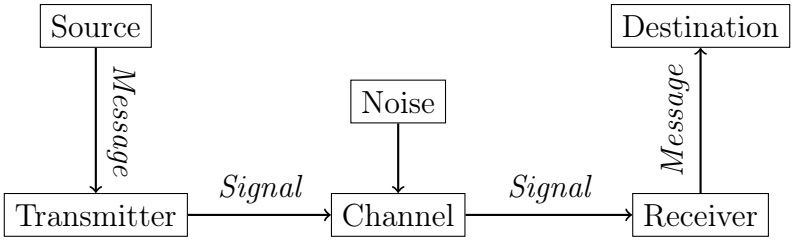


Figure 4.2: Schematic diagram of a general communication system from Shannon (1948)

Entropy, in the mathematical theory of information, is perhaps best understood as a measure of the degree of randomness in some data set. Understood in this way, it follows that more entropy means a higher degree of randomness, and less entropy means *higher predictability*. Suppose X is a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x) = p_X(x) = \Pr\{X = x\}, x \in \mathcal{X}$.[15] The definition for Shannon entropy is given in 4.5.

Definition 4.5. *Shannon Entropy*:

The (Shannon) *entropy* $H(X)$ of a discrete random variable X is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x). \tag{4.1}$$

The base of the logarithm, b, determines the unit of measure. For $b = 2, e, 10$, the unit of information is given by *Bit*, *Nat*, or *Hart*, respectively. As is a standard convention, we will assume that $0 \log 0 = 0$.[16] Note that the entropy of a discrete RV does not depend on the alphabet since it is a function of the *distribution* of X; therefore, it depends solely upon the probabilities underlying this distribution.

[15] In this case, $p(x)$ and $p(y)$ refer to two different random variables—indeed, two different probability mass functions, $p_X(x)$ and $p_Y(y)$. See discussion in Cover and Thomas (2006).

[16] This is often justified by the fact that $x \log x \to 0$ as $x \to 0$.

Definition 4.5 satisfies several intuitive properties—for example, entropy is non-negative, $H(X)$ is a continuous and concave function of X, entropy is additive, etc.

This can be extended to define the entropy of a pair of random variables, X and Y, as shown in Definition 4.6.

Definition 4.6. *Joint Entropy*:

The *joint* entropy, $H(X,Y)$ of two discrete random variables, $(X,\ Y)$ with a joint probability distribution $p(x,y)$ is defined by

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_b p(x,y), \tag{4.2}$$

and their conditional entropy is defined as in 4.7.

Definition 4.7. *Conditional Entropy*:

The *conditional entropy*, $H(Y|X)$ is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log_b p(y|x) \\ &= H(X,Y) - H(X). \end{aligned} \tag{4.3}$$

The entropy of a random variable, in general, is best described as a measure of how much information is required, on average, to describe the random variable fully. For example, if we consider the set of states in the 2×2 signalling game with unbiased nature as a discrete random variable $S = \{s_0, s_1\}$, such that $p(s_0) = p(s_1) = 1/2$, $H(S)$ tells us that we need, on average, 1 bit of information to describe S.

Relative entropy—also known as Kullback-Leibler (KL) Divergence—is understood as a measure of the similarity of two probability distributions, p and q.[17] Put another way, relative entropy is a measure of how inefficient it is to assume that the distribution is given by q when it is in fact given by p (Cover and Thomas, 2006). The relative entropy of two distributions is given in Definition 4.8:

Definition 4.8. *Relative Entropy (Kullback-Leibler Divergence)*:
The *relative entropy*, or the *Kullback-Leibler distance*, between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$\begin{aligned} D(p \parallel q) &= \sum_{x \in \mathcal{X}} p(x) \cdot (\log_b p(x) - \log_b q(x)) \\ &= \sum_{x \in \mathcal{X}} p(x) \log_b \frac{p(x)}{q(x)} \end{aligned} \tag{4.4}$$

This quantity is always non-negative and is equal to zero just in case $p = q$. Further, D_{KL} is a convex function of P. Note that this is not technically a metric, since it is not *symmetric*, nor does it satisfy the *triangle inequality*.[18]

We can further define the *mutual information* between two discrete random variables, X and Y, as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_b \frac{p(x,y)}{p(x)p(y)} = H(X) - H(X|Y). \tag{4.5}$$

This is, intuitively, a measure of the amount of information that X any Y share—i.e., it is a measure of the amount of information that Y affords over and above X. Once X is known, the conditional entropy gives a measure of the remaining uncertainty of Y.

[17] It is technically not a measure of the *distance* between two distributions, since KL divergence lacks symmetry, and so is not properly a metric. There are ways of remedying this, but it is unimportant for our purposes here.

[18] The fact that KL-Divergence is non-negative is shown in a theorem by J. Willard Gibbs; further, it is zero just in case the two probability distributions are equivalent (almost) everywhere.

There is a close relationship between mutual information and Kullback-Leibler divergence. Namely, mutual information can be expressed as a KL-divergence of the product of marginal distributions, $p(x) \cdot p(y)$, of the random variables X and Y, from their joint distribution, $p(x, y)$. Formally,

$$I(X;Y) = D_{KL}(p(x,y) \parallel p(x) \cdot p(y)) \tag{4.6}$$

4.3.2 Semantic Information and Signalling

Shannon information depends upon discrete random variables. However, we note that the elements of the signalling game can be understood as a set of discrete random variables, $\{S, M, A\}$. S is a static random variable with some associated probability distribution—uniform, in the simplest case.

Skyrms (2010b) points out that the information that a signal carries is information *about* what state obtains. When signals are entirely informative, the receiver has complete information about the state of the world, and so can act as though she had observed the state directly.

However, this depends upon a comparison between the (conditional) probability that we are in a particular state *given that* a signal was sent, and the likelihood that we are in that state simpliciter. This key quantity is

$$\frac{p(s_i|m_j)}{p(s_i)}$$

As an example, consider the probability distributions for an atomic 2-game at the outset of game-play. Then $p(s_1) = p(s_2) = p(m_1) = p(m_2) = p(a_1) = p(a_2) = 1/2$. Our key quantity for s_1 and m_1 is

$$\frac{p(s_i|m_j)}{p(s_i)} = \frac{1/2}{1/2} = 1$$

Similarly, suppose our initially random system has evolved into the signalling system shown in Figure 1.1, in Chapter 1. Then the key quantity gives

$$\frac{p(s_i|m_j)}{p(s_i)} = \frac{1}{1/2} = 2$$

Skyrms points out that the first situation should output 0 since there is intuitively no information carried by the signal about the state. This is achieved by taking the logarithm of this key quantity.

Thus, we can define the *quantity* of information carried by a signal, m_j (i.e., about a particular state, s_i) as

$$H(m_j) = \log_2 \frac{p(s_i|m_j)}{p(s_i)}.$$

Now, our first key quantity results in 0 information. In contrast, the quantity at the signalling system results in m_1 carrying 1 bit of information—this intuitively makes sense, because this corresponds to a reduction of uncertainty from two possibilities to one.

Skyrms point out that signals may carry some information about different states. Thus, to get a real sense of the amount of information that a signal carries, we can take a weighted sum of the probabilities of being in any of the particular states conditional upon the specific signal. Thus, we obtain the following measure of the quantity of information that is carried by a particular signal, m_j, about the states:

$$\underset{states}{I(m_j)} = \sum_{i=1}^{|S|} p(s_i|m_j) \cdot \log_2 \left(\frac{p(s_i|m_j)}{p(s_i)} \right)$$

Note that this is just the KL-Divergence of the two probability distributions $P = p(s|m), Q = p(s)$. Signals can also carry information about the acts:

$$\underset{acts}{I(m_j)} = \sum_{i=1}^{|A|} p(a_i|m_j) \cdot \log_2 \left(\frac{p(a_i|m_j)}{p(a_i)} \right)$$

In this context, the relative entropy of a particular signal can be understood as a measure of *additional bits gained* by moving from a prior to a posterior distribution, in a Bayesian sense.

$$I(M) = \sum_{i,j} p(s_i|m_j)H(m_i)$$

This gives us a notion of the *quantity* of information in a signal, but Skyrms (2010a) additionally uses this notion to define the informational *content* of a signal. The informational content of a signal on this account is just a vector which specifies the information that the signal gives about each state. This vector is given by

$$\left\langle \log_2 \left(\frac{p(s_1|m_j)}{p(s_1)} \right), \log_2 \left(\frac{p(s_2|m_j)}{p(s_2)} \right), \cdots, \log_2 \left(\frac{p(s_n|m_j)}{p(s_n)} \right) \right\rangle \tag{4.7}$$

for the content about the states of a particular signal, m_j.

Thus, if we suppose that there are four states, which are initially equiprobable, then the informational content about the states of each signal at the outset is given by the following vectors.

$$\begin{aligned} I(m_1) &= \langle 0, 0, 0, 0 \rangle \\ I(m_2) &= \langle 0, 0, 0, 0 \rangle \\ I(m_3) &= \langle 0, 0, 0, 0 \rangle \\ I(m_4) &= \langle 0, 0, 0, 0 \rangle \end{aligned} \tag{4.8}$$

That is, none of the signals carries any information about the states, and so their content is empty everywhere. If we further suppose that the sender and receiver evolve to a signalling system where signal i is sent only in state i, then the informational content of each signal at

that signalling system is given by the following vectors.

$$\begin{aligned} I(m_1) &= \langle 2, -\infty, -\infty, -\infty \rangle \\ I(m_2) &= \langle -\infty, 2, -\infty, -\infty \rangle \\ I(m_3) &= \langle -\infty, -\infty, 2, -\infty \rangle \\ I(m_4) &= \langle -\infty, -\infty, -\infty, 2 \rangle \end{aligned} \tag{4.9}$$

That is, each signal carries precisely 2 bits of information about the state of nature. The $-\infty$ components tell us which signals end up with probability zero conditional on which states.

This account, Skyrms (2010a) notes, is more general than the traditional account in the philosophy of language—where the (at least declarative) content of a signal is a proposition, and a proposition is a set of possible worlds. He highlights that a proposition can just as well be specified by giving the set of states that the true state is not in, and this is precisely what the $-\infty$ component of the vector does. Therefore, the notion of propositional content as a set of possible worlds is *contained* in this richer information-theoretic account of the content of a signal.

Furthermore, the *quantity* of information in a signal can be obtained by averaging over the components of the informational *content* vector. Thus, the quantity of information in a signal is a summary of the informational content of that signal. If this is averaged again, then we obtain the *mutual* information in the signals, given the relationship previously discussed between mutual information and KL-divergence. The maximum of this over signalling systems gives us the information-transfer capacity of a particular signalling game. Thus, Skyrms highlights 'There is a seamless integration of this conception of content with classical information theory' (42).

Nonetheless, this notion of content depends upon how probabilities are *moved* (Skyrms, 2010a). Godfrey-Smith (2011) suggests that the content of the signal should say something about *the world* rather than how much the probability of a particular state was moved by the signal's being sent. The informational content of a signal is going to be given as in 4.7 above—each vector entry is the quantity of information about a particular state, provided by the signal. However, the actual state of the world can be given a similar distribution. Suppose the content of a specific signal in a 3×3 signalling game, out of equilibrium, is given by $\langle 0.2, 0.5, 0.3 \rangle$. The actual state of the world—e.g., state 2—can be given by the distribution $\langle 0, 1, 0 \rangle$ for states 1, 2 and 3. Godfrey-Smith (2011) suggests that the *distance* between these two distributions might provide us with a measure of *how close* the content of the signal is to the truth. Several measures could be used, including Kullback-Leibler divergence. In this case, supposing P is the probability distribution of the states of the world, and Q is the probability distribution of the states conditional on the signal, we have $-\log_2 P(s_i|m_j)$, where s_i is the *actual* state of the world. Thus, the message with content $\langle 0.2, 0.5, 0.3 \rangle$ is precisely 1 bit of information away from the truth—Namely, s_2.

A nice property of this measure is that it has a minimum value of 0, when $Pr(S_i|m_j) = 1$, and no upper bound. Thus, if we are at a signalling system, the vector of probabilities for the signal is going to be equivalent to the vector for the states: 1 for the actual state, and 0 everywhere else.

4.3.3 Note on the Problem of Error

Birch (2014) highlights the fact that Skyrms' account of informational content falls prey to the *problem of error* (in the same way as the information-theoretic approach to content in Dretske (1981)).[19] If we consider what it would take for a signal, say m_j, to have *false* propositional content, two conditions need to be satisfied. On the one hand, there needs to

[19]See Fodor (1984b); Godfrey-Smith (1989); Crane (2003).

be a state, s_i, that the signal rules out. Thus, $p(s_i|m_j) = 0$. This implies that m_j is never sent when s_i obtains. Further, it has to be the case that on at least one occasion, m_j is sent when s_i does obtain so that on such an occasion, the propositional content of m_j can be said to be false. However, these two conditions cannot hold simultaneously, since the first requires that $p(s_i|m_j) = 0$ and the second requires that $p(s_i|m_j) \neq 0$.

Birch (2014) proposes a solution to this problem based on a notion of *fidelity conditions* (Stampe, 1977) for a signal such that we can allow one to say when a signal is being used misleadingly—what is necessary is to specify such fidelity conditions in a non-arbitrary way.

Skyrms and Barrett (2018) suggest that these fidelity criteria might instead be defined by a fully common-interest interaction (sub-game) that stabilises signalling. They suggest that the *content* of 'wolf' does not derive from the boy who cries wolf, even if this false signal becomes more prevalent. Thus, they separate signalling *contexts*—the content of the signal is determined by the context of common interest signalling, and usage in these contexts crystallises the meaning of the signal. Once this convention is in place, they posit that the signal may be used as a lie in a separate context.

For example, *Photinus* is a genus of firefly that flash as a mating signal. It is in this context that the signal comes to have meaning (in the sense of informational content). However, female *Photuris* fireflies engage in aggressive mimicry—they imitate the *Photinus* mating signal to lure male *Photinus* prey for consumption.[20]

Alternatively, Shea et al. (2018) suggest separating the informational content of a signal from its functional content. However, while these insights and suggestions are theoretically valuable, we will ignore them for now—better to focus on getting the simple cases right before moving on to the unsolvable cases.

[20]Stanger-Hall et al. (2007) suggest that the insincere flashing bioluminescent signals of *Photuris* seem to have evolved independently of the *Photinus* genus and was eventually adapted to those of *Photinus* (or *Pyractomena*).

4.4 Measuring Compositionality

Given the semantic notion of information discussed in Section 4.3.2, we can make exact the argument that syntax alone does not give rise to compositionality. This captures the complaints of Franke (2016); Steinert-Threlkeld (2016), that composite signals are interpreted atomically, and so cannot be compositional in the sense that they do not capture intuitions about generalisability conditions for compositional signalling.

Suppose we have a 4×4 syntactic signalling game, with two senders and one receiver. Each of the senders can send one of two messages, and the receiver is sensitive to which sender sent which message. Suppose further that the senders and receiver have evolved a signalling system, as shown in Figure 4.3. This is a signalling system, though it perhaps does not

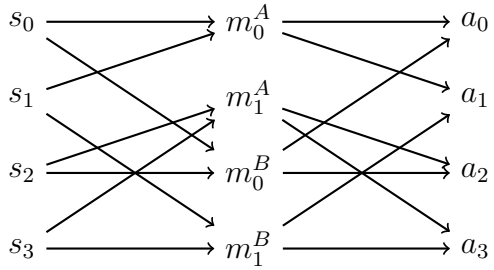


Figure 4.3: Signalling system for a syntactic signalling game

look like one at first. What happens is that each sender's signal partitions nature into two sets—$\{s_0, s_1\}$ and $\{s_2, s_3\}$ for sender A's signals and $\{s_0, s_2\}$ and $\{s_1, s_3\}$ for sender B's signals—and the combination of these signals determines the state via the intersection of these two sets. See Figure 4.4.

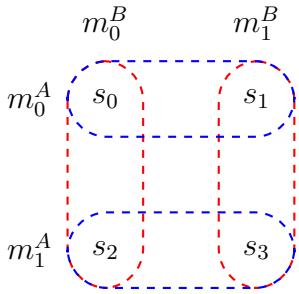


Figure 4.4: Fully partitioning states via set intersection

Note that from a theoretical point of view, we know the maximal entropy of the system, from Definition 4.5. This is given by

$$\begin{aligned} H(S) &= -\sum_{s \in \mathcal{S}} p(s) \log_2 p(s) \\ &= -\log_2 \left(\frac{1}{4}\right) \\ &= 2 \text{ bits.} \end{aligned}$$

Thus, an entirely informative length-two signal carries 2 bits of information because it reduces the possible states from 4 to 1. We defined the *informational content* of a particular signal with respect to the states as a vector. Therefore, we can give the entire informational content of all of the signals explicitly as a matrix. Each row is the informational *content*, as defined in Skyrms (2010a), of a particular message; see Table 4.1.

		States			
		s_0	s_1	s_2	s_3
Informational Content	m_0^A	1	1	$-\infty$	$-\infty$
	m_1^A	$-\infty$	$-\infty$	1	1
	m_0^B	1	$-\infty$	1	$-\infty$
	m_1^B	$-\infty$	1	$-\infty$	1

Table 4.1: Complete informational content about the states at a signalling system in a 4×4 syntactic signalling game.

Further, we can see that a particular state is *wholly determined* by all and only the messages that carry information about that state. Therefore, s_0 is entirely determined by the

combination of m_0^A and m_0^B, rather than, e.g., the combination of m_0^A and m_1^B, because the latter carries no information about state 0 when in combination with m_0^A. The syntactic combination of a complex length-two signal, as was suggested above, fully partitions nature, and so carries complete information about a particular state; see Table 4.2.

		States			
		s_0	s_1	s_2	s_3
Informational Content	$m_0^A{}^\frown m_0^B$	2	$-\infty$	$-\infty$	$-\infty$
	$m_0^A{}^\frown m_1^B$	$-\infty$	2	$-\infty$	$-\infty$
	$m_1^A{}^\frown m_0^B$	$-\infty$	$-\infty$	2	$-\infty$
	$m_1^A{}^\frown m_1^B$	$-\infty$	$-\infty$	$-\infty$	2

Table 4.2: Complete informational content in simple signals about the acts at a signalling system in a 4×4 syntactic signalling game.

Now, suppose that sender B spontaneously changes her signal m_0^B to a new signal, $m_?^B$. We can account for two possible situations. Either, sender B simply uses a novel signal in lieu of m_0^B, in which case the meaning of these two distinct signals is equivalent—this is similar to cue reading, in the sense that sender B's new signal has a fixed meaning, which the receiver needs to learn. Or, we might imagine that sender B forgets the meaning of signal m_0^B, in which case she needs to re-coordinate so that the new signal successfully partitions nature when combined with sender A's signal—this is akin to the normal signalling context since the second sender must re-learn when to send this novel signal (given the meanings of all the other signals are fixed, the correct strategy is to send the new signal in the same context as that in which the prior signal was used), and the receiver must additionally learn the meaning of the novel signal.

In the urn-learning metaphor, these two situations might be modelled in a variety of ways; however, each of these is functionally equivalent under the assumption that the meaning of the *other* signal does not change, as we shall see. If we suppose that the novel signal just means the same thing as the old signal, this corresponds to taking every ball labelled m_0^B in each of the state urns for sender B and re-labelling them $m_?^B$. Since the senders

already convened upon a signalling system that perfectly partitions the states of nature, this re-labelling does not change that. On the other hand, if we assume that the meaning of m_1^B remains the same, and only m_0^B changes, then sender B's urns for states s_1 and s_3 should, hypothetically, remain unchanged. However, this means that 'forgetting' the meaning of signal m_0^B consists in 'emptying' all of the balls from the s_0 and s_2 urns and adding a novel ball labelled $m_?^B$ to those urns. However, since the meaning of m_1^B is fixed, it follows that even if we 'reset' the urns for s_0 and s_2 with one each of $m_?^B$ and m_1^B, the conditional probability that s_0 obtains given that m_1^B is sent is effectively 0. Therefore, under either interpretation, the informational content vectors about the states remain unchanged. This is shown in Table 4.3.

		States				
		s_0	s_1	s_2	s_3	
Informational Content	m_0^A	1	1	$-\infty$	$-\infty$	
	m_1^A	$-\infty$	$-\infty$	1	1	
	$m_?^B$	1	$-\infty$	1	$-\infty$	$\leftarrow$ Novel Signal
	m_1^B	$-\infty$	1	$-\infty$	1	

Table 4.3: Complete informational content about the states at a signalling system in a 4×4 syntactic signalling game with a novel signal identical to the old signal.

However, the signals also carry information about the acts. The role of the receiver is asymmetric for the following reason. If we assume that message m_0^B is replaced with message $m_?^B$, this can only be modelled by effectively throwing out the receiver urns that have a token of m_0^B and *creating* new urns that are labelled identically to the old urns, except with each token of m_0^B replaced with $m_?^B$—the new signal from sender B. Therefore, we can calculate the information that each of the concatenated signals contains about the acts, as before. We obtain the complete information matrix shown in figure 4.4. That is to say, any composite signal containing a token of the novel signal *carries no information.*

However, if the concatenated signals we compositional, this should not happen. Consider that, regardless of what the new signal means, m_0^A is only sent for a_0 or a_1. Therefore, the

		Acts			
		a_0	a_1	a_2	a_3
	$m_0^A{}^\frown m_?^B$	0	0	0	0
Informational	$m_0^A{}^\frown m_1^B$	$-\infty$	2	$-\infty$	$-\infty$
Content	$m_1^A{}^\frown m_?^B$	0	0	0	0
	$m_1^A{}^\frown m_1^B$	$-\infty$	$-\infty$	$-\infty$	2

Table 4.4: Complete informational content in simple signals about the acts at a signalling system in a 4×4 syntactic signalling game.

conditional probability that a_2 or a_3 should obtain, given that the receiver has received a length-two string starting with m_0^A, is 0. Similarly, for a_3. The probability of a particular act being appropriate simpliciter is still the chance probability, 0.25. What does this mean for the informational content of the concatenated signal? It is given by

$$\left\langle \log_2 \left(\frac{p(a_0 | m_0^A{}^\frown m_?^B)}{p(a_i)} \right) \right\rangle, \qquad i \in \{1, 2, 3, 4\}.$$

Substituting the values for the conditional and unconditional probabilities, we have

$$\left\langle \log_2 \left(\frac{1/2}{1/4} \right), \log_2 \left(\frac{1/2}{1/4} \right), \log_2 \left(\frac{0}{1/4} \right), \log_2 \left(\frac{0}{1/4} \right) \right\rangle,$$

which resolves to the informational content vector

$$\langle 1, 1, -\infty, -\infty \rangle .$$

However, this makes no sense: m_0^A alone gives us 1 bit of disjunctive information—namely, about $a_0 \vee a_1$. If the receiver were interpreting the concatenation of m_0^A and $m_?^B$ compositionally, indeed, the second part of the length-two signal would not give her any *new* information regarding the disjunction $a_0 \vee a_1$—namely, unlike before, where the novel signal provides disjunctive information so that the union of the two signals uniquely determines a single state. There is no reason why changing the second signal should take information away from

the entire composite signal. But, as we have seen, the receiver interprets the signal as an atomic whole, which provides no information about the act. See row 1 of Table 4.4.

This shows that the signals are not *interpreted* compositionally. However, it also highlights that they are compositional for the senders (or, for the states, if you prefer). This is because there is a notion of *independence*—concerning the information that the signal carries about the states—that does not hold for the information that the signal carries about the acts.

We assumed that the states were fixed in the previous example, and only one of the signals changed its meaning. We saw that this has no effect on the informational content of the signal concerning the states, but the receiver counter-intuitively loses the information that should have been contained in the unchanged signal. What if we suppose that we *extend* the lexicon, rather than merely altering it? The same argument holds, however, if we assume that a novel state, a novel signal to represent that state, and a novel action to perform in that state are introduced into the signalling game.

To tell an intuitive story, we might suppose that sender A sends a verb, and sender B sends a *noun*. Suppose there are two distinct action-contexts and two distinct object contexts. Thus, we have the 4×4 syntactic signalling game, as before. Suppose now that a *novel object* context is added to the game. The noun-sender accommodates this by adding a new signal to her lexicon and sending that in the novel context.

The receiver, again, must learn what is appropriate given this new signal; however, given that the verb context has not changed, she should gain *some* information. This argument captures precisely the intuition noted above regarding what systematicity is supposed to achieve: if the receiver knows the meaning of 'pick up x' and the meaning of 'the book', but not the meaning of 'put down x', then she might understand the command 'pick up the book', though she does not understand the meaning of 'put down the book'. Even so, she may still understand that the latter expression has *something* to do with the book.

Though this argument specifically concerned the syntactic signalling model given in Barrett (2006, 2007, 2009), the same considerations apply to the model for combinatorial systems of communication proposed by Scott-Phillips and Blythe (2013). Since they explicitly focus on composition *qua* syntactic composition, this system cannot give rise to a genuine notion of compositional signalling. The same is true for spill-over reinforcement (Franke, 2014, 2016). If we add a novel signal to a pre-established signalling system that has evolved via spill-over RL, the receiver loses information in any string containing the novel signal; therefore, Brochhagen (2015) is correct in pointing out that the agents are not sensitive to our generalisation condition for compositionality—namely, the relations between constituent parts are not generalisable.

The model for functional negation proposed by Steinert-Threlkeld (2016) does not fall prey to this argument, however. This is because he presupposes the functional capacity of a special (function) signal, $\boxminus(\cdot)$. That is, the receiver interprets $\boxminus m_i$ as a 'minimal negation' in the sense that, when it receives a composite signal $\boxminus m_i$, she looks at what act she *would* have performed had she received the atomic message m_i—e.g., a_j—and then performs act $f(a_j)$ in response. Obviously, if a new signal, $m_?$, is introduced into the pre-established signalling system, the receiver need only learn the meaning of this new signal; then she is immediately able to interpret the signal $\boxminus m_?$—the information of the pre-understood functional component of the compositional signal does not change even when we add novel signals to the lexicon. i.e., once the receiver learns the meaning of $m_?$, she immediately understands the meaning of $\boxminus m_?$, so this model is sensitive to generalisation. However, recall that Steinert-Threlkeld (2016) does not explain, indeed is not concerned with, *how* such function words might evolve in the first place.[21]

Brochhagen (2015) highlights that, for a complex signal to be compositional, there needs to be a systematic association between simplex elements and the complex elements of which

[21] Though see Steinert-Threlkeld (2020).

they are constituents. To account for productivity, structural properties that are common between components of complex signals must be recognisable (and indeed recognised) for it to be possible to learn how to (de)compose two such elements in such a way that this can be generalised over their classes. In particular, as we have seen, if each combination of parts needs to be learned case by case and mentally stored in a lexicon for interpretation, then this will not provide any advantage to the receiver.

This is enough for *syntactic* compositionality—composition on the part of the sender. However, this leads to, what Steinert-Threlkeld (2018, 2020) calls, *trivial compositionality*. He forwards the following definition:

> *Trivial Compositionality*
> A communication system is trivially compositional just in case complex expressions are always interpreted by the intersection (generalised conjunction) of the meanings of the parts of the expression. (Steinert-Threlkeld, 2020, 3-4)

He highlights that the models that we discussed in Chapter 2 all share the following underlying assumptions:

(A1) Agents communicate about a fixed set of states.

(A2) Optimal communication consists in correctly identifying the true member of the state space.

(A3) Messages are fixed-length sequences of signals from fixed sets.

He then *proves* that if one's model carries all three of these assumptions, the composition that emerges from the model will necessarily be trivial.[22] However, non-trivial composition

[22]Note that Steinert-Threlkeld (2014, 2016) and Barrett et al. (2018) drop assumption (A3); Steinert-Threlkeld (2020) drops (A1).

is a necessary precondition for the emergence of function words. This is captured in the following:

> *Generalisation Condition* (Brochhagen, 2015):
> The relation between simplex elements and the complex elements of which they are constituents must be generalisable.

Barrett et al. (2018) do show how such non-trivial compositionality might evolve. Let us consider their model briefly. (See also the discussion in Chapter 6.) In the simplest case (the special composition game), there are two basic senders, one executive sender, one basic receiver, and one executive receiver. The executive sender and receiver—called *hierarchical agents*—can learn to influence the behaviour of the basic senders and receiver—called *basic agents*.

The state of nature consists in two *properties*—COLOUR $= \{black, white\}$ and ANIMAL $= \{dog, cat\}$—and a *context* $(= \{colour, animal, both\})$. That state of nature is the combination properties—black dog, white dog, black cat, and white cat—and the context indicates to which of these the receiver needs to attend to perform the correct action.

In the simplest case, they assume that each basic sender is assigned a particular property and only has access to that aspect of the state of nature—in this sense, this is similar to the 4×4 syntactic signalling game. One basic sender sees the colour, and the other sees the animal. The executive sender sees the context, and determines whether the colour sender, the animal sender, or both will send their signals to the receiver—note that the executive sender must learn *which* type of signal a particular context (determined by nature) demands; this is not presupposed.

The basic receiver sees the signal(s) sent by the basic senders and is sensitive to which sender sent which signal. The executive receiver determines whether the basic receiver

should interpret the signal as colour, animal, or both. The receiver performs an action, which can be one of black, white, cat, dog, black dog, black cat, white dog, white cat.

A play is counted as a success *just in case* (1) the receiver performs the correct action given the context and (2) the senders only sent the signals required for success, given the context. Therefore, as Barrett et al. (2018) point out, success requires that the receiver's action matches the state of nature *and* that the senders are as efficient as possible. Agents learn by simple reinforcement learning, as usual. On each play, nature chooses a value for each of the two properties and the context randomly and with uniform probabilities. Thus, each colour and animal have probability 0.5 of being chosen—each combination of colour and animal has probability 0.25—and each context has probability $\frac{1}{3}$.

The executive sender has an urn for each of the three contexts, each containing a ball for each type of sender—colour sender, animal-sender, and both. The executive sender sees the context and chooses a ball at random from the appropriate urn. This determines which sender will send a message. The colour sender is equipped with an urn for each possible colour property, and each urn contains a ball labelled 0 and 1. The animal sender is equipped with an urn for each possible animal property, and each urn likewise contains a ball labelled 0 and 1.

The basic receiver has four urns—again, as was the case in the 4×4 syntactic signalling game. Each urn is labelled for every ordered pair of signals that she might receive—$00, 01, 10, 11$. Each urn contains a ball for each of the colour-animal pairs. If the receiver receives a length-two message, she selects a ball at random from the appropriate urn. However, if she receives a length-one message (i.e., if only one sender sends a signal), then the receiver chooses randomly (with unbiased probabilities) from one of the two urns that correspond to the sender's signal and draws a ball at random from that urn.

The executive receiver is equipped with urns labelled colour-sender, animal-sender, and both. Each of these initially contains a ball labelled 'colour', 'animal', and 'both'. The ball chosen by the executive receiver determines *how* the base receiver will interpret the type of signal received. This interpretation, in conjunction with the ball drawn, determines how the receiver will act. If the executive receiver draws the 'both' ball, then the basic receiver performs the action corresponding to the ball that she drew; if the executive receiver draws the 'animal' or 'colour' ball, then the basic receiver performs the action corresponding to the appropriate property from the ball she drew. Therefore, the ball drawn by the executive receiver and the base receiver jointly determine the action—black, white, cat, dog, black dog, white dog, black cat, white cat. Reinforcement, in this case, works as follows: If a play of the game is successful, then each agent who was involved in that particular play returns the ball she drew to the urn from which she drew it and adds another a ball of the same type to that urn.

Barrett et al. (2018) report that on simulation with this simple reinforcement learning set up, in almost every case (0.97 of runs) the agents collectively evolve a maximally efficient and successful compositional language. Furthermore, their model of hierarchical compositionality *also* does not fall prey to the prior argument: if we add a new *animal* state, for example, then the receiver loses no information when she is sent a combinatorial signal of colour and animal, and the context is, e.g., colour.

However, it is not the compositionality of the signals *itself* that drives compositionality in this signalling system—instead, it is the *reflexivity* and *modularity* of the executive sender and receiver that drives compositionality in this context—the ball that the executive sender chooses *refers* to a component of the base-game. This can be seen by the fact that the base-game (constituted by the base-senders and base-receiver) is functionally equivalent to the 4×4 syntactic signalling game, which does *not* evolve compositional signalling, as we have seen.

4.5 Discussion

Why should information be the essential criterion for understanding compositional signals? Information transfer is an excellent measure for meaning in the sense that it can also be understood in a theory-neutral way. Skyrms (2010a,b) takes a stronger position than this, and suggests that meaning just *is* (semantic) information—in a duplex sense of both quantity and content; see also Dretske (1981). However, we can remain neutral about meanings in the sense that, no matter what we take meaning to *be*, in meaningful exchanges, information is transferred. Further, many researchers already assume this notion, and by defining compositional signals in terms of information, these assumptions are made explicit. For example, the following selections appeal to a notion of information without actually defining what information is.

> '[signals are] behavioural, physiological, or morphological characteristics fashioned or maintained by natural selection because they convey information to other organisms' (Otte, 1974, 385);
>
> '[communication] consists of the transmission of information from one animal to another' (Green and Marler, 1979, 73);
>
> '[communication is] any sharing of information between entities' (Smith, 1997, 11);
>
> 'Signals carry certain kinds of informational content, which can be manipulated by the sender and differentially acted on by the perceiver' (Hauser, 1996, 6);
>
> 'the function of most signals is to provide information .... If this provision of information benefits both sender and receiver, mutations in either party that refine and improve the process will be favored over evolutionary time' (Bradbury and Vehrencamp, 2011, 4);

> 'We define a "signal" as any act or structure that alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved. ... the signal must carry information,—about the state or future actions of the signaler, or about the external world—that is of interest to the receiver' (Maynard Smith and Harper, 2003, 3);
>
> 'Honest signals are those which accurately (but not necessarily perfectly) convey information about some relevant quality of the signaler (e.g. its species, sex, size, condition, etc.) or environment' (Fitch, 2008, 385).

In this chapter, I suggested that, due to the inherent ambiguities and complexities of natural languages, the question of whether or not languages are compositional is grossly underspecified. As such, an alternative approach to discussing the compositionality of language from an evolutionary standpoint is to discuss simple communication systems to determine the conditions under which they would be taken to be compositional. In particular, if compositionality is a necessary condition for the generative nature of languages, and if languages evolved from simpler communication systems, then compositionality itself evolved. Thus, to clarify how this sort of compositionality might have evolved, it is necessary to determine what counts as a compositional *signal*.

I suggested that, in light of evolutionary considerations, we should not appropriate a notion of compositionality from natural languages, but rather analyse complex signals in a simpler communication context. Thus, we built a simple notion of compositionality from the ground up, as it were. This helps to avoid many of the conceptual difficulties arising from the discussion of compositionality in natural languages, in the same way that simple models of the world avoid the complexities of the actual world for conceptual clarity and tractability.

Moreover, Armstrong (2018) highlights that we must distinguish between compositional communication and social coordination with compositionally determined meanings—compositional

systems of communication form a mere subset of systems of social interaction that are mediated by compositionally structured internal representations. He highlights that compositionality plays a more significant role in social and cognitive phenomena over and above the power that compositionality might bestow upon communication. For example, baboons are capable of (*i*) generating discrete representations of individuals in their troops, (*ii*) merging those representations of individuals to form complex representations of families, (*iii*) embedding representations of both individuals and families under a hierarchical relation (e.g., *dominant with respect to*), and (*iv*) deploying those representations in flexible and socially situated ways (Armstrong, 2018).[23] He also points out that though baboons have compositionally rich cognitive and social structures, their communication system lacks compositionality—thus, these other forms of compositionality may provide a necessary, though certainly not a sufficient condition, for compositional communication.

Armstrong (2018) proposes that the human language faculty evolved as the product of complex feedback mechanisms that gradually diversified and changed humans (perhaps hominins) into different kinds of animals from other living primates. This feedback loop, on his account, involves interrelations between social organisation, complex cognition, and environmental modification.

Relevant semantic enrichments involve extended capacities for tracking mental states or theory of mind; cognitive capabilities supporting extractive foraging, including the use of tools; cognitive capacities supporting more extensive and more variable habitat ranges; cognitive capacities supporting more coalition partners in larger groups; and cognitive capacities enabling the formation of planned actions with variable components.[24] Additional phonological enrichment might include an extended capacity for signal learning in the case of both sounds and gestures; extended volitional control for both vocal and motor production; and motivations to use existing abilities for expression in new ways (Armstrong, 2018).

[23]See also, Cheney and Seyfarth (2007); Seyfarth and Cheney (2018).
[24]See also, Steedman (2009).

Chapter 5

The Correction Game

> *[I]t seems to us as though in this case the instructor imparted the meaning to the pupil—without telling him it directly; but in the end the pupil is brought to the point of giving himself the correct ostensive definition. And this is where our illusion is.*
>
> — Wittgenstein, *Philosophical Investigations*

In this chapter, I present a model of learning that varies the reward for coordination in the signalling game as a function of the agents' actions. The model takes advantage of the type of communicative bootstrapping processes that were suggested in Part I—namely, how previously evolved capacities might help to more efficiently evolve new capacities, via reflexivity.

Recall the simple reinforcement learning dynamic that was presented in Chapter 1. Propensities for a particular action under this dynamic are proportional to the accumulated rewards for those actions. Thus, previous successes make it more likely that a specific action will be chosen in the future. An urn-learning process illustrated this: when agents are successful in coordinating signals to state-act pairs, they reinforce their behaviour, thus shifting the

probabilities that the same action will be chosen in the same context in future plays. When the sender and receiver miscoordinate, they do not reinforce their behaviour. Indeed, they may be punished for miscoordination, resulting in the *reduction* of probabilistically choosing an action that previously failed to achieve coordination on a future round.

We might note that coordination for communication is generally *goal-directed*. This need not be understood in terms of something as high-level as, e.g., *intentions*; rather, this may be as simple as understanding cue reading or sensory manipulation (in the sense discussed in Chapter 3) as 'goal-directed' behaviour. For example, in the cue-reading game, the sender has a fixed set of dispositions which the receiver must learn to interpret (in the sense of 'react to') in the appropriate way. Thus, the receiver's *goal* is to understand the sender's signals; the sender, on the other hand, is static—she does not have a goal in the way the receiver does, but reacts fixedly to the states of nature, regardless of whether the receiver interprets her actions correctly.

Might the sender in the cue-reading game not also have a goal—namely, for the receiver to *understand* her fixed signalling disposition? How might she achieve this goal? If she can communicate *that* her signal *means* such-and-such, then this would help the receiver toward the goal of interpreting the signal appropriately. However, this would be putting the cart before the horse, so to speak: the entire premise of the signalling [cue-reading] game is that we do not presuppose the sender can communicate the meaning of her signal [cue]; instead, it is precisely the 'intended' meaning, based on the sender's signalling disposition, that the receiver must learn. In the signalling game, the meaning of the signal co-evolves as a function of *both* the sender's and receiver's respective dispositions. If the sender could communicate her disposition, then the sender and receiver would have already arrived at a signalling convention.

Suppose that the sender and receiver have already evolved a signalling system in some *other* context. Might the sender not then use *those* communicative capacities to try to express

to the receiver what her meaning is in the new context? This does not presuppose that the sender and receiver have already solidified the meanings of the signals in the *main* context; instead, when the receiver fails to perform the action that the sender *wants* her to perform (i.e., the one that is appropriate for the particular state), the sender may be able to communicate *that* the receiver did something wrong.[1]

The following story makes more explicit the sort of phenomena that I have in mind here. Consider two actors in a signalling context. Suppose they have already evolved up some rudimentary communicative capacity. For example, they may have learned how to communicate some simple command for an action. This might be interpreted, at least for this story, as a command that represents some holophrastic binary distinction—e.g., 'stop/go', 'yes/no', 'correct/incorrect', 'true/false', 'right/wrong', etc.[2]

In such a context, there are two relevant states of the world, with corresponding appropriate actions, and there are two possible signals to represent these state-act pairs. One or the other signalling system will evolve with certainty, given that this is a 2×2 signalling game. Now, if we imagine this sort of communicative context has already evolved, it stands to reason that individuals in a new signalling context (where no dispositions have yet evolved) might learn to take advantage of their previously evolved communication convention in the following sense. Suppose the sender and receiver are in a novel signalling context, where they must evolve dispositions from scratch. In the standard signalling game model, they may learn to coordinate upon a signalling convention by merely trying things and reinforcing those actions that led to success.

[1]There is a lot of *intentional* talk in this paragraph; even so, I take this to be harmless for the reasons given by Dennett (1971, 1987); according to Dennett (1971), presupposing beliefs and desires on the part of such an agent—one who is not rational, *per se*—is a form of 'conceptually innocent anthropomorphizing' (93). It should be fairly clear that I am not presupposing, as a matter of fact, that a sender or receiver in the signalling game have any human-level cognitive capacities.

[2]Note that equally a command, or imperative, in a signalling system can be interpreted as an indicative statement. We will not worry about this distinction too much here, but see the discussion in Harms (2004a,b); Millikan (2005); Huttegger (2007b); Zollman (2011).

However, since (*ex hypothesi*) they have already evolved a communicative disposition to communicate *that* an action is appropriate or not, they already have at their disposal a signalling game which takes a correct or incorrect action as input and outputs a signal that states that the action was correct or incorrect. As such, when an agent performs an incorrect action, the 'state of the world' is such that it would be appropriate for the sender to send the 'no/stop/wrong' signal, which the receiver will appropriately interpret—since she already understands *this* signal. Thus, miscoordination in our new context is an appropriate input for the pre-evolved context. Furthermore, the signal becomes reflexive in this context, since it has been appropriated to talk about the very disposition upon which the sender and receiver are learning to coordinate. This is precisely the notion of modular composition that was discussed in Chapter 3. Therefore, on the presupposition that the actors have already evolved such a capacity, they need only to be able to compose the two separate games into a single game to communicate that, e.g., *corrective* action should take place.

Below, I present several variations of a base model, which I will call the *correction game*, that are built on this intuitive story, and analyse the results of this ability to take advantage of a previously evolved disposition. In particular, I compare learning rates and occurrence of suboptimal partial-pooling equilibria with the atomic signalling game of the same dimension, where the individuals do not take advantage of a previously evolved disposition, to show whether and in what ways this is advantageous to the players. I conclude this chapter by discussing related work and grounding the empirical plausibility of my model in terms of theoretical linguistic work in the evolution of language.

5.1 The Correction Game Model

Before getting into the details of how the model works, we might consider the following 'high-level' interpretation of what is going on here. Suppose two agents want to communicate.

We model this with a signalling game of some arbitrary dimension, depending upon the case under consideration. The sender sends a signal to the receiver in an attempt to transfer information about the state of the world. Suppose the receiver performs an action that is inappropriate for the state under consideration. In the standard signalling game model, they move on, and a new state of the world is chosen for a new attempt. However, suppose with some probability the sender tries to correct the action. Note that this, in a sense, presupposes that the sender 'knows' what the correct action is; even so, this is not problematic because if the sender were the one attempting the action, she has perfect information about the state of the world and so, even if she does not know *a priori* what the correct action is, she could hit upon the right action quickly via some simple trial-and-error experimentation—the purpose of the signalling game model is to show how such state-act pairs might become associated with signals, thus giving rise to information transfer. So, rather than moving on to a new round, with some probability the sender will attempt to take advantage of a previously evolved communicative capacity for 'correcting' the inappropriate action of the receiver.

We will start by supposing that the agents in the signalling game have already evolved up some command capacity, which we will take to be analogous, in some respect (i.e., the intended outcome action of the command), to 'stop/go' or 'right/wrong' or 'yes/no', etc.

This model is built upon the base of a normal atomic n-game, as it was presented in Definition 1.4, Chapter 1. The correction game proceeds as the atomic n-game usually does: nature picks a state of the world without bias, the sender chooses a signal at random, and the receiver chooses an action at random. If they coordinate, they receive payoff 1 and shift their dispositions proportional to their accumulated rewards. However, the correction game diverges from the atomic n-game when the sender and receiver miscoordinate. When the actors fail to coordinate, the sender attempts to 'correct' the action in question, with some probability, μ. Namely, with probability μ, the agents take advantage of the previously evolved capacity to direct actions via some command—i.e., the sender takes the failure as

input for the sub-game and sends the pre-evolved signal corresponding to 'wrong'. With probability $(1 - \mu)$, they simply move on to the next play of the game, as they usually would, with payoff 0.

Thus, the reward is 1 for one-shot coordination, and if the actors fail and abandon their failure (probability $(1 - \mu)$), then the reward is 0. This 'segment' of the correction game is just the standard signalling game procedure with payoff 1 for coordination and no punishment for miscoordination. The main difference between the correction game and the signalling game is that there is a chance (μ) that the receiver attempts a new action, with the state and the signal remaining fixed. This is under the assumption that, in light of the failure, the sender sends the additional signal that the receiver has done something 'incorrect', as it were. (Note that if the underlying command is already evolved to a signalling system, then we can assume the sender and receiver always coordinate on *this* signal—i.e., the receiver knows how to react to the additional command, by, e.g., trying something new, since this is a pre-evolved disposition.) This extra command from the sender gives the receiver complete knowledge that the particular action she chose was incorrect for the state; nonetheless, she still lacks full knowledge about *which* of the remaining actions *is* appropriate for the state. Thus, this set-up does not presuppose anything about the meanings of the signals being evolved in the main game, nor the sender's ability to communicate these meanings.

There are several possible ways of modelling this process. I will suggest the following. For each run, we will take the reward on the first 'cycle' to be the usual reward for coordination: $u(s, a)$. If the sender tries to correct the receiver's action, due to miscoordination, then the sender and receiver will get some discounted reward conditional upon coordination. This will be given by a discount factor, $\gamma \in [0, 1]$. This discount might be understood as decreasing marginal utility for the additional cost of having to play an extra game—i.e., taking the time to try to correct the receiver's action. For $\gamma = 1$, we have cost-less correction. For $\gamma = 0$, this

extended game reduces to the normal atomic signalling game. Thus, the full specification of the reward is given by $R_{t_n} = \gamma^n \cdot u(s, a)$—the reward, R, on the nth cycle, t_n.

To make clear what I have in mind here, consider the following possible play. Suppose we have a discount factor, $\gamma = \frac{1}{2}$, and a base-payoff, $u(s, a) = 1$, for coordination. The signalling game begins as normal. Nature picks a state of the world; the sender picks a signal; the receiver picks an action. If they coordinate, then they both receive a payoff of 1, and they move on to the next play. If they miscoordinate, then with probability $(1 - \mu)$, they receive a payoff of 0, and they move on to the next play. However, with probability μ, if they miscoordinate, then they play a 'correction game', which can be understood as the sender utilising a previously evolved capacity to inform the receiver that she did something wrong. Here, we assume they always coordinate on the correction game, since it is a pre-evolved 2×2 game, so the receiver tries a new action—namely, if A is the set of actions available to the receiver, and if action $a_i \in A$ led to a miscoordination, the receiver samples stochastically from the set $A - \{a_i\}$, with an associated re-normalised probability distribution $\Delta(A - \{a_i\})$. Suppose the action chosen on the first repetition is a_j.

If the sender and receiver coordinate on the first repetition, then they both receive a discounted payoff $R_{t_1} = \gamma^1 \cdot u(s, a) = \frac{1}{2}$. If they miscoordinate, then, again, with probability $(1 - \mu)$, they abandon the attempt to coordinate, receive payoff 0, and move to the next play. Still, with probability μ, the sender tries to correct the receiver a second time. The receiver tries a new action, sampled from the set $A - \{a_i, a_j\}$, with an associated probability distribution $\Delta(A - \{a_i, a_j\})$. If coordination occurs on the second retry, the sender and receiver get a discounted payoff of $R_{t_2} = \gamma^2 \cdot u(s, a) = \frac{1}{4}$. This continues, with the general discounted reward given by $R_{t_n} = \gamma^n \cdot u(s, a)$ for n attempts to correct the action. See Figure 5.1.

Note that the sender strategy (and the state of nature) are fixed during the correction component of the game; only the receiver tries to correct her action.

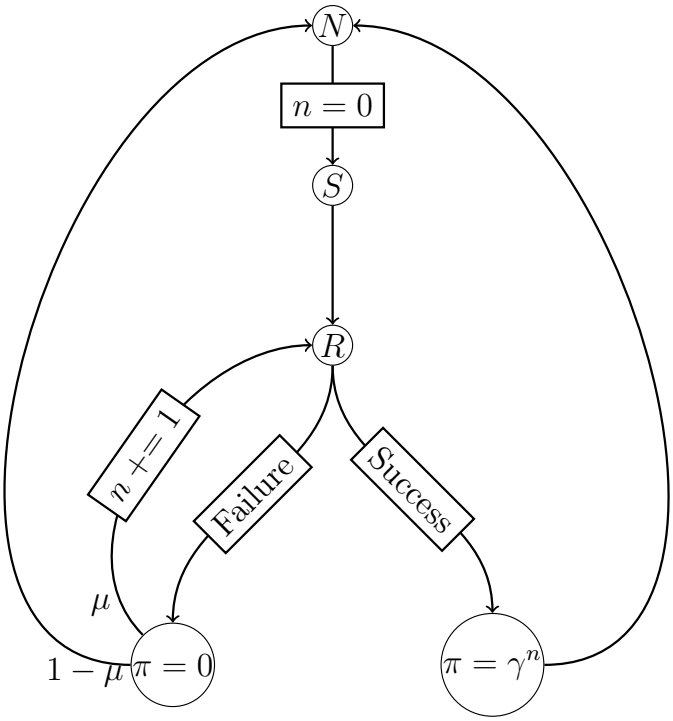


Figure 5.1: Basic *correction game*, with $\pi = R_{t_n} = \gamma^n u(s,a)$, and $u(s,a) = 1$. N denotes nature, S denotes the sender, and R denotes the receiver. $\mu \in [0,1]$ is a probability. n is the 'counter' that is used to discount the rewards.

5.2 The Simple Correction Game: Cue-Reading

I begin by examining a correction game where the sender's dispositions are already fixed. This is, in effect, a cue-reading game, with the possibility for the sender to attempt to correct the receiver's action when her chosen action fails to achieve coordination on that particular state.

We examine an 8×8 cue-reading game. States are equiprobable, and the payoff for success is $u(s,a) = 1$. The sender begins with dispositions such that

$$P(m_i|s_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

Each run consists of 500 individual plays of the game.[3] We examine the results of 1000 runs.[4]

We have two new parameters that can be varied. First, the probability, μ, with which the sender and receiver repeat a failed play; second, the cost for repetition, γ. To get a reasonable picture of how these parameters affect learning in the underlying cue-reading game, we examine the 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$.

5.2.1 Results

We must be careful in interpreting the results of our simulations. What is common is to calculate the *cumulative success rate* of a particular run by merely counting the number of plays where the sender and receiver successfully coordinated and dividing this by the total number of plays. Early failures get washed out as the number of plays per run increases. We can then examine the *proportion* of runs that have a cumulative success rate surpassing some threshold.

The threshold for success is not arbitrary. The 8×8 cue-reading game has a large number of partial pooling equilibria. These are polymorphic traps where the sender and receiver might get caught. The most efficient sub-optimal strategy for the receiver (given the sender's dispositions are fixed in the cue-reading game) occurs when the receiver performs the appropriate action for 7/8 of the signals and pools her strategy on the 8th signal. These pooling equilibria allow for a maximum communicative success rate (and a maximum expected payoff) of

[3]Note that this is an extremely low number of plays, but individuals learn quickly under reinforcement learning when the sender's dispositions are already fixed. In an 8×8 cue-reading game, after $10,000$ plays, the sender and receiver have a cumulative success rate greater than 0.95 on almost all (0.975) of the runs, and every run results in a cumulative success rate greater than 0.90. As such, signalling systems are guaranteed in a fairly short amount of time in this particular case—thus, we examine shorter-run results to see whether we cannot arrive at signalling even faster with correction. The question of partial pooling is less of a concern here.

[4]The simulations were run in Python 2.7, and the resultant data was compiled using MatLab.

0.875. Thus, we ought to set our threshold for success to at least 0.875 to see whether the sender and receiver have escaped these polymorphic traps.[5]

However, we note that when $\mu = 1$, the receiver will necessarily retry actions until she hits upon a successful one. Thus, we should expect that for $\mu = 1$, the sender and receiver will *always* surpass the threshold for success. Indeed, this is precisely what happens (after 1000 plays per run), as shown in Table 5.1. More complete data are shown, for comparison, in Figure 5.2.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.000**		
Repetition	0.25	0.000	0.000	0.000	0.000
Probability	0.50	0.003	0.021	0.082	0.214
μ	0.75	0.954	0.999	0.999	1.000
	1.00	1.000	1.000	1.000	1.000

Table 5.1: Proportion of successes for short-run simulation results for correction game with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (10^3 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

Now, one might worry that the cumulative success rate is not accurately capturing successes in the cue-reading game with correction, since 10^3 plays are really $10^3 + C$ plays—where C is the number of repeat attempts at success which take place on a given run. The average number of repetitions in each case is shown in Table 5.2. This further highlights the effects of cost-less correction—because the accumulated rewards are shifted more for lower-cost correction, the likelihood of choosing the correct action in a future play is increased more than when correction is expensive. Repetitions increase monotonically as the probability of repeating increases, and they decrease monotonically as the discount factor increases—i.e., as the cost for correction decreases. For example, a reduction in the number of repetitions

[5]It is possible that a suboptimal random walk spends some time above this threshold before settling in to a polymorphic trap. The probability that this happens decreases significantly as the threshold increases, or as the number of plays increases.

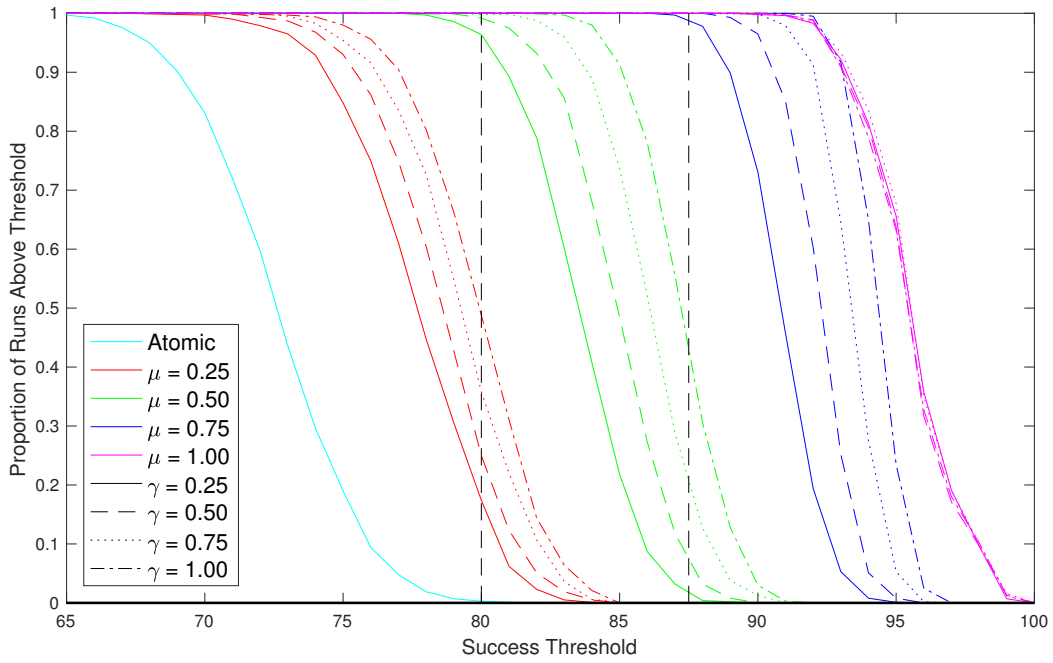


Figure 5.2: The proportion of successful runs above thresholds $[0.65, 1.00]$ shown for each combination of parameters, $[\mu, \gamma]$. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic			**0**	
Repetition	0.25	78	76	73	70
Probability	0.50	176	161	149	138
μ	0.75	303	264	228	199
	1.00	484	386	307	247

Table 5.2: Average number of repetitions made in the cue-reading correction game

when the sender is guaranteed to try to correct the receiver's behaviour ($\mu = 1.00$) implies that the sender and receiver are failing to coordinate less often. In the worst case, we see almost a 50% increase in 'plays'.

There is perhaps good reason not to interpret our data this conservatively: for one, a repeat does not constitute a full play of the game to the extent that no new state nor signal is chosen during a repeat. Even so, we can correct for this in the following way.

The simulations were re-run, and new data was gathered thus. The sender and receiver are allowed 500 plays to try to learn a signalling convention. We then let them communicate according to whatever convention they have settled upon (or begun to settle upon) for 1000 plays. We count successes and failures during the communication period only, not during the learning period—thus, we ignore the failures that occur during learning. This approximates the expectation of success in the same way as looking at the urn contents after 500 rounds and calculating the exact expectation.

Since the strategies that evolve are going to vary stochastically, we take an average of 1000 runs. The adjusted success rates for 1000 runs under this success measure are shown in Table 5.3. The data vary significantly from those of Table 5.1. In particular, there appears

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.000**		
Repetition	0.25	0.006	0.017	0.040	0.091
Probability	0.50	0.017	0.089	0.371	0.701
μ	0.75	0.045	0.359	0.874	0.991
	1.00	0.113	0.752	0.995	1.000

Table 5.3: Adjusted proportion of successes for short-run simulation results for correction game with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (5×10^2 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

to be less pooling across all discount factors when the correction probability $\mu = 0.25, 0.50$, whereas more pooling (than the data in Table 5.1) seems to be exhibited for more probable repetitions, $\mu = 0.75, 1.00$. However, we should note three things here: first, there are half as many plays where learning occurs (500 as opposed to 1000); second, initial miscoordination during learning is not counted in the latter case—thus, we should expect slightly more successes than if initial failures are counted; finally, since the sender and receiver do not take advantage of the correction capacity during the communication period, successes here really do constitute successes. The general qualitative results still hold: fixing the discount factor, γ, an increase in μ corresponds to an increase in success; fixing the repetition probability,

μ, a decrease in cost for repetition also corresponds to an increase in success. Thus, these results are robust regardless of what one counts as a success. Again, more complete data are shown, for comparison, in Figure 5.3. Note that the successes for the atomic case are

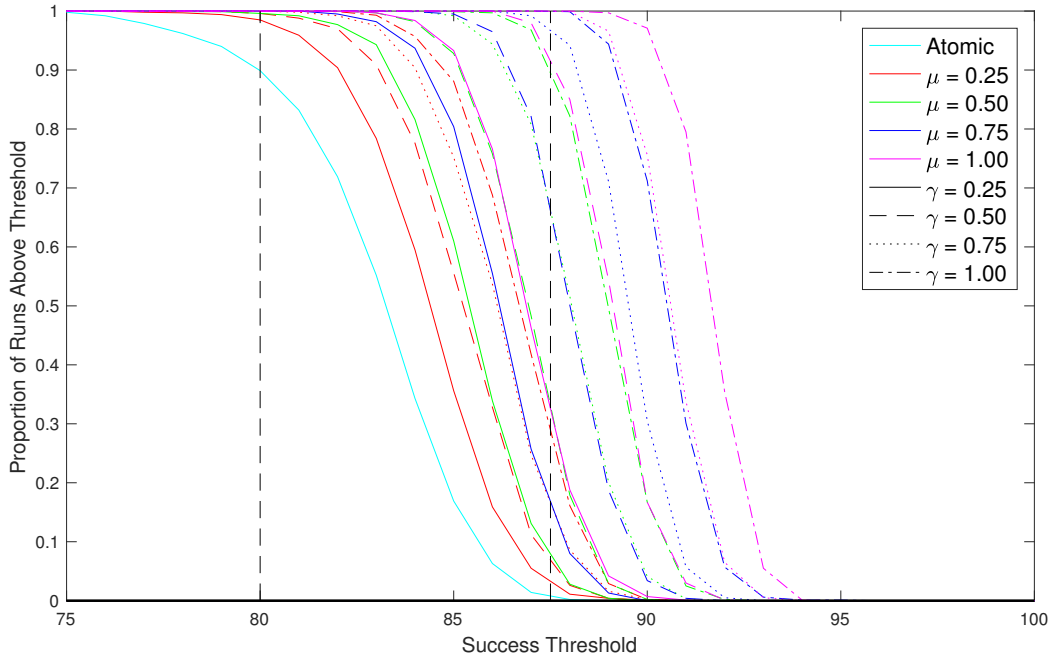


Figure 5.3: The adjusted proportion of successful runs above thresholds $[0.75, 1.00]$ shown for each combination of parameters, $[\mu, \gamma]$. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively

shifted up since we are not counting the initial failures during the learning period. The data is less clearly differentiated; again, holding fixed one parameter, we see a monotonic increase in successes as we vary the other parameter (either holding fixed the cost and increasing the probability of correction, or holding fixed the probability and decreasing the cost).

We might note that this cue-reading game with correction is really only half of the model suggested by the story at the outset. Correction (modelled as a pre-evolved 'yes/no' meta-game) only occurs when the receiver fails to coordinate with the sender's intended meaning. This is the 'no' component: the receiver is corrected with some probability only when she does something wrong. We might add the 'yes' half of the correction as follows: suppose the

receiver coordinates on the first try; with some probability, μ, the sender further reinforces this behaviour by telling the receiver that she did something *right*, by using their pre-evolved disposition. Thus, the receiver receives an additional payoff, given by $\gamma \cdot u(s, a)$. Note that there is no repetition when the action is successful, so there is a one-shot reinforcement, which occurs with probability μ.[6]

In this case, even under the worst parameter combinations, every combination of parameters resulted in 100% of the runs exceeding the pooling-threshold cumulative success rate of 0.875 after only 500 plays. Indeed, most combinations do significantly better than this. The proportion of runs resulting in a cumulative success rate greater than 0.95 are shown in Table 5.4.[7]

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.000**		
Repetition	0.25	0.670	0.792	0.877	0.933
Probability	0.50	0.820	0.950	0.981	0.998
μ	0.75	0.883	0.987	0.999	1.000
	1.00	0.947	0.994	1.000	1.000

Table 5.4: Adjusted proportion of successes for short-run simulation results for correction game ('yes' *and* 'no') with pre-evolved sender dispositions (cue-reading) under a variety of discount factors and repetition probabilities (5×10^2 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.95

This model was built upon a cue-reading game, rather than a signalling game; thus, the sender's dispositions were fixed at the outset. Can the sender and receiver co-evolve their strategies, while taking advantage of their pre-evolved corrective dispositions?

[6]We might imagine that the normal payoff for coordination is given by nature, as is the case in the atomic signalling game, whereas this additional payoff is given by the sender; nonetheless, it need not be the sender who tries to correct the receiver's behaviour—see the discussion in Section 4.5 below.

[7]Note the increase in the success threshold; every run results in 100% of the plays having a cumulative success rate greater than 0.9 (and so greater than 0.875). No pooling whatsoever occurs after 500 plays.

5.3 The Simple Correction Game: Signalling

In this section, we examine the effects of combining the possibility for correction with the full signalling game, as opposed to the asymmetric cue-reading game. However, some care is required here. If the sender and receiver begin correcting too early, then correction will obviously not help them to evolve a signalling convention, because the sender would effectively be trying to correct the receiver's behaviour while she is still not yet fixed upon what her signal actually means. Early on, she might use m_0 to mean s_0, and correct the receiver when she chooses an act other than a_0. Later in the game, she might use m_0 to mean s_1. The propensities are highly variable at the outset. Thus, we must allow the sender and receiver to start to learn a signalling convention before they can utilise the correction game. Unfortunately, the answer to the previous question is decidedly: *no*. The efficiency and efficacy seen in the cue-reading game with correction do not generalise to the signalling game, so the results here are limitative. Even so, what happens is somewhat subtle, so it is worth going through with some care.

5.3.1 Results

Here we examine the short-term results for a simple correction game built on top of a full atomic 8-game, under a variety of parameters. States are equiprobable, and the payoff for success is $u(s, a) = 1$. Each run consists of 10^5 individual plays of the game, and we examine the results of 1000 runs. The sender and receiver are allowed a learning period of $25,000$ plays before trying to correct behaviour using their pre-evolved dispositions. Again, we examine the correction game with 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$. Again, when *either* $\mu = 0$ or $\gamma = 0$, the correction game is equivalent to the atomic signalling game. In the former case, the probability of repetition is 0, so the sender and receiver never retry. In the latter case, any number of repetitions

results in a payoff of 0, so even if the sender and receiver repeat until a success, they do not reinforce that success.

The cumulative success rates, with a threshold of 0.875 for success, of these several parameters are shown in Table 5.5. In general, it appears that correction helps the sender and

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.331**		
Repetition	0.25	0.413	0.432	0.406	0.428
Probability	0.50	0.765	0.812	0.805	0.765
μ	0.75	0.917	0.914	0.901	0.892
	1.00	1.000	1.000	1.000	1.000

Table 5.5: Proportion of successes for short-run simulation results for correction game under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

receiver to learn a signalling convention; however, this is again under the assumption that a 'success' is just coordination on a given play, ignoring the repetitions. Thus, when the repetition probability is 1, the proportion of successes is going to be 1 trivially—the sender and receiver repeat a failure until it turns into a success. We can obtain more accurate results of whether the sender and receiver are avoiding pooling by examining their success during a communication period, after an initial learning period.

Successes are re-calculated as follows: The sender and receiver have an initial learning period of $25,000$ plays where they learn atomically. They learn for the rest of the 10^5 plays by using correction. Finally, we count successes during a 1000-play 'communication period', which approximates the actual *expectation* of success; the results of 1000 runs are examined. These adjusted data are displayed in Table 5.6 Note, first and foremost, that the successes in the atomic case are increased. This is because we are not counting the failures during the initial learning period.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.548**		
Repetition	0.25	0.468	0.409	0.341	0.260
Probability	0.50	0.385	0.249	0.164	0.078
μ	0.75	0.349	0.183	0.037	0.000
	1.00	0.279	0.076	0.000	0.000

Table 5.6: Adjusted proportion of successes for short-run simulation results for correction game with co-evolved sender dispositions (signalling) under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

In comparison to the atomic game under this success measure, the correction game does categorically worse. In a way, however, this makes sense. The sender is correcting behaviour without herself knowing what a signal is supposed to mean. Thus, correction is too aggressive. Note also that the correction game here performs worse when the cost for payoff is decreased. This is the opposite of what happens in the cue-reading game. Again, this is because cost-less correction has a more substantial effect on propensities, which, we have now seen, is detrimental when the sender's disposition is not yet fixed. This is further highlighted by the fact that the number of repetitions *increases* as the cost of repeating goes down in the full signalling game. The average number of repetitions in each case are shown in Table 5.7.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0**		
Repetition	0.25	2136	2158	2199	2154
Probability	0.50	4608	4459	4572	4731
μ	0.75	6889	6924	7223	7534
	1.00	9409	10044	10317	12239

Table 5.7: Average number of repetitions made in the signalling correction game.

We are obtaining a clearer picture of how and when correction, in the form of a pre-evolved disposition, might positively affect learning a new disposition. In the cue-reading game, the sender is determined that the signal means such-and-such, so correction is appropriate. In the signalling game, she is also learning a conventional meaning for her signals, so it makes

little sense for her to insist very early on that the receiver has done something wrong. This is further highlighted by the fact that, when we include the 'yes' component so that the sender reinforces correct behaviour on the receiver's part, the results are even worse than those in Table 5.6. This is because the correcting behaviour on the full correction game is even more aggressive than the behaviour on the correction game with only the 'no' component.

These results are more subtle than just that they fail to help avoid pooling. In particular, even though the sender and receiver end up pooling their strategies more often when the sender is too aggressive, the *expected payoff* remains largely unchanged in every case. These data are shown in Table 5.8 The variance between the expected payoff between these 1000

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.8982**		
Repetition	0.25	0.8975	0.8961	0.8926	0.8944
Probability	0.50	0.8910	0.8927	0.8890	0.8844
μ	0.75	0.8936	0.8909	0.8850	0.8780
	1.00	0.8947	0.8849	0.8797	0.8587

Table 5.8: Average expected payoff for short-run simulation results for correction game under a variety of discount factors and repetition probabilities (10^5 plays per run, 1000 runs).

runs is effectively equivalent—approximately 0.005—in every case. Indeed, when the sender and receiver escape pooling equilibria, the correction game does no worse than the atomic signalling game. However, they tend to get caught in pooling more often the more aggressive the sender is in trying to correct the receiver's behaviour.

If we decrease the initial period in which the sender and receiver learn atomically, they do even worse still. This is because their dispositions are even less fixed than when they start with a period of atomic learning. If we increase the initial period where they learn atomically, then, as the period of atomic learning approaches the total number of plays, the results limit toward the atomic results. Thus, it is not possible that the sender and receiver do better

than the atomic case when they have a pre-evolved corrective disposition at their disposal. The best they can do is as good as the atomic case.

5.4 The Simple Correction Game: Signalling with Invention

The reason why the results of the general signalling game were limitative, it was suggested, is because the sender is unwarranted in attempting to correct the receiver's behaviour: she is also learning what the signals mean, and so it makes little sense for her to insist upon a particular meaning at the outset when the meanings of the signals are still fluctuating. In this section, we examine the general signalling game *with invention*, which is a modified version of the Hoppe-Pólya urn model (Hoppe, 1984) of *neutral* evolution—where many mutations do not convey a selective advantage. This extension was mentioned briefly in Chapter 1.[8]

In the atomic case, the signalling game with invention works in this way. Suppose we have 8 states of nature and 8 appropriate actions. The sender begins with no signals; she has 8 urns for each of the states, and each urn contains one black ball—the *mutator*. The receiver begins with no urns. On a particular round, nature picks a state of the world with some probability—again, we assume nature is unbiased, so each state is equiprobable. The sender sees the state and selects a ball at random from the corresponding urn. If she selects the black ball, she invents a new signal, by placing a ball for that signal in the urn. This is the signal that she sends to the receiver. The receiver is attentive to new signals: when the signal sent is novel, she creates a new urn for that signal, containing 8 balls for each of the possible actions and then selects an act from that urn. When the sender and receiver coordinate, they reinforce by adding another ball of the same type to the urn from which it

[8]See also the discussion in Skyrms (2010a); Alexander et al. (2012).

was chosen. The game is then repeated with a new state of nature. This is a Hoppe-Pólya urn model with *differential* reinforcement.

Note that the sender never reinforces her propensity to invent, so the rate at which the sender invents new signals decreases over time. Thus, when a state of nature is seen for the first time, the sender invents a signal to communicate with the receiver. If they coordinate, then there is a 2/3 probability in the future that the same signal will be sent in that state and 1/3 probability that a brand-new signal will be invented in that state. If they miscoordinate, then there is a 1/2 probability that the sender will retry the same signal in that state, and 1/2 probability that the sender will send a new signal in that state.[9]

5.4.1 Results

Here we examine the short-term results for the full correction game built on top of a full atomic 8-game, with invention, under a variety of parameters. The sender begins with no signals. States are equiprobable, and the payoff for success is $u(s, a) = 1$. Each run consists of 1.5×10^4 individual plays of the game, and we examine the results of 1000 runs. The sender and receiver begin with no atomic learning period since this was the worst-case in the general signalling game with correction. Again, we examine the correction game with 16 combinations of $\mu = [0.25, 0.50, 0.75, 1.00]$ and $\gamma = [0.25, 0.50, 0.75, 1.00]$, as compared with the atomic signalling game with invention.

The sender and receiver learn signalling dispositions with invention over the course of 1.5×10^4 individual plays, and then, to gain a more accurate representation of what counts as a success, they communicate for 1000 plays. We calculate the average number of successes over the course of the communication period, where the threshold for success is 0.875. These results are shown in Table 5.9 As opposed to the atomic signalling game, where the sender's

[9]This is related (though due to the differential reinforcement not equivalent) to the *Chinese Restaurant Process*; see, e.g., Aldous (1985); Pitman (1995).

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**0.071**		
Repetition	0.25	0.166	0.378	0.528	0.664
Probability	0.50	0.316	0.551	0.547	0.409
μ	0.75	0.446	0.518	0.025	0.004
	1.00	0.495	0.313	0.007	0.000

Table 5.9: Adjusted proportion of successes for short-run simulation results for correction game ('yes' *and* 'no') with co-evolved sender dispositions (signalling) plus invention under a variety of discount factors and repetition probabilities (1.5×10^4 plays per run, 1000 runs). A run is counted as a success if the proportion of successful plays for that run in greater than 0.875

unjustified aggressiveness in correcting the receiver is detrimental to them both, we see that correction again helps the sender and receiver to coordinate, when the correction is not too often and not too inexpensive. In the case where the sender always tries to correct the receiver, and correction is cost-free, they do worse. Again, this should be unsurprising, given that the sender and receiver are still learning to coordinate. Thus, the sender's being too aggressive is still detrimental to them both; however, in almost every other case, correction has a significant impact on learning to signal. More complete data are displayed in Figure 5.4. We look more closely at the results that are beneficial in Figure 5.5, centred about the success threshold. Correction almost always helps when the sender invents new signals. Even so, it is also known that inventing new signals can help to avoid pooling equilibria in general (Alexander et al., 2012). There are further subtleties to the signalling game with invention; we can look at the average number of signals invented in each case to see how efficiently the sender is inventing while the sender and receiver are learning. The average number of signals invented in each case are shown in Table 5.10. As is evident, correction not only helps the sender and receiver to avoid pooling equilibria, it helps them to do so more efficiently—i.e., by creating fewer signals at the outset. In the best case in terms of the proportion of successes ($[\mu, \gamma] = [0.50, 0.75]$), they are almost 8 times more successful than in the atomic case, and they can achieve this rate of success more efficiently, with 2/3 the number of signals. With

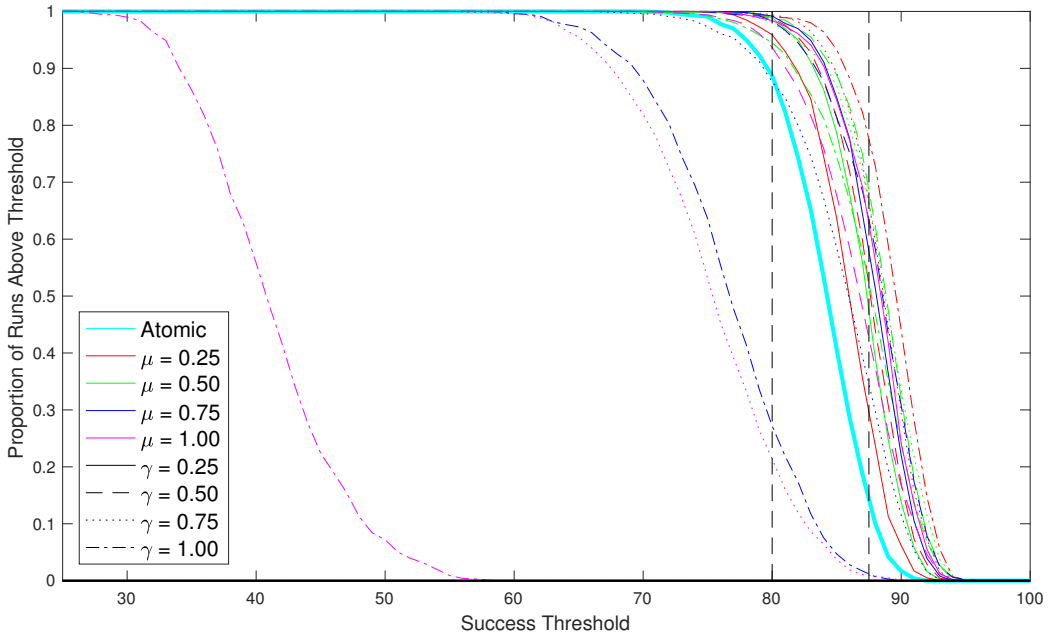


Figure 5.4: The proportion of successful runs above thresholds $[0.25, 1.0]$ shown for each combination of parameters, $[\mu, \gamma]$, compared with the atomic game. The vertical dashed lines indicate the thresholds 0.80, and 0.875, respectively. The atomic case is bold for clarity.

		Discount Factor, γ			
		0.25	0.50	0.75	1.00
	Atomic		**89**, (100.00%)		
Repetition	0.25	83, (93.25%)	80, (89.89%)	74, (83.15%)	69, (77.53%)
Probability	0.50	78, (87.64%)	69, (77.53%)	62, (69.66%)	57, (64.04%)
μ	0.75	70, (78.65%)	59, (66.29%)	47, (52.81%)	38, (42.70%)
	1.00	72, (80.90%)	58, (65.17%)	48, (53.92%)	38, (42.70%)

Table 5.10: Average number of signals at the end of 1.5×10^4 plays of the signalling game with invention across a variety of parameters, and comparison with atomic case

about half of the signals, they can coordinate more than 3 times as often than in the atomic case ($[\mu, \gamma] = [0.75, 0.75]$).

Note further that the invention of signals captures a notion of communicative development which is *diachronic* rather than synchronic. Recall that one of the charges against the assumptions in the signalling game model was that the sender and receiver start with a fixed number of messages (Hurford, 2012). The correction game from signalling with invention

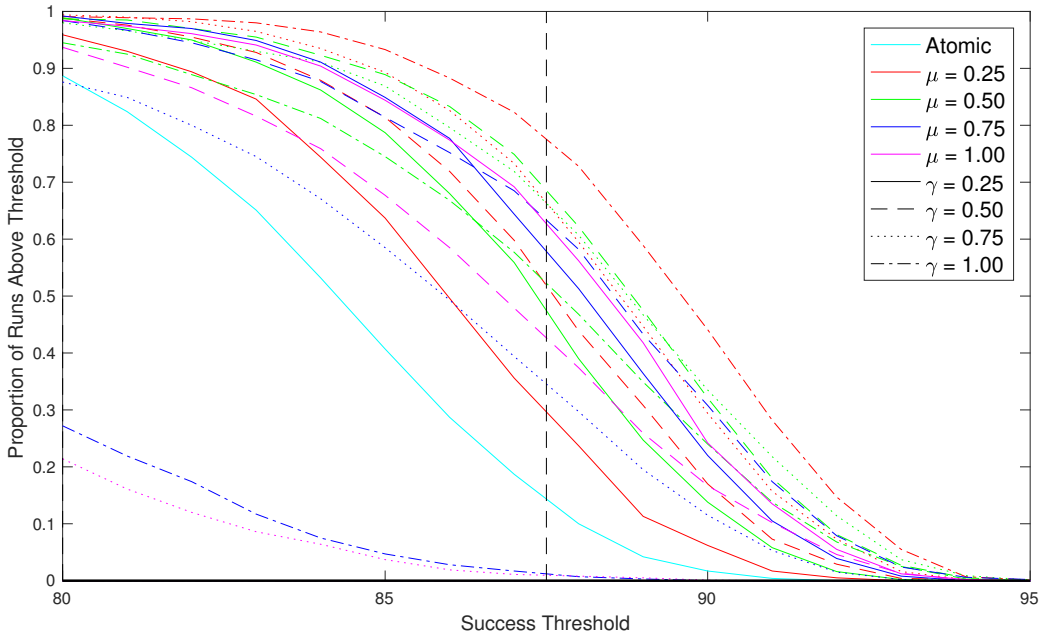


Figure 5.5: The proportion of successful runs above thresholds $[0.8, 0.95]$ shown for each combination of parameters, $[\mu, \gamma]$, compared with the atomic game, and centred about the pooling threshold, 0.875 (dashed vertical line)

captures a subtle process, which we might take to be more realistic than the signalling game with fixed signals, in the following sense. The sender might invent a signal for representing a particular state. When the sender invents a new signal, holding everything else fixed, she is *not* just choosing randomly; instead, she is creating to communicate a particular thing. If the receiver fails to understand, it stands to reason that, again, holding everything fixed, the sender might *insist*.

5.5 Discussion

In summary, correction is helpful in several cases, though this is not universally true. I repeat the claim made at the outset: this process presupposes little over and above the standard signalling game, as far as cognitive sophistication goes. When the sender tries to

communicate that the receiver does X, and the receiver fails to understand the meaning of X, the sender corrects—'no, no no, do X!'—the *meaning* of X is *still* unknown, but the meaning of *no* is known, by the pre-evolved disposition. Thus, the receiver communicates perfect information that the action tried was incorrect, but the receiver still needs to *learn* which act *is* correct.

When the sender has pre-evolved dispositions, correction not only helps to avoid pooling equilibria in the 8×8 case, but it allows the receiver to learn how to coordinate to the sender's fixed disposition very quickly—in the best case, the sender and receiver have surpassed the pooling threshold every time after only 500 plays.

However, these results do not generalise to the signalling game, where the sender and receiver both learn their dispositions at the same time. This should not come as a surprise—early on in the game, when the sender's dispositions are highly variable, it makes little sense for her to insist that the receiver has done something wrong. Even in this case, though, those runs that surpassed the pooling threshold perform no worse than in the atomic case. This highlights that correction is helpful, as long as the sender knows what it is that she is correcting.

Finally, in the signalling game with invention, we saw that correction once again has a significant impact on learning, both in terms of speed and avoiding pooling equilibria, in most cases. The caveat here is that if the sender is too aggressive in trying to correct behaviour, it can be detrimental to learning; however, correction was only detrimental in two cases: when the sender always or almost always corrects the receiver's behaviour, and there is no cost to correction. In every other case, correction helps learning. Furthermore, the sender and receiver can learn to signal more efficiently in the sense of requiring fewer synonymous signals: often, they end up inventing between 2/3 and 3/4 of the signals invented in the atomic case.

Note also that we interpreted the correction model as the sender correcting the receiver. Nothing necessitates this interpretation. Indeed, in social animals, correction may not be done by the sender, but rather by a conspecific who is a bystander. This may occur in the case of adults correcting juveniles.[10] Thus, several different interpretations are allowed by the generality of this model. Each of the agents in the model may be understood as functional components of a social group or an individual agent.[11] Further, the agents in the meta-game need not be the same as the agents in the base game. We might imagine an observer watching a signalling interaction, keeping track of what signals and actions are correct given the conventions of the agents in the base-game, and occasionally correcting those actions that are inconsistent with previous behaviour.

5.5.1 Relation to Previous Work

It was assumed in every case that the correction game involved a *pre-evolved* disposition, of which the sender variably takes advantage. We might wonder whether the sender and receiver can co-evolve this disposition as they are learning to signal.

Barrett (2016) examines how a metalanguage might co-evolve with the language it describes. In the first model he describes, the meta-game co-evolves to indicate the success and failure of the base-game agents as they evolve signalling dispositions, in the sense of the atomic signalling game. In the second model he describes, the sender attends to the co-evolving conventional use of expressions in the base game. Thus, the meta-game evolves to track whether the expressions of the base-game are *true* (in a simple, pragmatic sense), and so provides a sense in which the base language might be understood to have evolved propositional content (1–2).

[10] With thanks to Brian Skyrms for pointing this out.
[11] See Barrett et al. (2018).

The base game that Barrett (2016) describes is an atomic 4-game, where the agents evolve their dispositions under simple reinforcement learning. The meta-game is an atomic 2-game, which takes the success or failure of the sender and receiver in the base-game as input. The meta-game sender and receiver also learn by simple reinforcement. The state of nature, which the sender observes, may be obtained either by looking at the state and act of the base-game to see whether they match or examining whether or not the sender and receiver in the base-game received a payoff, for example. Actions in the meta-game correspond to success or failure, and the meta-game sender and receiver reinforce just in case the meta-game receiver's action matches the meta-game state. Thus, the meta-game agents learn from their observations of the evolving dispositions of the base-game agents. Barrett (2016) reports that on simulation, the meta-game receiver exhibits a cumulative success rate of better than 0.95 on better than 0.99 of the runs of the model, with 1000 runs of 10^6 plays per run.

This may seem unsurprising, given the results of Argiento et al. (2009) for the atomic 2-game. However, what happens here is slightly more subtle: Barrett (2016) points out that even the 2×2 signalling game can get stuck in sup-optimal pooling equilibria when nature is biased.[12] This is relevant because as the base-game evolves, the successes become more frequent. But the input (nature) in the meta-game is just the successes and failures of the base-game; thus, nature in the meta-game is unbiased at the outset but becomes biased as the base-game evolves. Nature in the meta-game is strongly biased toward success over time.

Even so, the base game is more complicated than the meta-game. Therefore, the meta-game evolves more quickly than the base game. Hence, by the time nature in the meta-game becomes strongly biased toward success, the meta-game sender and receiver have already evolved a signalling system, which clearly demarcates the two input states. This happens 0.98 of the time, with 1000 runs of 10^6 plays per run.[13]

[12]See the discussion in Skyrms (2010a). See Hofbauer and Huttegger (2008) for a proof in the context of the replicator dynamic. These points were discussed in detail in Chapter 1.

[13]In a second model, the meta-game sender tracks whether the base-game sender used the signal that is *customary*, given what the agents in the base game have been doing. This simple game can be extended

It was assumed, in this chapter, that the sender and receiver in the correction game have already evolved a disposition to communicate some binary 'yes/no' signal successfully. In the correction game, they *use* this pre-evolved disposition to try to correct the receiver's action when the sender and receiver miscoordinate. I take the results of Barrett (2016) to be sufficient for an affirmative answer to the question of whether or not the correction meta-game can co-evolve alongside the signalling game. The set up for a co-evolutionary correction game, where the meta-game agents learn the 'yes/no' signalling disposition where the input (states of nature) for the meta-game are given by the success of failure to coordinate in the base-game, is almost equivalent to the 'true/false' model that Barrett (2016) presents. The main difference in the correction game is that the *output* of the meta-game also affects the dispositions in the base game. However, we can imagine that the meta-game evolves in the atomic period of the correction game. Since the meta-game is significantly less complex than the base-game, the meta-game will evolve faster. Thus, by the time the sender and receiver start to *utilise* their dispositions, they should have already coordinated upon a signalling convention in the meta-game.

Due to the subtleties described in Barrett (2016), these results will also be limitative: the meta-game will not necessarily evolve for significantly more complex base-games, like the 8×8 signalling game. This is because, in this case, there are 8^{16} (almost 300 trillion) possible combinations of strategies, but only $8!$ (slightly more than $40,000$) of these are signalling systems. At the outset, it is significantly more probable that the input for the meta-game will be a failure rather than a success (0.875 in the atomic case).

Indeed, when we examine the co-evolution of a 2×2 meta-game that tracks truth and falsity, taking successes and failures from the atomic 8×8 base game as input, only around 20% of

to include the co-evolution of a pragmatic sense of *probability*; such a model is discussed further in Barrett (2017).

the meta-game plays surpass the pooling threshold of 0.75.[14] However, what is typical here is that the sender perfectly partitions the states of nature (success and failure) by the two available signals, and the receiver learns the meaning of the failure signal but is indifferent between her actions for the success signal. Thus, as successes become more frequent, it is probable that she will eventually learn the meaning of this signal—that is, she does not bias her action toward 'failure' for each signal. This is for the atomic 8-game. It remains to be seen whether or how the signalling game with invention affects these meta-game propensities, given that the output of the meta-game feeds back into the base game.

5.5.2 Affirmation and Negation from a Linguistic Perspective

Why is this particular pre-evolved disposition relevant to the evolution of communicative capacities? Negation is a universal category of human language Dahl (1979)—every natural language at least can express clausal negation; however, the way that different languages negate varies. In English, and other Indo-European languages, sentence negation is frequently realised by the negative participle 'not'.[15] For example,

(1a) Atlas believes that Sarah is not home.

(1b) Atlas does not believe that Sarah is at home.

In some languages, though, sentence negation is expressed by a negative verb. For example, in Tongan, the negator *ikai* acts as a higher verb which takes the corresponding affirmative clause as its complement, and *ke* is a subjunctive marker, which marks the complement clause as subordinate (Churchward, 1953, 56):

[14]This is for $25,000$ plays per run and $10,000$ runs—corresponding to the atomic learning period in the model presented in Section 5.3. The results are essentially equivalent when we increase this to 10^7 plays per run.

[15]This is typically referred to as 'standard negation'. This terminology originates in Payne (1985).

(2a) na'e 'alu 'a siale.
PST *go* ABS Siale
'Siale went'.

(2b) na'e 'ikai ke 'alu 'a siale.
PST NEG SBJN *go* ABS Siale
'Siale did not go'.

Even so, Miestamo (2007) notes that this type of negation is marginal.

Along with truth-functional negation, a large range of word-formation processes can be used to coin negative meanings. For example, in English, these word-formation processes include *prefixation*, *suffixation*, *compounding* and *conversion*. Morphologically, negation is quite complicated. For example, in English negation may be expressed through several *negative derivational affixes*: *de-*, *dis-*, *in-*, *non-*, *un-* and *-less*.

In most languages, negation systematically either precedes or follows the verb. Dryer (1988) studies the placement of the marker of sentential negation in relation to the subject (S), object (O) and verb (V)—three main clausal elements—in a worldwide sample of 345 languages. His results suggest that SOV languages are most commonly either SOV*Neg* or SO*Neg*V. *Neg*SOV and S*Neg*OV languages are infrequent. SVO languages are most commonly S*Neg*VO, and V-initial languages are almost always *Neg*V (i.e. *Neg*VSO or *Neg*VOS). In 70% of the 325 languages surveyed, Dryer (1988) finds that the negation marker is placed before the verb.[16]

It has been claimed that no animal communication system has a notion of negation (Horn, 1989; Jackendoff, 2002). Even so, it is suggested that some variety of *pre-logical* negation might be available in the cognitive representation of higher animals—this is consistent with the view that non-human animal communication systems lack recursion (Hauser et al., 2002);

[16]See also the discussion in de Swart (2010).

bona fide truth-functional negation in natural language is recursive to the extent that, semantically, it takes an arbitrary proposition, ϕ, and creates a new proposition, $\neg\phi$, where ϕ may itself be a negated proposition.

Negation, in natural languages, is complicated for a variety of reasons. First, the logic of affirmation and negation is asymmetric: negations are generally less valuable, less specific, and less informative than affirmations (Plato, 1921b). Aristotle (1995b) held that affirmations have ontological, epistemological, psychological, and grammatical priority over negations (996b14–16). Further, negations are morphosyntactically more *marked* and psychologically more difficult to parse (Just and Carpenter, 1971; Horn, 1989). In some sense, negation presupposes affirmation: 'the feeling is as if the negation of a proposition had to make it true in a certain sense in order to negate it' (Wittgenstein, 1953, §447).[17] Finally, affirmation usually introduces a proposition into the 'discourse model'; in contrast, negation—in its 'chief use" (Jespersen, 1917, 4), its 'most common use' (Ayer, 1952, 39), its 'standard and primary use' (Strawson, 1952, 7)—is directed at a proposition that is already in, or that can be accommodated by, the discourse model.[18]

Protolanguages need not contain propositions nor truth-functions, though these would at least need to emerge somewhere in the transition from protolanguage to language. In a review of the relevant literature, Heine and Kuteva (2007) suggest that trained animals can develop notions of rejection and refusal, and even of non-existence.[19]

In addition to the omnipresence of negation in natural languages, negation and affirmation may have evolved early on, and so serve as linguistic 'fossils' of a one-word stage of the evolution of language, wherein single utterances serve holophrastic purposes and are not integrated into a more extensive combinatorial system (Jackendoff, 1999).[20] It is irrelevant

[17]See also Givón (1978).

[18]See also the discussion in Horn and Wansing (2017).

[19]See also Patterson (1978); Premack and Premack (1983); Herman and Forestell (1985); Savage-Rumbaugh (1986); Pepperberg (1999); Zuberbühler (2002).

[20]See also the discussion in Progovac (2015).

that no known animal communication system contains a generalised negation; instead, what is important is the *signal* understood as a proto-command of encouragement or negation.

Several such one-word utterances exist in language: Jackendoff (1999) points to sudden, high-affect utterances, such as *ouch!*, *dammit!*, *wow!* and *oboy!*, and suggests that

> These exclamations have no syntax and therefore cannot be integrated into larger syntactic constructions[.] ... They can remain in the repertoire of the deepest aphasics, apparently coming from the right hemisphere. There also exist situation-specific utterances such as *shh*, *psst*, and some uses of *hey* that have almost the flavor of primate alarm calls. Though the *ouch* type and the *shh* type both lack syntax, they have different properties. ... Further single-word utterances include the situation-specific greetings *hello* and *goodbye* and the answers *yes* and *no*.

Hurford (2012) highlights the fact that such one-word phrases (along with pragmatic inference) allow for the possibility of conveying propositional information without the benefit of syntax.

This is precisely the type of linguistic fossil that is suggested by a pre-evolved disposition for correction.

5.5.3 Future Work

There are several variants of this simple correction model that might be of interest. For example, it was supposed that if the sender and receiver abandon an attempt to correct after n repetitions, then they receive a payoff of 0. However, we might suppose that there is a (time/effort) cost for correction, such that the payoff is discounted even when the repetitions do not end in a success—i.e., the result is a negative payoff when the sender and receiver attempt to correct action and fail repeatedly. This sort of extension incorporates varying punishment for failure to coordinate in the same way that the correction game incorporates

a varying (positive) payoff as a function of the number of repeat attempts made. This amounts to varying a parameter of the underlying model; several other such extensions could be made. For example, we might vary the initial payoff, $u(s, a)$, or we might add a punishment parameter for failure, even when the attempt is not repeated, and then vary the punishment to increase with an increase in repetitions. This is in addition to the usual parameters that might be varied—e.g., the dimension of the game, the underlying dynamic itself, including punishment in general, etc. In this case, we used a single, well-studied dimension and the most straightforward learning dynamic for illustrative purposes.

Several questions arise for an analysis of the model that was presented here. For example, we might look at different choice rules for how the receiver chooses her action in the event of a repetition. It was supposed that the receiver reduces the set of possible actions by abandoning previous actions that resulted in a failure to coordinate. This was taken to be the most parsimonious decision for how the sender and receiver play this modified signalling game: given that we assume that the sender and receiver have already coordinated on a 'yes/no' signal, it makes sense that the receiver would 'understand' that she should not retry the action that led to a failure. However, we might relax this assumption by allowing the receiver to randomise over the entire set of possible actions repeatedly—this might be plausible to the extent that individuals might keep trying something incorrect even when they are told it is incorrect.

In this case, there is some nonzero probability that the correction cycle will continue indefinitely, for any $\mu > 0$—especially as μ gets arbitrarily close to 1. Note that the probability on a given round that the sender and receiver end up in a loop of repetitions is 0, in the limit.[21] However, for any particular n, the probability that they miscoordinate n times in a

[21]For this particular case (the 8×8 correction game), assume the probability of repetition is 1; then, for example, at the outset when all dispositions are equiprobable, the probability that the sender and receiver miscoordinate, given a fixed state and signal, is 7/8. Thus, the probability that they miscoordinate n times in a series of n repetitions is $\left(\frac{7}{8}\right)^n$, since each miscoordination is independent of anything that has happened

row has non-zero probability—though the likelihood of this happening decreases quickly as n increases.[22]

Nonetheless, in the case of getting caught in pooling equilibria, the probability that the sender and receiver get caught in such an infinite loop may increase considerably. Thus, in terms of modelling, we might want to have an upper bound on how many times the sender bothers to try to correct the action of the receiver. Note that, given the modelling assumptions that were made in this chapter, there will always be a maximum number of repetitions possible—namely, $(n-1)$ times for the $n \times n$ game.

Finally, we assumed a fixed μ and γ for any particular set of simulations that were run; there may well be an optimal combination of parameters. Another extension would be to see whether or not the sender and receiver can coordinate on such an optimal combination. That is, we might model the game in a way such that the sender *learns* a probability parameter for attempting to correct μ. This will likely be most effective when μ is very low to start—to allow the sender and receiver to begin moving toward a coordination equilibrium—and then gradually increasing as time goes on. It might be the case that when the sender only tries to correct the receiver's behaviour when it is *salient* to do so.[23] For example, it might be more salient to try to correct when the sender has already clearly differentiated a particular signal, though the meanings of the other signals may still be in flux.

While this is all certainly food for thought, the purpose of this chapter was to show, in one particular case, how the composition of games might allow for the more efficient evolution of signalling dispositions. The correction game, as presented, does precisely that.

---

previously. Further,

$$\lim_{n\to\infty} \left(\frac{7}{8}\right)^n = 0.$$

[22]The probability of failing more than 6 times in a row is less than 0.5, and the probability of failing, e.g., 50 times in a row is around than 0.001

[23]See, e.g., LaCroix (2018); Barrett (2019).

Chapter 6

Using Logic to Learn More Logic

> *Logic takes care of itself; all we have to do is to look and see how it does it.*
> — Wittgenstein, *Journal Entry - 13 Oct. 1914*

Evolutionary game theorists often take for granted that individuals can be characterised as being involved in a game in the first place. For the modeller, a game-theoretic model is just a neat and tidy idealisation that is amenable to some analysis—in this sense, such underlying assumptions are not different than, e.g., frictionless planes in physics. However, Barrett and Skyrms (2017) ask how such games might evolve in the first place. In bringing to light natural processes by which games themselves may evolve, they suggest that it is worthwhile to ask how individuals in the world might come to interact in such ways that can be usefully characterised *as* a game. This is the main focus of their notion of *self-assembly*, which was discussed in detail in Chapter 3. Under the description of self-assembly, in such cases, individuals with prior strategies for solving decision problems might interact. These interactions may compose to form games. Once such simple games have arisen, they may themselves compose to form more complex games—ones that are, perhaps, capable of dealing with novel phenomena in ways that are more efficient than learning new dispositions from

scratch. This process, as we have seen, may variously involve template transfer, analogical reasoning, and modular composition; see Chapter 3.

Barrett and Skyrms (2017) note that, while their inquiry is concerned primarily with signalling, this framework applies equally well to other types of social interactions, including, e.g., a public goods game evolving to work in tandem with a bargaining game. Barrett and Skyrms (2017) discuss, specifically, the composition of addition and ordering judgements, as well as efficacy considerations concerning template transfer in the context of binary logical operators—specifically, NAND. In the linguistic context of a signalling game, we might ask how pre-evolved *communicative* dispositions might compose to form more complex communication systems. As we have previously seen, there are several ways in which this might happen—specifically, without relying upon a *linguistic* notion of compositionality, but rather a notion of *modular* composition; see Chapter 3 and Chapter 4. In Chapter 5, we saw how the evolution of a simple 'yes/no' type of command might help signalling systems to evolve, and how successful evolution depends inherently upon the game under consideration.

I consider how complex logical operations might self-assemble in a signalling-game context via composition of simpler underlying dispositions—in particular, with an emphasis on reflexivity. On the one hand, agents may take advantage of pre-evolved dispositions; on the other hand, they may co-evolve dispositions as they simultaneously learn to combine them to display more complex behaviour. In either case, the evolution of complex logical operations can be more efficient than evolving such capacities from scratch. Showing how complex phenomena like these might evolve provides an additional path to the possibility of evolving more or less rich notions of compositionality. This helps provide another facet of the evolutionary story of how sufficiently rich, human-level cognitive or linguistic capacities may arise from simpler precursors.

I will examine the evolution of logical operations in more detail. The case of logical operations provides a nice toy example to play with, and an excellent testbed for comparing results,

in the sense that it is well-structured, unambiguous, and widely applicable. Interpreting the states and acts as the input and output of logical operations allows for a distinct set of parameters under which a logical operation might be said to be successful. Barrett and Skyrms (2017) begin to examine this sort of structure in the context of self-assembling games. This simple logical game is used to give a clear example of the relative efficacy of template transfer in a direct way. Barrett (2019) highlights that this models a situation in which agents *co-evolve* to represent facts about the world and apply logical operations to those facts.

The dynamic that I will consider throughout this chapter is simple reinforcement learning; see Chapter 1. The main reason for this, as has been discussed before, is twofold: on the one hand, reinforcement learning is often considered the simplest dynamic that we can study. Thus, if interesting phenomena can arise under this dynamic, it should only arise more quickly when the agents are supplemented with further computational resources. On the other hand, this is the dynamic upon which Barrett and Skyrms (2017) focus; thus, utilising the same dynamic for each of the games will allow for easy comparison of results across cases.

In their conclusion, Barrett and Skyrms (2017) point out that '[t]he evolution of strategies in a given game is a vibrant area of ongoing research. But the question of the evolution of games themselves is important and deserves to be explored. Here we have taken a few initial steps' (351). The purpose of this chapter is to take their analysis one or two (or perhaps more) steps further and, by doing so, to give a concrete example of the themes that have been explored throughout this dissertation. This analysis is complementary to the direction in which Barrett (2019) goes to consider how complex logical operations might evolve via self-assembly and salience.

6.1 Simple Signalling Games for Unary Logical Functions

I will begin with an excessively simple model. Whereas Barrett and Skyrms (2017) use the binary operator NAND to compare the evolution of new dispositions—taking two values as input (the states) and returning a single value as output (the act)—we will see that we can understand such a binary operation as the modular composition of two unary logical operators. Consider the following four (exhaustive) unary operations on a single input (proposition) in the context of a two-player sender-receiver game: identity (ID), negation (NEG), tautology (TAUT), and contradiction (CONT); see Table 6.1.[1]

p	ID(p)	NEG(p)	TAUT(p)	CONT(p)
1	1	0	1	0
0	0	1	1	0

Table 6.1: Unique outputs for unary functions

Each of these unary functions can be modelled as a signalling game, with the differentiating feature being the composition of the payoff matrix for that game. To be clear in the exposition that follows, I will denote the game that models the ID function as the 'ID game', and similarly for the other unary functions. The payoff matrices for the games that correspond to each of these unary logical operations are given in Table 6.2. Here, I will follow the convention that 's_0' ['a_0'] corresponds to input [output] 0, and 's_1' ['a_1'] corresponds to input [output] 1. Since the payoffs are also denoted by '0' and '1', this is meant to help keep the state, act, and payoff disambiguated. I will refer to the input [output] as either '0' and '1', or 's_0' and 's_1' ['a_0' and 'a_1'] depending upon the context and which notation allows for the most clarity.

[1]Note that each of these *functions* can be interpreted as logical propositions themselves, where ID(p) = p, NEG(p) = ($\neg p$), TAUT(p) = ($p \vee \neg p$), and CONT(p) = ($p \wedge \neg p$). The key interpretation here is that although, e.g., ($p \wedge \neg p$) can be interpreted as a binary operation—AND, in this case—each of these depends only upon one proposition—namely, p—therefore, they can be understood atomically.

	a_0	a_1
s_0	1	0
s_1	0	1

(a) ID Game

	a_0	a_1
s_0	0	1
s_1	1	0

(b) NEG Game

	a_0	a_1
s_0	0	1
s_1	0	1

(c) TAUT Game

	a_0	a_1
s_0	1	0
s_1	1	0

(d) CONT Game

Table 6.2: Four possibilities for payoffs matching states to acts: (a) is a coordination game, which corresponds to the unary identity function; (b) is an anti-coordination game, corresponding to the unary negation function; (c) and (d) are, what we might call, pooling games, corresponding to the tautology and contradiction functions, respectively.

On a given play of the game, nature picks a truth value for the proposition, randomly and without bias. The sender sees the state and chooses a message randomly from the appropriate urn for the current state. The receiver sees the message but not the state; she chooses an action randomly from the urn for the current message. If the action 'matches' the state (as defined by the payoff tables in Table 6.2), then the play is counted as a success. In this case, the players each replace the ball they chose on that round to the urn from which it was chosen and add another ball of the same type to that urn. If the play was not successful, then they return the ball to the urn from which it was chosen. Thus, over time, their propensities to act shift proportional to their past successes.

Given that the ID game just is a 2×2 signalling game, we know that the players will (eventually, but usually quickly) coordinate upon one or the other signalling system with probability 1, when nature is unbiased (Argiento et al., 2009). Indeed, the NEG game is functionally equivalent to the ID game—modulo a permutation of the payoffs, or a re-labelling of the states or acts. Thus, it follows immediately that the NEG game also gives rise to one or the other signalling system with probability 1 (subject to the same caveats). Further, the receiver in the TAUT- and CONT games has an action available to her which is strictly dominant. Therefore, each of these will converge to an optimal strategy under simple reinforcement learning since this dynamic converges on a dominant strategy if there is one (Beggs, 2005).

Accordingly, success is guaranteed eventually in each of these four simple unary logic games under reinforcement learning.

However, analytic results do not capture what happens in the short term, or how quickly these effective strategies might arise. On simulation, the average cumulative success rates (1000 runs each) for the ID- and NEG games are both around 0.70 after 10^2 plays per run. In contrast, by 10^2 plays per run, the average cumulative success rates for the TAUT- and CONT games are already near 0.92. A comparison of the cumulative success rates for each of these games for several thresholds is given in Table 6.3. Figures 6.1 and 6.2 illustrate

	ID game		NEG game		TAUT game		CONT game	
	10^2	10^6	10^2	10^6	10^2	10^6	10^2	10^6
μ	**0.692**	**0.999**	**0.700**	**0.998**	**0.916**	**1.00**	**0.917**	**1.00**
0.99	0.00	0.98	0.00	1.00	0.05	1.00	0.04	1.00
0.95	0.00	1.00	0.00	1.00	0.20	1.00	0.21	1.00
0.90	0.03	1.00	0.03	1.00	0.76	1.00	0.78	1.00
0.80	0.27	1.00	0.29	0.98	1.00	1.00	1.00	1.00

Table 6.3: Comparison of cumulative success rates for unary logic games with a variety of thresholds for success

the difference in speed of convergence between these games graphically. The relative speed of the TAUT- and CONT games will be necessary for explaining the results discussed in Section 6.2 below. Therefore, it is instructive to discuss why the TAUT- and CONT games should evolve faster than the ID- and NEG games.

In the ID- and NEG games, the players must co-evolve their strategies to reach a maximally-effective strategy set—namely, the 'signalling systems' of these sender-receiver games.[2] However, there are more maximally-effective sets of pure strategies available to the players in

[2]Technically, the maximally-effective sets of strategies are only signalling systems in the ID- and NEG games since these are properly signalling games, which require coordination on the part of the agents. In the TAUT- and CONT games, the players need not coordinate to achieve maximal payoff—signals need not carry any information—so it makes little sense to talk of 'signalling systems' or 'coordination conventions' in these latter contexts.

Figure 6.1: Comparison of average cumulative success rates for unary-input logic games in the short term (10^2 plays per run)

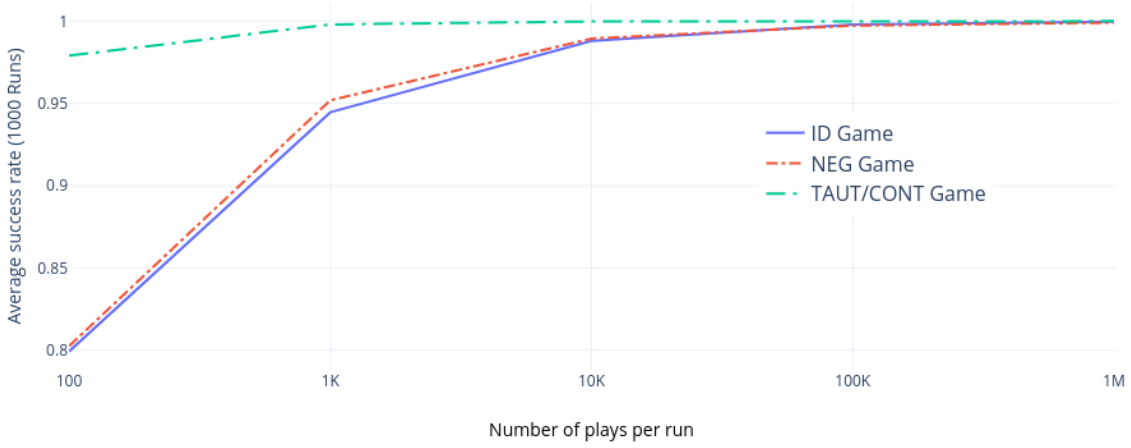


Figure 6.2: Comparison of average cumulative success rates for unary-input logic games in the long term (10^6 plays per run)

the CONT- and TAUT games.[3] Taking account of mixed strategies implies a continuum of

[3]In the atomic n-game, there are n^{2n} strategy combinations, $n!$ of which are signalling systems. Thus, for the ID- and NEG games there are $2^4 = 16$ combinations of pure strategies, and $2! = 2$ of these are signalling systems. However, in the TAUT game and the CONT game, there are 4 maximally-effective sets of (pure) sender-receiver strategies.

maximally-effective strategies for the TAUT- and CONT games: every mixed strategy for the sender in conjunction with the appropriate pure total-pooling strategy for the receiver will result in a maximal payoff in the TAUT- and CONT games, whereas no mixed strategy proffers maximal payoff in the ID- and NEG games.

Thus, it is sufficient, but not necessary, that the players co-evolve their strategies in the TAUT- and CONT games to achieve maximal payoff; the receiver may learn that the signal does not matter—she should always choose action 0 in the CONT game and action 1 in the TAUT game.

Unary functions are simple enough that they are virtually guaranteed to emerge in a sender-receiver context under simple reinforcement learning.[4] With these initial results outlined, I describe how a binary logical operator can be composed out of unary logical operators.

6.2 Composing Unary Functions for Binary Inputs

In this section, I demonstrate how actors in a signalling context might compose unary dispositions into more complex binary dispositions. I begin with the relevant background, against which I will compare the novel models in this chapter. This includes learning to evolve a binary disposition from scratch via 'syntactic' signalling (Barrett, 2006, 2007, 2009) and appropriating a pre-evolved binary disposition for a novel context via 'template transfer' (Barrett and Skyrms, 2017). In Section 6.2.1, I present a novel means for agents to utilise pre-evolved dispositions (the unary logic dispositions discussed in Section 6.1) to learn more complex novel dispositions (a binary-input NAND disposition). This model presupposes that the agents have already learned the unary logical operations and further learn to combine them into binary logical operations. However, the assumption that the unary logical opera-

[4]This guarantee requires that nature is not too biased for the ID- and NEG games (Hofbauer and Huttegger, 2008); no such assumption is required for the TAUT- and CONT games.

tions are pre-evolved is relaxed in Section 6.2.2 using a hierarchical model similar to the one employed in (Barrett et al., 2018). This too involves several simplifying assumptions, which are incrementally relaxed in Sections 6.2.3, 6.2.4, and 6.2.5.

Each of the models I discuss uses only simple reinforcement learning, where the propensities for actions are proportional to the accumulated rewards for prior actions. Nature is unbiased, and players' initial weights are always 1 so that the probability distribution over any player's actions is uniform to start. Finally, the rewards are always 1 for successful actions. There is no discounting, no error rate, no punishment, and no bounds in any of these models. In terms of our urn-learning metaphor, each urn always starts with one ball of each type, and a single ball of the appropriate type is added when agents act successfully.[5]

Since Barrett and Skyrms (2017) examine template transfer for NAND, I will use the same operator to analyse the evolution of binary logical functions via modular composition of the unary operations that were discussed in Section 6.1.[6] There are several ways that we might model the evolution of a NAND game using reinforcement learning. The first, which I will refer to as the 'atomic two-sender NAND game', is shown in Figure 6.3. There are two senders, called 'sender A' and 'sender B'. There are two values for the state of the world, $s_0 = 0$ and $s_1 = 1$, and a full input state is an ordered pair. Thus, there are four input states, $\langle 0, 0\rangle, \langle 0, 1\rangle, \langle 1, 0\rangle$ and $\langle 1, 1\rangle$.

To be clear on notation, I will use s_i to indicate the state-value $i \in \{0, 1\}$ and s_{ij} to denote the full binary state, $\langle i, j\rangle$. I will use s_A to refer to the state seen by sender A and s_B to refer

[5]This is the most straightforward case, but there is good reason to think that the results presented will be robust to variation of these parameters. For example, Barrett et al. (2017a) discuss a low-rationality hybrid of simple reinforcement and the 'win-stay/lose-randomise' learning dynamic and show that it is reliable, stable, and exceptionally fast for learning in signalling contexts. LaCroix (2018) discusses a novel learning rule which helps avoid partial pooling, even in complex games. Similarly, adding punishment or forgetting can help agents evolve optimal signalling conventions (Barrett and Zollman, 2009).

[6]Barrett (2018) also discusses the evolution of NAND in the context of a 'sender-predictor' game. However, his simulations use bounded reinforcement with punishment, which is radically different (and importantly more sophisticated) than the generic, straightforward dynamic I consider. Still, his general remarks regarding the effectiveness of appropriation of logical operations are relevant here—specifically, the empirical case of appropriation in the context of rule-following in pinyon and scrub jays (Bond et al., 2003).

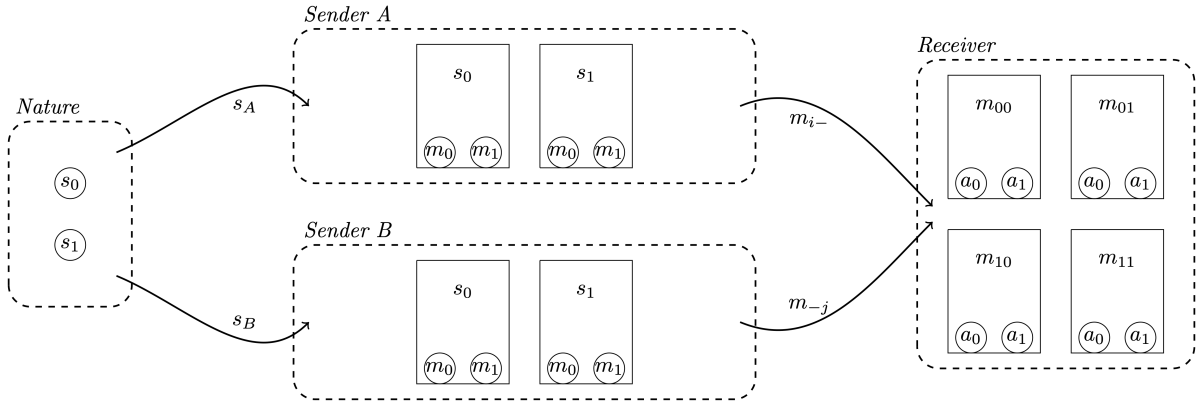


Figure 6.3: Reinforcement learning model for simple binary-input logic game with two senders. Nature chooses twice from one of two states; each sender chooses one of two messages from the urn matching the state received; the receiver chooses one of two actions from the urn matching the two-input signal received. Both the senders and receiver reinforce just in case the act corresponds (s_A NAND s_B)—namely, the state shown to sender A and the state shown to sender B, respectively.

to the state seen by sender B. Nature chooses each state independently, and each sender only sees one aspect of the full state.[7] m_{i-} denotes A's message, and m_{-j} denotes B's message. m_{ij} denotes the full 2-bit message that the receiver observes. Consistent with previous work on two-sender signalling games, I assume that the receiver knows which sender sends which message.[8] Further, the senders' messages are independent in the sense that neither sender knows which message the other sent. Finally, there is no requirement that the two senders be interpreted as distinct agents—they might be understood as functional components of a single organism.

The atomic two-sender NAND game involves learning the appropriate outputs for NAND from scratch. Nature chooses a state-value randomly to send to A and separately chooses another state-value to send to B. The ordered combination of values $\langle s_A, s_B \rangle$ constitutes the full

[7]This game can also be modelled so the senders each see the full state of nature and must coordinate on how they partition nature, to encode complete information about the state. Barrett (2018) highlights that a truth-functional rule is more likely to evolve when the senders each have access to different, independent states of nature than when they have access to the full state. However, such a constraint is not artificial since the receiver must use information from both senders; hence, this allows for a generalisable truth-functional operation. In either case, I discuss role-free senders in Section 6.2.5.

[8]See (Skyrms, 2000a, 2010a; Barrett, 2006, 2007, 2009, 2018, 2019; Barrett and Skyrms, 2017).

state, which can be interpreted as the binary input for the logical function in question. Each sender chooses a message from the urn that matches the state that she sees and sends that message to the receiver. The receiver sees which message each sender sent, and who sent the message, and chooses an output action, $a_0(=0)$ or $a_1(=1)$, from the urn corresponding to the full 2-bit message. In the NAND game, a play is successful just in case the receiver chooses the act corresponding to (s_A NAND s_B)—namely, she should choose a_0 when both state-values are 1, and she should choose a_1 otherwise. The payoff table for the NAND game, with both the state and act labels, s_{ij} and a_k, and the actual input-output values, 0 or 1, is shown in Table 6.4.

		a_0	a_1
		0	**1**
s_{00}	$\langle 0,0\rangle$	0	1
s_{01}	$\langle 0,1\rangle$	0	1
s_{10}	$\langle 1,0\rangle$	0	1
s_{11}	$\langle 1,1\rangle$	1	0

Table 6.4: Payoff table (states and acts) for atomic NAND game

On simulation, after 10^6 plays per run, the cumulative success rate is 0.9053, on average (1000 runs), when the players must learn a NAND disposition from scratch. More often than not (0.54), the players achieve a cumulative success rate higher than 0.95, and about one-quarter of the time (0.26), they achieve a near-perfect cumulative success rate (≥ 0.99). In approximately one-quarter (0.26) of the runs, the agents appear to get caught in a partial-pooling equilibrium. Here, they fail to learn a maximally-efficient signalling convention for a reduced expected payoff of 0.75. In such cases, the receiver always chooses a_1.[9]

[9]On 10^7 plays per run, the agents still get caught in partial pooling at a rate of about 0.25.

Once the agents have learned a NAND disposition, they may appropriate this disposition, via template transfer, for use in a novel context (Barrett and Skyrms, 2017). This process is an order of magnitude more efficient than learning the same disposition in the new context from scratch. A schematic for the template-transfer model is shown in Figure 6.4. We

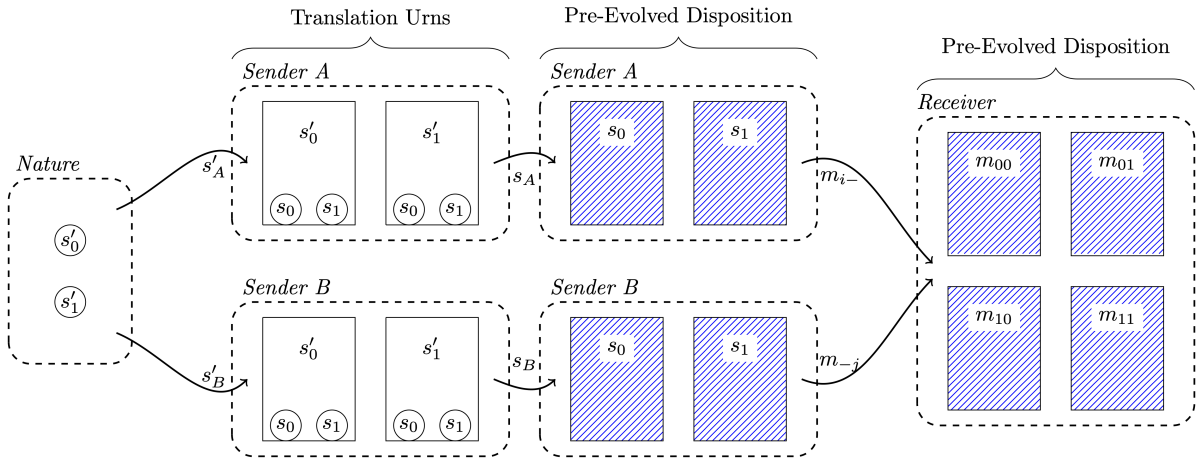


Figure 6.4: Transfer learning model for simple binary-input logic game with two senders. Nature chooses twice from one of two states; each sender chooses one of two messages from the urn matching the state received; the receiver chooses one of two actions from the urn matching the two-input signal received. Both the senders and receiver reinforce just in case the act corresponds to (s_A NAND s_B).

suppose that the agents have already coordinated upon a convention for outputting (s_A NAND s_B) on input $\langle s_A, s_B \rangle$. Thus, the urns for these pre-evolved dispositions are already populated and fixed. Now, given a new context with novel state-values, s'_0 and s'_1, the senders learn to appropriate the previously evolved disposition by translating the novel states into their analogues in the prior context. This model additionally shows how individuals who have already learned NAND can quickly learn a different logical operation, such as OR. See Figure 6.5.

Barrett and Skyrms (2017) report that on 1000 runs with 10^5 plays per run, 0.78 of the runs exhibit a cumulative success rate of better than 0.80, 0.61 of the runs better than 0.90, and

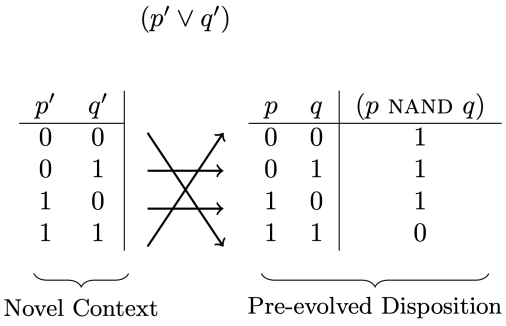


Figure 6.5: Example of translating a novel context (OR) into a pre-evolved disposition (NAND) via template transfer

0.50 of the runs better than 0.95. This is roughly the same level of success that the atomic two-sender NAND game achieves on 10^6 plays per run.

The template-transfer model for NAND is meant to be suggestive. However, this set up has several short-comings—for example, though the players might evolve OR via template transfer on a pre-evolved NAND disposition, it is less clear how (or whether) they might be able to use this pre-evolved NAND disposition to evolve, for example, AND, since this operation has a different number of '0' values. I will discuss these and other considerations in more detail in Section 6.3 below. But first, we will see how modular composition can be applied to the evolution of binary logical operations out of the unary logical operations that were presented in Section 6.1.

6.2.1 Utilising pre-evolved dispositions

In this section, I present one of the key insights of this chapter: the payoff table for the atomic NAND game (Table 6.4) can be understood as the composition of two unary logic games—in this particular case, TAUT and NEG; see Table 6.5. This insight is crucially important for understanding the subsequent models presented in this chapter, as each of these builds off of and relaxes certain assumptions of the basic model shown here. The sense in which I mean

		a_1	a_2	
		0	**1**	
s_{00}	$\langle \mathbf{0}, 0 \rangle$	0	1	} TAUT Game
s_{01}	$\langle \mathbf{0}, 1 \rangle$	0	1	
s_{10}	$\langle \mathbf{1}, 0 \rangle$	0	1	} NEG Game
s_{11}	$\langle \mathbf{1}, 1 \rangle$	1	0	

Table 6.5: Payoff table for NAND game as the composition of two unary games

that the binary logical operation NAND can be understood as the composition of the unary logic operations TAUT and NEG is as simple as this: the truth table for NAND (Table 6.4) *just is* the truth table for unary TAUT stacked on top of the truth table for unary NEG (Table 6.5). This compositional idea is at the core of every model that follows in this section.

This may seem somewhat trivial; however, the order of the inputs plays a particular role in this game: the second input corresponds to the input of the unary sub-game, whereas the first input differentiates the two sub-games that compose the NAND game. Put another way, the first input tells us whether we are in the top half or bottom half of the (binary-NAND) truth table, and the second input codes for which output is appropriate given the context that is differentiated by the first input. Therefore, if sender A codes for the first input, and sender B codes for the second input, then the receiver might learn to interpret the first signal as specifying which unary game should take the state that is encoded by the second signal as input, and then output the appropriate value for that game and input.

If we assume that players have pre-evolved the unary sub-game (so their dispositions are already fixed), then the binary-input NAND game with pre-evolved unary sub-games can be modelled as in Figure 6.6. I will refer to this model as the 'pre-evolved composition' model. There are two senders, as with the atomic NAND game (Figure 6.3). A's message codes for the first input from Nature—namely, the unary logic sub-game that ought to be played. The receiver must learn the meaning of this message. B's message codes for the second input

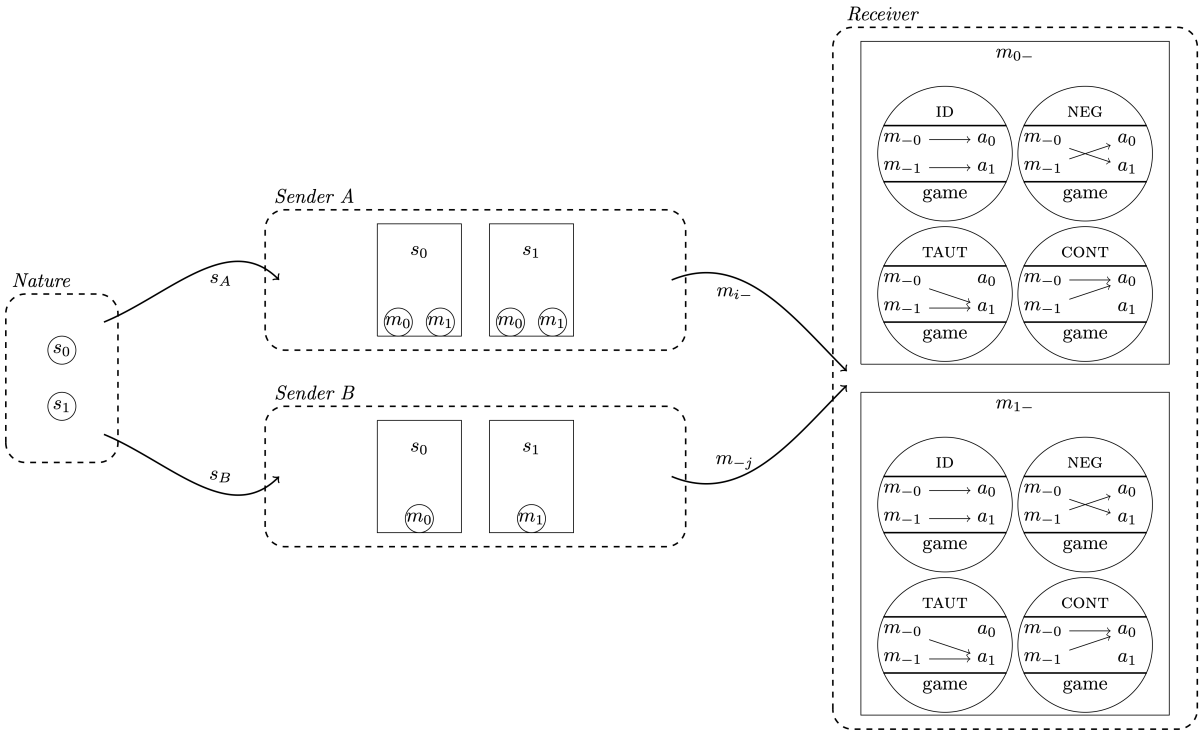


Figure 6.6: Reinforcement learning model for binary-input logic game assuming a pre-evolved unary-input logic game. Nature chooses one of two states for the sender; the sender chooses one of two messages to send to the receiver; the receiver acts according to her pre-evolved disposition for the message received. The solid rectangles (the urns) are determined by the input seen by a particular player. Once an urn is determined, the circles (the balls) show what choices are available to the agent.

from Nature—namely, the *input* of the unary sub-game. We assume that B's dispositions are fixed so that she always sends m_i on input s_i. Thus, the receiver must choose a unary sub-game to make use of B's message; however, once the receiver chooses a unary sub-game, she already knows how to proceed to choose an output based on B's message since the four unary sub-games are pre-evolved.[10]

On a given play of the game, nature chooses a state-value randomly to show to A and separately chooses a state-value to show to B. A chooses a message from the urn matching

[10]This model assumes the senders' roles are fixed—namely, A codes for the *game* and B codes for the *input*. However, this assumption is not necessary. A role-free game with pre-evolved unary sub-game dispositions has almost identical results as the fixed-role game presented here. Hence, I have opted to show the simpler of these models; however, I discuss role-free senders in detail in the co-evolutionary game presented in Section 6.2.5.

the observed state to send to the receiver. B's choices are fixed. The receiver sees each of the messages and chooses, from the urn matching A's message, a unary sub-game to play. Once the sub-game is chosen, everything else is determined: the receiver performs the action corresponding to the message received from B, given the sub-game which she has chosen to play. There are ways in which the players can miscoordinate for a partial payoff in this game. The full payoff table is shown in Table 6.6.[11] As with the atomic binary-input NAND game,

		a_0 ID	a_1 NEG	a_2 TAUT	a_3 CONT
s_{0-}	$\langle \mathbf{0}, - \rangle$	0.5	0.5	1	0
s_{1-}	$\langle \mathbf{1}, - \rangle$	0	1	0.5	0.5

Table 6.6: Payoff table for NAND game as the composition of two unary games

when the players act randomly, the chance payoff is 0.5; if they fail to evolve a signalling system, efficient pooling strategies have a success rate of 0.75.

On simulation, after 10^6 plays per run, the cumulative success rate for the pre-evolved composition game is 0.9440, on average (1000 runs). The players often (0.66) achieve a cumulative success rate higher than 0.95, and about one-third of the time (0.31) achieve a near-perfect cumulative success rate (≥ 0.99). Less than 10% of the time (0.08), they fail to evolve a maximally-efficient signalling convention and get caught in pooling equilibria—in these rare cases, the receiver fails to learn the differentiating feature of the first sender's signal and always chooses the TAUT game as the unary sub-game.

[11] Compare this with Table 6.5. Recalling that the unary logic sub-game dispositions are pre-evolved, and so fixed, consider the following situation. Suppose $s_A = 1$ and the receiver chooses to play the NEG game. Then regardless of what s_B is, the receiver will choose the correct action, affording each player a payoff of 1. Suppose, however, that the receiver chooses the ID game. Then regardless of what s_B is, the receiver will choose the wrong action, affording each player a payoff of 0. Finally, suppose that the receiver chooses to play the TAUT game. Then, if s_B was 0, she outputs 1, which was the correct action (since she should have chosen the NEG game, which outputs 1 on input 0); however, if s_B was 1, then she again outputs 1, which is the incorrect action (since she should have chosen the NEG game, which outputs 0 on input 1). Therefore, the players will average a payoff of 0.5 since the states are uniformly distributed. The same is true if the receiver chooses the CONT game when she should have chosen the NEG game.

For ease of comparison, the results for 10^5 and 10^6 plays per run, for each of the atomic, template-transfer, and pre-evolved NAND games, are displayed in Table 6.7.[12] As with tem-

	Atomic NAND		Template Transfer		Pre-Evolved Composition	
	10^5	10^6	10^5	10^6	10^5	10^6
0.95	0.35	0.54	0.50	—	0.47	0.66
0.90	0.50	0.63	0.61	—	0.64	0.79
0.80	0.66	0.74	0.78	—	0.84	0.92

Table 6.7: Comparison of evolutionary efficacy for learning NAND (a) as a novel disposition, from scratch; (b) via template transfer on a pre-evolved NAND disposition; and (c) via simple reinforcement on pre-evolved unary dispositions

plate transfer, composition that takes advantage of pre-evolved unary dispositions appears to allow the agents to learn the binary disposition an order of magnitude faster than learning it from scratch. When the agents learn to compose pre-evolved unary dispositions, they do comparably well to template transfer on a pre-evolved binary NAND disposition. However, although template transfer and pre-evolved unary composition both take advantage of prior dispositions, the model presented here is more efficient at evolving a NAND disposition than with template transfer, in the following sense.

The atomic NAND game sometimes (0.25 on 10^7 plays per run) fails to evolve NAND, but pools strategies for an expected payoff of 0.75. Barrett and Skyrms (2017) report that their template-transfer game fails to evolve NAND at about the same rate (0.23 on 10^7 plays per run). Therefore, when the players do learn the disposition, they learn more quickly with template transfer than atomically. However, they fail to learn as often with template transfer as they do atomically. In contrast, utilising pre-evolved unary sub-games to learn a composed NAND disposition fails less often than either template transfer or atomic NAND—only 0.08 of the runs fail to evolve NAND on 10^6 plays per run, and only 0.16 of the runs fail to evolve NAND on 10^5 plays per run. Therefore, the agents in a pre-evolved NAND game learn

[12]The results for the template-transfer game are as reported in (Barrett and Skyrms, 2017); they do not report results for 10^6 plays.

as quickly as in template transfer, but they also learn significantly more often. Composing pre-evolved dispositions is both as efficient and more effective.[13]

Since the discussion of template transfer in (Barrett and Skyrms, 2017) presupposes that the underlying disposition is already evolved, the assumptions in my pre-evolved composition model seem relevantly justified—at least for comparing these results.[14] However, by cashing out binary operators in terms of the composition of (pre-evolved) unary operators, this model can do slightly more.

I mentioned previously that Barrett and Skyrms (2017); Barrett (2019) suggest that an OR disposition can easily be transferred from a NAND disposition by dint of the parity of their truth tables—both OR and NAND have three inputs yielding '1' and one input yielding '0'. However, it is not so obvious how this translation can be generalised: NAND cannot be transferred to learn an XOR disposition effectively since its truth values do not exhibit this sort of parity with NAND—XOR has two inputs yielding '1' and two inputs yielding '0'. Thus, understanding binary operations in terms of the composition of unary operations has at least this theoretical virtue over and above simple template transfer.

We have seen how NAND can be composed of unary TAUT and NEG operations; similarly, OR can be composed by unary ID and TAUT operations, XOR can be composed by unary ID and NEG, and so on. Since there are 4 distinct unary operators, and a binary operator consists of some particular permutation of these (with replacement, so, for example, we can account for a 2-input tautology or contradiction), we have $2^4 = 16$ unique permutations of unary operations, which correspond precisely to the 16 unique binary logical operators.

[13]There is a subtle point to be made clear here: recall that the cumulative success rate gives a measure of all of the plays over the course of all of the runs. If agents are slow to learn, then early failures may not be truly washed out. However, the communicative success rate is not history-dependent in the same way. In this case, the pre-evolved NAND game has a failure rate between 0.02 and 0.05 by 10^6 plays per run, as compared with a failure rate between 0.15 and 0.25 in the atomic case. This implies that some of the runs that count as failures in the pre-evolved game are just slow to learn. However, as was noted above, increasing the number of runs in the atomic case does not change the failure rate—the runs that are counted as failures appear genuinely to be caught in partial-pooling equilibria.

[14]Note that the roles of the senders are also fixed in the template-transfer model.

Furthermore, we know from the results of Section 6.1 that all of the unary operations are a 'sure thing' when nature is not too biased, and TAUT and CONT are a sure thing regardless of whether nature is biased. Therefore, it is not unreasonable to assume that these simple dispositions come pre-evolved. In such a case, the agents can learn a complex disposition by merely learning how to code for which unary disposition is appropriate in which context. Though the assumption that the unary sub-game comes pre-evolved is perhaps theoretically justified, we would like a more general picture that does not necessarily presuppose such favourable circumstances are already in place. In Section 6.2.2, this simple model is extended to account for dispositions that co-evolve.

6.2.2 Co-evolving logical dispositions

The co-evolutionary logic game is a variant of the special composition game presented in (Barrett et al., 2018); it is a cooperative game with two base senders (whom we will call A and B) and one base receiver. The agents in this game must evolve a particular sort of signalling system to be uniformly successful. Additionally, there are two 'hierarchical' agents—an 'executive' sender and an 'executive' receiver—who can learn to influence the behaviour of the base agents. See Figure 6.7.[15]

A complete specification of the state is given by two *properties* (or *features*) and a *context*. The properties are *game* (the unary sub-game) and *input*. Each of the properties has two *values*. Considering NAND as a concrete example, the value of the *game* property can be TAUT or NEG, and the value of the *input* property can be 0 or 1. Therefore, the state on a particular play of the game will be either $\langle$TAUT, 0$\rangle$, $\langle$TAUT, 1$\rangle$, $\langle$NEG, 0$\rangle$, or $\langle$NEG, 1$\rangle$. The context indicates which aspect of the state—that is, which of the two properties—needs to

[15]Note that I shift the notation for representing the first index from nature as '0' to explicitly representing it as the game. It should be clear that 's_{00}' '$\langle 0, 0\rangle$', and '$\langle$TAUT, 0$\rangle$' are different representations for the exact same thing. See Table 5.

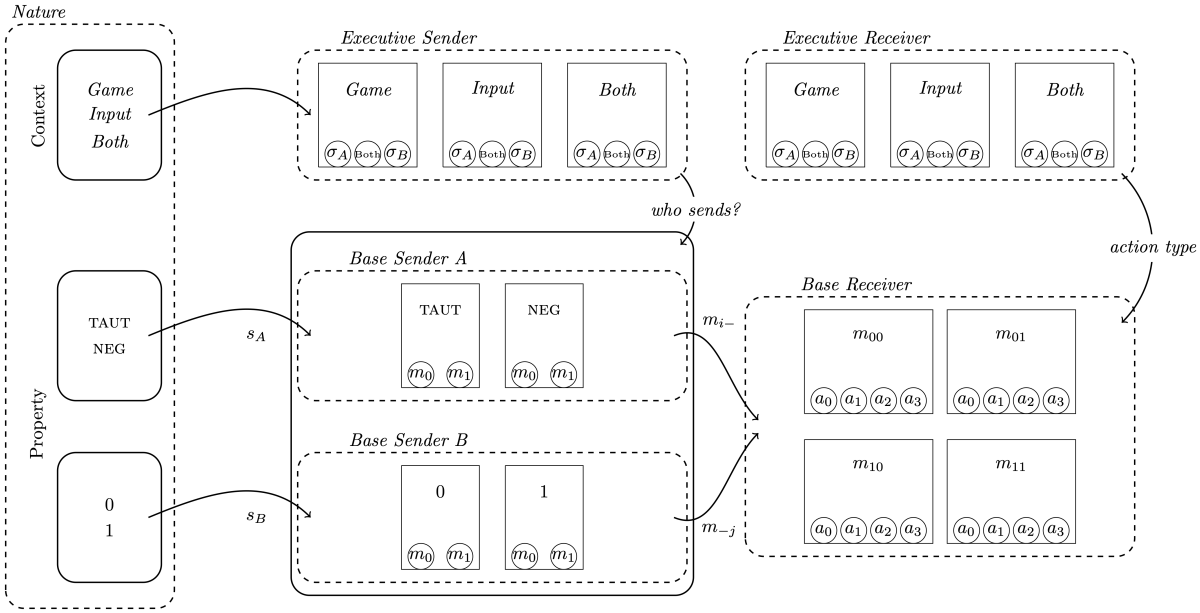


Figure 6.7: Reinforcement learning model for hierarchical, co-evolutionary binary-input logic game.

be known by the receiver for her to perform a successful action on that particular play of the game. Thus, the values for the context are *game*, *input*, or *both*.

The game is played by two base senders, an executive sender, one base receiver, and an executive receiver. Each base sender is assigned a particular property and only has access to that aspect of nature. (This condition is dropped in Section 6.2.5, where I discuss role-free senders.) As such, one base sender sees the *game*, and the other base sender sees the *input*. Initially, the executive sender randomly determines whether the game sender (A), the input sender (B), or both will send a signal. Over time, the executive sender may learn what type of signal the current context demands—namely, a unary logical operation or a binary logical operation.

The base receiver sees the signals sent by the base senders and knows which sender sent each signal. The executive receiver also sees who sends the signals, and she determines whether the base receiver will interpret the signal as a 1-bit message—*game* or *input*—or a 2-bit

message—*both*. The receiver performs an action based upon her interpretation. The actions for the receiver are represented by ⟨*game*, *input*⟩-pairs, matching the states of nature. We assume, for now, that the receiver 'just knows' what the output is for this representation. (This condition is dropped in Section 6.2.3.)

This can be interpreted as follows: the receiver understands the appropriate output—0 or 1—for a given state; however, she does not have access to the current state. Thus, the players must co-evolve a communication system whereby the receiver gains knowledge about the current state. Once the receiver knows the current state, she automatically knows the correct output. Thus, the actions are represented by ⟨TAUT, 0⟩, ⟨TAUT, 1⟩, ⟨NEG, 0⟩, and ⟨NEG, 1⟩. Note that, on this model, the receiver does not have access to representations for CONT or ID. (This condition is dropped in Section 6.2.4.)

The agents are successful on a particular play of the game just in case (1) the base receiver performs the correct action given the current context and (2) the base senders only sent the signals required for success given the context. Thus, to be successful, the receiver's action must be appropriate for the state, and the senders must be as efficient as possible—that is, they must not send any irrelevant signals. The 'efficiency' condition requires the base senders to coordinate, additionally, on something like a pragmatic maxim of relation—namely, the signal sent must be relevant to the current type of context (Grice, 1975). This condition is what drives the co-evolution of the unary dispositions and their composition.[16] See Figure 6.8 for an example.

[16]For example, when the value of the context is *game*, the input is irrelevant, so the sender need only pick an action having to do with the game-value of the state on that round; and, *mutatis mutandis* when the value of the context is *input*. It may seem counterintuitive to 'successfully' play a game without knowing the appropriate input, or to 'successfully' choose an output without knowing the game. However, recall that 'game' and 'input' are peculiar to the structure of the logic game being discussed. The game itself, as was mentioned in the introduction, is only meant to serve as a relatively clear testbed for the types of compositional processes of interest here. Nonetheless, I discuss real-world interpretations in Section 6.5. An alternative way of enforcing efficiency is to posit a cost for signals. However, to maintain consistency with the parameters of the other models presented here—that is, no punishment—a more stringent condition is placed on what counts as a success.

Nature	Base Senders	Base Receiver	
Context	**Who Sends?**	**State**	**Action**
Both	*Sender A* *Sender B*	$\langle \mathbf{neg}, \mathbf{0} \rangle$	$\longrightarrow \langle \mathbf{neg}, \mathbf{0} \rangle$
Game	*Sender A Only*	$\langle \mathbf{neg}, 0 \rangle$	$\longrightarrow \begin{cases} \langle \mathbf{neg}, 0 \rangle \\ \langle \mathbf{neg}, 1 \rangle \end{cases}$
Input	*Sender B Only*	$\langle \textsc{neg}, \mathbf{0} \rangle$	$\longrightarrow \begin{cases} \langle \textsc{neg}, \mathbf{0} \rangle \\ \langle \textsc{taut}, \mathbf{0} \rangle \end{cases}$

Figure 6.8: Example of success conditions for the three types of context. Each row specifies what is required for success in that row's context. Not pictured here are the roles of the executive players; however, they play a part in the success conditions, insofar as the executive sender determines which sender sends a signal, and the executive receiver determines how the sender interprets the signal.

Again, agents learn via simple reinforcement. On each play, nature determines a state by choosing a value for each of the two properties and the context randomly and with uniform probabilities. The executive sender is equipped with an urn for each of the three context-values—*game*, *input*, and *both*. Each urn begins with one ball of each type: *Sender A*, *Sender B*, and *Both*. The executive sender observes the context and randomly draws a ball from the corresponding urn. The drawn ball determines who will send a signal.

A is equipped with an urn labelled TAUT, and an urn labelled NEG; each initially contains a ball labelled m_0, and a ball labelled m_1. If the executive sender draws a ball requiring A to send a signal, then A randomly draws a ball from the urn corresponding to the property she observes, and she sends the corresponding signal. Similarly, B is equipped with an urn labelled 0, and an urn labelled 1—each initially containing a ball labelled m_0, and a ball labelled m_1. If required by the executive sender, she draws a ball from the urn corresponding to the property that she sees and sends that signal to the receiver.

The receiver has four urns, one for every ordered pair of signals she might receive from A and B, respectively: m_{00}, m_{01}, m_{10}, and m_{11}. Each urn begins with one ball for each of the game-input pairs: $\langle$TAUT, 0$\rangle$, $\langle$TAUT, 1$\rangle$, $\langle$NEG, 0$\rangle$, or $\langle$NEG, 1$\rangle$. If both senders send a signal, then the receiver draws a ball randomly from the corresponding urn. If only one sender sends a signal, then the receiver randomly chooses, with unbiased probabilities, one of the two urns corresponding to the sender's signal then draws a ball randomly from that urn.

The executive receiver determines how the receiver will interpret the type of signal she received. This interpretation, in conjunction with the ball the receiver drew, determines how the receiver will act. The executive receiver is equipped with a *game-sender* urn, an *input-sender* urn, and a *both* urn. Each of these initially contains a *game* ball, an *input* ball, and a *both* ball. The ball drawn by the executive receiver determines what type of act the receiver takes as salient given the signal(s) that she has received.

If a play of the game is successful, as per the conditions described above, then each agent who was involved in that particular play returns the ball she drew to the urn from which she drew it and adds another ball of the same type to that urn. Otherwise, each agent simply returns the ball she drew to the urn from which she drew it.

On simulation, the agents nearly always evolve a successful and optimally efficient communication system. After 10^6 plays per run, the cumulative success rate is 0.9716, on average (1000 runs). The players usually (0.88) achieve a cumulative success rate higher than 0.95, and most of the time (0.63) they achieve a near-perfect cumulative success rate ($\geq$ 0.99). Rarely (0.04), they fail to evolve a maximally-efficient signalling convention and get caught in pooling equilibria.[17]

[17]Again, if we examine the 'snapshot' measure of the communicative success rate, the results are slightly better. The average expected payoff after 10^6 plays per run is 0.9747. More than three-quarters of the time, the agents achieve a near-perfect communication convention for a payoff greater than 0.99. Still, 0.04 runs fail to exceed a payoff that could be got by a partial-pooling convention.

Comparing the cumulative success rates, we see that the co-evolutionary NAND game evolves an order of magnitude faster than the pre-evolved NAND game, which in turn evolved an order of magnitude faster than the atomic NAND game on the same learning dynamic. This is despite the fact that the chance payoff for the co-evolved NAND game is less than the chance payoff for the atomic or pre-evolved games. (Since the receiver has twice as many options, the chance payoff is 0.25, rather than 0.50.) Thus, this game starts with a significant handicap, but still outperforms learning NAND from scratch by at least an order of magnitude. See Table 6.8. Part of the reason for this is the full set of conditions for what success consists in.

	Atomic NAND		Template Transfer		Pre-Evolved Composition		Co-Evolved Composition	
	10^5	10^6	10^5	10^6	10^5	10^6	10^5	10^6
0.95	0.35	0.54	0.50	—	0.47	0.66	0.64	0.88
0.90	0.50	0.63	0.61	—	0.64	0.79	0.79	0.92
0.80	0.66	0.74	0.78	—	0.84	0.92	0.89	0.96

Table 6.8: Comparison of evolutionary efficacy for learning NAND (a) as a novel disposition, from scratch; (b) via template transfer on a pre-evolved NAND disposition; (c) via simple reinforcement on pre-evolved unary dispositions; and (d) via co-evolved unary dispositions

The payoffs structure the dispositions of the agents so that they cannot be successful in any way unless they are successful in every way. This is discussed in more detail in Section 6.3.

There are a significant number of simplifying assumptions made in this model, which one may worry are allowing for the high rates of success that we see on simulation. First, I assumed that the receiver 'just knows' how to interpret an action, such as $\langle$TAUT, 1$\rangle$—namely, by outputting 1. Second, I assumed that the only possible states for the game were TAUT and NEG. This is a simplification, given that there are two additional unary operations which the agents may well learn in a general framework, but which happen not to be useful for producing the appropriate action in the NAND context. Finally, I assumed that each base sender is assigned a particular property—*game* or *input*—and only has access to that aspect of nature. Thus, the roles of the senders are fixed.

I relax the first assumption in Section 6.2.3, where the receiver must also learn which output, 0 or 1, is appropriate for which state. The second assumption is relaxed in Section 6.2.4, where I extend the composition game to account for the full action space of unary operations. Finally, in Section 6.2.5, I drop the assumption that the roles of the base senders are fixed.[18] These results are discussed in more detail in Section 4.5.

6.2.3 Learning appropriate outputs

In this section, I drop the condition that the receiver 'just knows' what to do with the 'action' she has chosen from her urn. This game is modelled precisely as the co-evolutionary logic game, except for the following modification. Instead of 4 balls with the state labels, each of the receiver's urns has 8 balls with the state-labels plus an output—0 or 1. Thus, each ball has a three-part label corresponding to the *game* component of the state, the *input* component of the state, and the *output* component of the state—thus, balls on this interpretation are labelled $\langle$*game*, *input*, *output*$\rangle$.

Now, it is not presupposed that the receiver 'just knows' what action she ought to perform when she draws the ball $\langle$TAUT, 0$\rangle$ since she has balls labelled $\langle$TAUT, $0, 0\rangle$ and $\langle$TAUT, $0, 1\rangle$. Thus, she must learn which output is correct, given the complex state.

On this model, the players are successful in coordinating their actions just in case (1) the receiver performs the correct action for the given context, and (2) the senders only sent the signals required for success given the context, and (3) the receiver chooses the correct *output* given the action selected. The rest of the game is as was described before.

[18]Each of these assumptions is dropped independently of the others. This is meant only to be suggestive concerning the effects of these individual assumptions on the simulation results. It would be ideal, though due to space constraints impractical, to look at dropping combinations of assumptions to see whether there are interaction effects between these several parameters.

On the co-evolutionary logic game with learned inputs, the agents still effectively always learn a successful and optimally efficient communication system. After 10^6 plays per run, the cumulative success rate is 0.9313, on average (1000 runs). More than half the time (0.57), the senders and receiver achieve a cumulative success rate higher than 0.95, though they rarely (0.03) achieve a near-perfect cumulative success rate ($\geq$ 0.99). In very few cases (0.07), they fail to evolve a maximally-efficient signalling convention and get caught in pooling equilibria. These results are comparable to the basic co-evolutionary NAND game, where it is assumed that the receiver knows what to do given the state, even though the players start at a significant disadvantage—the chance payoff at the outset is half that of the basic co-evolved NAND game and one-quarter that of the pre-evolved and atomic NAND games.

6.2.4 Taking account of the full state-space of unary games

In this section, I drop the assumption that the only games available to the receiver are TAUT and NEG. Though the ID and CONT games are not appropriate for the NAND game, one might argue that the receiver must learn that these actions are inappropriate. This condition is dropped by extending the basic co-evolutionary NAND game (6.2.2) to a more general logic game. Now, there are 8 actions that the receiver might choose, corresponding to the eight combinations of TAUT, CONT, ID and NEG with the inputs 0 and 1.[19]

On the co-evolutionary logic game where the receiver needs to differentiate the appropriate action from the full state space of unary sub-games, the results are similar to the case where the agent needs to learn the correct output for a given state—it is slightly less efficient and slower to learn initially. Both of these facts make sense since the models are similar in complexity, but, in this case, there are fewer situations that constitute a success. After 10^6

[19]Formally, this model is similar to the model of Section 6.2.3, where the receiver must additionally learn the appropriate output for a given state. However, the success conditions are different when the context is *game* only; thus, these models are not functionally equivalent.

plays per run, the cumulative success rate is 0.9458, on average (1000 runs). More than three-quarters of the time (0.78), the agents achieve a cumulative success rate higher than 0.95, and they often (0.43) achieve a near-perfect cumulative success rate (≥ 0.99). In some cases (0.10), they fail to evolve a maximally-efficient signalling convention and appear to get caught in pooling equilibria.

6.2.5 Role-free composition

The final assumption that I examine is that the roles of the base senders are fixed. As was mentioned previously, assuming that the roles of the base senders are pre-assigned imposes a fair amount of structure on the hierarchical co-evolutionary NAND model. For one, this guarantees that the executive agents always learn to coordinate. On the other hand, a truth-functional rule is more likely to evolve when the senders each have access to different, independent states of nature than when they have access to the full state of nature.[20]

On the role-free co-evolutionary logic game, the base senders have no pre-assigned representational roles. Instead, they are both shown the full state of nature. In this case, the base senders must learn to coordinate their roles to partition nature fully, and the executive agents must learn what roles the base senders are playing. The executive agents thus co-evolve their dispositions even while the base agents are learning their representational roles.

Since there are no stipulated roles for the base senders, there is no stipulated game sender or input sender. On each play of the game, the senders are both shown one of the four states (but not the context-value). These, again, are $\langle$TAUT, 0$\rangle$, $\langle$TAUT, 1$\rangle$, $\langle$NEG, 0$\rangle$, and $\langle$NEG, 1$\rangle$. Each sender has an urn for each of these states, and each urn contains balls labelled m_0 and m_1. Each sender still has only two available messages, and so neither sender can convey full information about the state of the world. Thus, to be successful, they must coordinate so that

[20]See discussion in (Barrett, 2018).

they partition nature fully—that is, their signals ought to give complementary information about the full state of nature. So, one sender ought to learn to code for the *game*, and the other sender ought to learn to code for the *input*.

Since there are no pre-assigned roles in the role-free composition game, the conditions for success are also slightly different. A play now counts as a success just in case (1) the receiver performs the correct action given the current context (as before), and (2) Both senders send a signal if and only if the context given by nature requires both *game* and *input*.

After 10^6 plays per run, the cumulative success rate is 0.9450, on average (1000 runs). About two-thirds of the time (0.67), the agents achieve a cumulative success rate greater than 0.95, and they sometimes (0.21) achieve a near-perfect cumulative success rate (≥ 0.99). Rarely (0.05), they fail to evolve a maximally-efficient signalling convention and appear to get caught in pooling equilibria.

6.3 Discussion

6.3.1 Efficacy and efficiency of learning complex dispositions

Several subtleties should be noted about the results discussed in the previous sections. Notably, a distinction can be made between how effective the agents are at learning a signalling disposition and how efficient they are at learning that disposition. Efficacy is highlighted by the long-run results—particularly by the avoidance of partial-pooling equilibria. In this case, as we have already seen, composing simple dispositions is always more effective than learning a complex disposition from scratch to achieve a maximally-efficient signalling strategy in the NAND game. See Table 6.9.

	Atomic NAND		Template Transfer		Pre-Evolved Composition		Co-Evolved Composition	
	10^5	10^6	10^5	10^6	10^5	10^6	10^5	10^6
0.95	0.35	0.54	0.50	—	0.47	0.66	0.64	0.88
0.90	0.50	0.63	0.61	—	0.64	0.79	0.79	0.92
0.80	0.66	0.74	0.78	—	0.84	0.92	0.89	0.96

	Learned Outputs		Full State-Space		Role-Free Composition	
	10^5	10^6	10^5	10^6	10^5	10^6
0.95	0.43	0.78	0.10	0.57	0.31	0.67
0.90	0.65	0.84	0.41	0.82	0.60	0.85
0.80	0.79	0.90	0.77	0.93	0.87	0.95

Table 6.9: Comparison of evolutionary efficacy for learning NAND (a) as a novel disposition, from scratch; (b) via template transfer on a pre-evolved NAND disposition; (c) via simple reinforcement on pre-evolved unary dispositions; and (d) via co-evolved unary dispositions. These base models are compared with (e) learning the appropriate outputs, (f) learning from the full state-space, and (g) learning with role-free agents

This comparison is made clear in terms of the communicative success rate (average expected payoff) over the long-term (10^6 plays per run) course of these runs in Figure 6.9. The

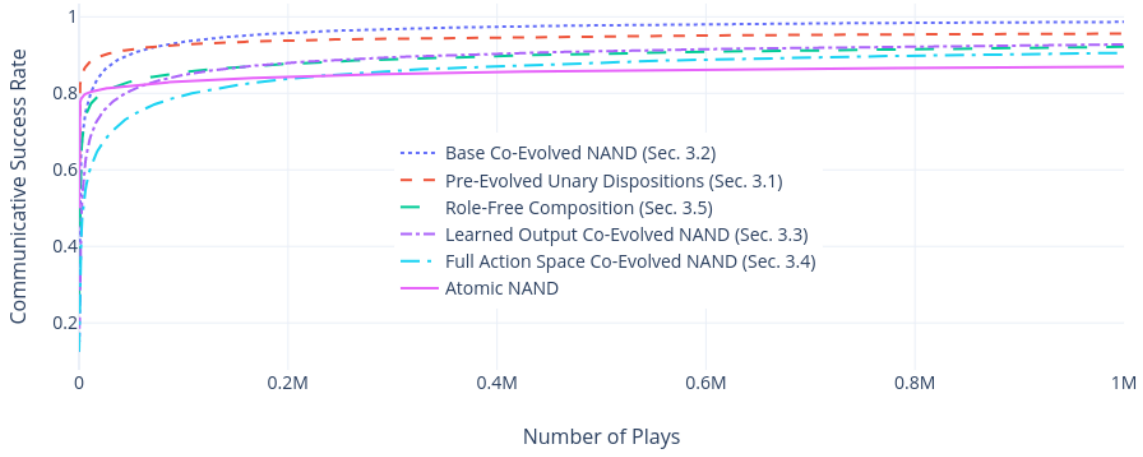


Figure 6.9: Comparison of average communicative success rates over 10^6 plays per run (efficacy considerations)

co-evolved disposition (6.2.2) is the most effective, followed by the model that utilises a pre-

evolved unary sub-game disposition (6.2.1). The extended models of Sections 6.2.3, 6.2.4, and 6.2.5 perform slightly worse in the long run than the base models; however, they all outperform the atomic case.

The reason for this has to do with the efficacy of the learning rule in the following sense. When the agents do learn a maximally-efficient signalling disposition in the atomic NAND game, they generally do as well as in any other game. However, they fail to evolve such a maximally-efficient communication system more often than in the other cases, which brings the average down. A signalling system obtains a maximal payoff of 1 in the limit; however, if 0.25 of the runs get caught in partial-pooling equilibria, then the average payoff will rise no higher than 0.937 in the limit. Thus, these averages tell us something about how effective each model is at avoiding partial-pooling equilibria.

Upon reflection, this ordering makes some sense. The co-evolutionary NAND game takes advantage of a more complex, hierarchical structure than the flat pre-evolved and atomic NAND game models. The base co-evolutionary model contains fewer situations that constitute success, which in turn enhances the ability of the agents to avoid pooling. Furthermore, we can analyse the game in terms of its structural components: there are situations in which only the sub-game dispositions are relevant. Since we saw in Section 6.1 that TAUT and CONT are learned more quickly than ID and NEG, this comes to bear in a significant way on the co-evolution of complex dispositions. That is, the complex disposition can be broken up into structural components which themselves vary in how difficult they are to learn. The extensions of the co-evolutionary model do slightly worse because the receiver has more choice points available to her; however, these complex games still perform better than the atomic case because they still have components which, in some rounds, are decoupled from the more complex game itself. In the atomic case, no such decomposition is available to the agents.

These long-run results say nothing about how quickly the agents learn such dispositions. To get a sense of how efficient learning is, we can compare the short-run results of each of these models. Although the atomic NAND game is the least effective of these models, it is also one of the most efficient. However, because it is not as effective as the other models, it is eventually surpassed in every case. This comparison is made clear in terms of the average expected payoff over the short term (10^4 plays) in Figure 6.10. Note further that the

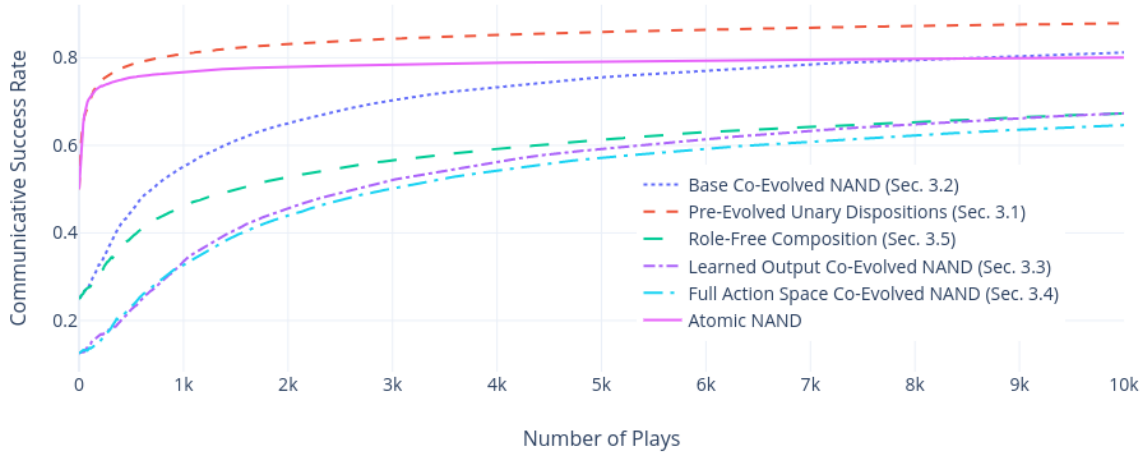


Figure 6.10: Comparison of average communicative success rates over 10^4 plays per run (efficiency considerations)

co-evolutionary games, due to their complexity, start at a significant disadvantage from the atomic game—the chance payoff is significantly lower in each of these cases. Thus, learning is slow and steady in each of the co-evolved instances, but it is also extremely effective.

A final note about the efficacy of composing unary functions compared to template transfer. Barrett and Skyrms (2017) suppose that a NAND disposition has already been evolved. So, evolving a NAND disposition, and learning to apply that disposition to a novel context, are modelled as independent processes. However, having learned a NAND disposition is a necessary condition for there to be a template to transfer to a novel context. As such, these should be understood as serial processes.

The simulation results for learning an atomic NAND game from scratch suggest that approximately 0.25 of the runs result in pooling equilibria. Suppose we have a population of pairs of senders and receivers. Then, around 0.75 of these pairs will learn a NAND disposition, on average. Those who failed to learn NAND would not be able to transfer these dispositions successfully. We also know from the simulation results that learning a NAND disposition is not sufficient for successful transfer in a novel context: approximately 0.25 of the template transfer runs also result in pooling equilibria, failing to learn NAND in the new context successfully. Thus, when considering the serial evolution of the complex disposition, slightly more than half of the population might be successful.

In contrast, we know that the unary NEG disposition is a sure thing. Therefore, every pair of individuals will learn this disposition. Further, TAUT is also a sure thing, so that every pair of individuals will also learn this disposition. Finally, based on the simulation results of Section 6.2.1, success in coordination is high so that approximately 0.92 of those individuals who learned the pre-evolved unary dispositions will also learn to combine them appropriately into a NAND disposition. Thus, almost everyone in the population will learn this disposition when we consider the evolution of these dispositions as a combined serial process.

6.3.2 Other binary operations

There is a subtle distinction that comes to light when we understand binary operators in terms of unary operators. We saw in Section 6.1 that TAUT and CONT evolve more quickly than ID and NEG. So, for example, the composition of TAUT and CONT together should evolve faster than ID and NEG since each of these dispositions is, independently, easier to learn in the former case than in the latter. I discussed the NAND game since this is the logical operation that Barrett and Skyrms (2017); Barrett (2019) discuss. However, the NAND game—which is composed of TAUT and NEG—should then evolve faster than, for example, an IFF game,

which is composed of ID and NEG. We should also expect a binary operation that is composed of two completely pooling unary operations—for example, a binary TAUT operator, to evolve more readily.

Simulation results suggest that this intuition is correct. A comparison of the cumulative success rates for IFF, NAND, and TAUT—being completely representative of the combinations of completely-separating, half-pooling, and totally-pooling sub-games—are shown in Table 6.10. This shows a clear ordering concerning how easy it is to (quickly) learn a

	Atomic IFF		Atomic NAND		Atomic TAUT	
	10^4	10^6	10^4	10^6	10^4	10^6
Average	**0.7502**	**0.8997**	**0.8326**	**0.9053**	**0.9968**	**0.9999**
0.99	0.02	0.56	0.00	0.26	1.00	1.00
0.95	0.25	0.72	0.14	0.54	1.00	1.00
0.90	0.35	0.76	0.29	0.63	1.00	1.00
0.80	0.50	0.80	0.53	0.74	1.00	1.00

Table 6.10: Comparison of three different ways of combining unary operations (cumulative success rate) over the short term (10^4 plays per run) and long term (10^6 plays per run)

binary disposition as a function of how easy it is to learn the constituent unary dispositions which underlie it.

The binary tautology (which takes two inputs and always outputs 1) is easiest to evolve because, in essence, the signal does not matter: the receiver needs only to react to any signal with the same disposition—a_1. This is quickly learned even in the atomic case. In the case of atomic IFF—a binary operation which has no pooling and thus requires co-evolution of strategies—the results are as expected. Specifically, NAND is more likely to get caught in partial-pooling equilibria than IFF and IFF is more efficient than NAND overall; however, because of the difficulty in co-evolving strategies, IFF is more difficult for agents to learn—the variance in the payoffs is significantly larger for IFF (1.74×10^{-1}) than it is for NAND (9.95×10^{-2}). Thus, Barrett and Skyrms (2017) do not look at the simplest case when they discuss template transfer, but they also do not look at the most challenging case.

6.3.3 To infinity, and beyond

How well do these results generalise? There are two different ways that agents can learn a ternary-input logical operation: they might learn to compose two binary-input logical operations appropriately, or they might learn to compose four unary input logical operations. These two possibilities are illustrated in Table 6.11 for evolving NAND.

		a_1	a_2	
		0	**1**	
s_{000}	$\langle \mathbf{0}, \mathbf{0}, 0 \rangle$	0	1	TAUT
s_{001}	$\langle \mathbf{0}, \mathbf{0}, 1 \rangle$	0	1	
s_{010}	$\langle \mathbf{0}, \mathbf{1}, 0 \rangle$	0	1	TAUT
s_{011}	$\langle \mathbf{0}, \mathbf{1}, 1 \rangle$	0	1	
s_{100}	$\langle \mathbf{1}, \mathbf{0}, 0 \rangle$	0	1	TAUT
s_{101}	$\langle \mathbf{1}, \mathbf{0}, 1 \rangle$	0	1	
s_{110}	$\langle \mathbf{1}, \mathbf{1}, 0 \rangle$	0	1	NEG
s_{111}	$\langle \mathbf{1}, \mathbf{1}, 1 \rangle$	1	0	

(a) Ternary logical operator as the composition of two binary logical operators

		a_1	a_2	
		0	**1**	
s_{000}	$\langle \mathbf{0}, 0, 0 \rangle$	0	1	2-TAUT
s_{001}	$\langle \mathbf{0}, 0, 1 \rangle$	0	1	
s_{010}	$\langle \mathbf{0}, 1, 0 \rangle$	0	1	
s_{011}	$\langle \mathbf{0}, 1, 1 \rangle$	0	1	
s_{100}	$\langle \mathbf{1}, 0, 0 \rangle$	0	1	2-NAND
s_{101}	$\langle \mathbf{1}, 0, 1 \rangle$	0	1	
s_{110}	$\langle \mathbf{1}, 1, 0 \rangle$	0	1	
s_{111}	$\langle \mathbf{1}, 1, 1 \rangle$	1	0	

(b) Ternary logical operator as the composition of two binary logical operators

Table 6.11: Two different ways of composing a ternary NAND operation. (a) shows the composition of four unary operations, whereas (b) shows the composition of two binary operations.

In the unary case, the last index of the ternary state provides the unary input for the unary sub-game. The first index of the ternary state distinguishes the top two possibilities from the bottom two, and the second index distinguishes the top unary sub-game from the bottom one (for each partition). In the binary case, there are 16 unique binary operations that we might take account of, assuming the order of the outputs matters. When the sender and receiver play a game which takes advantage of an underlying binary predisposition, we have a signalling game with two states, two signals, and 16 actions.

As with the binary-input NAND game, the players might co-evolve their strategies. The co-evolved ternary-input NAND game may also be modelled in several different ways, depending on whether or not the underlying co-evolved game is a logic game with unary inputs or binary inputs. As the dimension of the game increases, so too do the degrees of freedom concerning modelling decisions.

6.4 Relation to Other Work

6.4.1 Skyrms on Information Processing

Skyrms (2000b, 2004, 2010a) initially suggested that the Lewis signalling game could be modified to show how logical inference, in the context of information processing, might evolve. In one case, the sender's observations might include disjunctive information about states, giving rise to uncertainty; in another case, we might postulate multiple senders who observe different information to see whether the receiver is capable of synthesising these several messages to perform the correct action. In this case, logical inference might be attributed to the receiver to the extent that she can make use of multiple signals, each of which conveys perfect information about a *partition* of nature, but neither of which has full information about the state of nature on its own.

Barrett (2007, 2009) shows how the sender(s) and receiver can simultaneously learn to partition nature appropriately, and code for that partition. Senders learn to coordinate so that they jointly send maximal information to the receiver.[21] Though the examples that we have seen are relatively simple, Skyrms (2010a) sees this as a significant achievement: 'The inferences [that the receiver learns] are not just valid inferences, but also the *relevant* valid inferences for the task at hand' (141-2).

[21] See also the discussion in LaCroix (2020a).

When the senders and receiver co-evolve the unary and binary inputs, they are in general not pooling their strategies (in the case of NAND), but in fact learning the right kind of inputs for this mode of composition.

6.4.2 Steinert-Threlkeld on Logical Operations

Steinert-Threlkeld (2014) discusses a signalling game in which there are *disjunctive* states. That is, rather than nature being partitioned into n distinct states, and the sender being guaranteed perfect information about each of these unique states, we might suppose that the sender has imperfect information about the states, or that a *set* of states is actual. This gives rise to the idea of a *disjunctive* state—e.g., $s_1 \vee s_2$. In the simplest case, the payoffs are adjusted so that, e.g., a_1 chosen in the disjunctive state $s_1 \vee s_2$ is afforded a payoff of 0.5. Since the disjunctive state will always be mapped to one or the other of the acts corresponding to its disjuncts, the expectation is a payoff of 1 half of the time, and a payoff of 0 the other half of the time (when the states are equiprobable). Similarly, we might introduce the possibility for *disjunctive actions*—which Steinert-Threlkeld (2014) refers to as 'cautious' actions. These will, in general, receive a payoff in the interval $(0.5, 1.0)$—a cautious action performs better than a single action due to its being cautious, but worse than perfect, due to a lack of complete information.

On this set-up, Steinert-Threlkeld (2014) shows how a 'function-word', akin to negation, can evolve naturally. However, the results of his simulations are fairly limitative. Steinert-Threlkeld (2014) concludes that the setup he examines fails to address the *origin* of function words since the model 'effectively builds in' a particular signal *as* a function word. Other models that deal with syntax or composition—e.g., Barrett (2006, 2007, 2009); Franke (2013, 2014, 2016) lack functionality in the requisite sense.

Steinert-Threlkeld (2014) examines how function words might evolve in a simple Lewis signalling game and uses a 'nested urn' process for the sender. However, it is not the receiver that has nested urns, but the sender. His set-up is a basic six-state, six-act signalling game—where the states include atomic and disjunctive states: $s_0, s_1, s_2, (s_0 \vee s_1), (s_0 \vee s_2)$ and $(s_1 \vee s_2)$. The sender has three atomic messages and one message that comes to be used as a functional message. This is where the urn-nesting comes into play for the sender: if she chooses m_3, then she sends that message followed by another message chosen from $\{s_0, s_1, s_2\}$. The receiver's urn for m_3 does not contain an act, but a function f, so that when she receives two messages, she performs some function on the act corresponding to the second message. Steinert-Threlkeld (2014) assumes that the sender and receiver have already coordinated on a signalling system for the atomic states, signals, and actions—their urns are already populated at the outset of the game.

Further, Steinert-Threlkeld (2014) says in the discussion of his game that the results were reasonably limitative: 'Given the inability of our learning algorithm to ensure convergence in the presence of the constant $a_1 \vee a_2$ function even under these ideal conditions, one may doubt whether this is the right approach to study the evolution of function words'.

6.4.3 Barrett and Skyrms on Self-Assembly

One thing that came out in the discussion of unary functions is that the ID game and the NEG game are structurally more similar to one another than they are to either of the TAUT game and the CONT game, and vice-versa. This highlights a subtlety that is ignored in the analysis of Barrett and Skyrms (2017).

Note that there are eleven unique combinations of outputs for binary logical operators; see Table 6.12. However, Barrett and Skyrms (2017) suggest, in a footnote, that if one had five binary logical operators—namely, one for each possible number of false outputs; see

Input		*Output*															
p	q	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	O_{16}
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Table 6.12: Fine-grained unique output for binary operations

Table 6.13—one could get the other eleven by template transfer, and each would evolve an order of magnitude faster this way than on its own. They show this for evolving OR from NAND. While this is true, it is not clear how efficiently one might be able to evolve, e.g.,

p	q	O_1	O_2	O_3	O_4	O_5
1	1	1	1	1	1	0
1	0	1	1	1	0	0
0	1	1	1	0	0	0
0	0	1	0	0	0	0

Table 6.13: Coarse-grained unique output for binary operations

AND from NAND, given that they have a different number of falses in their truth tables. However, it is obvious how one would evolve each binary operator from combinations of two unary operators. For example, we saw that NAND is just the composition of TAUT and NEG. If we combine NEG and ID, we get IFF, for example; see Table 6.14.

		a_1	a_2
		0	**1**
s_{11}	$\langle 0,0\rangle$	1	0
s_{12}	$\langle 0,1\rangle$	1	0
s_{21}	$\langle 1,0\rangle$	0	1
s_{22}	$\langle 1,1\rangle$	0	1

(a) Payoffs for 2-ID game

		a_1	a_2
		0	**1**
s_{11}	$\langle 0,0\rangle$	0	1
s_{12}	$\langle 0,1\rangle$	1	0
s_{21}	$\langle 1,0\rangle$	1	0
s_{22}	$\langle 1,1\rangle$	0	1

(b) Payoffs for IFF game

Table 6.14: Payoffs for 2-ID game and IFF game

It was shown in Section 6.3.2 that NAND is simpler to evolve than, e.g., IFF. The reason for this is that the underlying unary logical operations that compose NAND are not equally difficult to learn—the TAUT game allows the sender and receiver to coordinate upon a con-

vention significantly more quickly than the NEG game. This distinction is ignored in Barrett and Skyrms (2017), and it was mentioned that though they do not look at the simplest case, they also do not examine the most difficult case.

It is important to note precisely *what* is being evolved in the NAND game under template transfer, as it is presented in Barrett and Skyrms (2017). In that game, it is presupposed that the senders and receiver have *already* evolved the NAND disposition to encode (s_A NAND s_B) successfully. What template transfer shows is that if we assume that this disposition is already evolved, and we present the players with a novel context, where anything might be appropriate, but NAND happens to be appropriate, they can learn very quickly to utilise their pre-evolved NAND disposition—an order of magnitude quicker than if they were to evolve the identical disposition from scratch again. This makes sense, given how individuals learn, and the argument is well supported by the empirical evidence that they cite.

However, it is also important to note what the NAND game does not do. In particular, the purpose of presenting the NAND game as it is presented is simply to show that utilising template transfer can be more efficient than learning novel dispositions from scratch. Their argument is not intended to show *how* such logical dispositions might evolve in the first place. NAND in a new context evolves quicker via template transfer when a previous NAND disposition was already present. This does not show, for example, how to evolve OR from NAND.

Nonetheless, Barrett and Skyrms (2017) do suggest that this is the sort of process they have in mind: 'Once NAND has evolved, it may be appropriated to a new context by template transfer to play the role of a different logical operator more efficiently than that operator might evolve on its own' (350-1). They report that evolving OR from NAND happens an order of magnitude faster than evolving OR from scratch. This is precisely because OR and NAND both contain exactly one false output. In a suggestive footnote, they point out that '[m]ore generally, if one had five binary logical operators, one for each possible number of

false outputs, one could get the other eleven by template transfer. And each would evolve an order of magnitude faster this way than on its own'. However, it would not do to have *four* binary logical operators, and it is unclear whether evolving, e.g., XOR from NAND would be any more efficient than evolving XOR from scratch. Further, template transfer does not show how to evolve a ternary-input logical operator from a binary-input logical operator. Though they do not suggest this, their process of modular composition should be sufficient for this purpose. Namely, one module might consist of a NAND game, with the other module consisting of a single-input.[22]

The evolution of more complex functions out of simpler ones, as has been presented in this chapter, shows how to evolve NAND (or any other logical function) in parts—e.g., evolving binary NAND out of two unary parts. This is analogous to the story that Barrett and Skyrms (2017) give for the evolution of signalling games in parts—e.g., out of cue-reading and sensory manipulation. The problem that they introduce is then how these separate games link up to form a full signalling game. The problem that has been introduced here, analogously, is how an evolutionary process might search the space of such logical structures to show how they can link up in the right ways.

This is precisely the process we have seen in this chapter. The main difference is that we started with *unary* logical operations and showed how these could compose to more efficiently evolve *binary* logical operations. Further, this process extends in a natural way to compose *ternary* logical operations from two BINARY logical operations, or four unary logical operations, and so on *ad infinitum*. Further, I showed how it is possible for these dispositions to *co-evolve* in the binary case.

In some sense, the efficacy of template transfer on a pre-evolved disposition should come as no surprise. Since the sender dispositions in the original context are *fixed*, the sender in the new context is effectively playing a *sensory-manipulation* game. Furthermore, since the two

[22]See Barrett (2019).

senders are completely independent, these branches constitute two independent 2×2 sensory-manipulation games—the work of composing these two games has already been achieved by the pre-evolved disposition.

6.4.4 Barrett (et al.) Hierarchical Models of Compositionality

The type of compositionality that evolves in the co-evolutionary NAND game is hierarchical, in the sense that the binary logical operator, NAND, is a composition of two unary logical operators, TAUT and NEG. Barrett et al. (2018) have suggested a hierarchical model of compositionality to explain how compositional signalling might arise in nature. Their models of hierarchical composition include a basic component, which consists of a two sender, one receiver signalling game, and an 'executive' component, which includes an executive sender and an executive receiver. The basic idea is that the executive agents can learn to influence the behaviour of the basic agents. They present three models in decreasing strength of modelling assumptions. Their first model assumes that the basic senders have pre-established representational roles; their second model relaxes this assumption and shows how the senders can co-evolve representational roles; the third model shows how costly signalling alone can lead agents (without pre-established roles) to evolve a compositional communication system.

In their simplest model, where the agent roles are pre-established, nature chooses two *properties* and a *context*. The properties are *animal* (cat or dog) and *colour* (black or white), so that there are four 'states'—*black dog*, *white dog*, *black cat*, and *white cat*. The context determines which information from the state is relevant for a successful action—*colour*, *animal*, or *both*. The sender and receiver are successful in coordination their actions just in case (1) the receiver performs the correct action for the given context and (2) the senders only sent the signals required for success given the context. So, the receiver must match the action to the state of nature, and the senders must communicate as efficiently as is possible.

The reason their model is *hierarchical* is that the executive sender and receiver constitute a higher 'layer' in the signalling game. Thus, the game itself is hierarchical. In a sense, the co-evolutionary game that I have presented here is also hierarchical. However, the structure of the game is slightly more complex than the game given by Barrett et al. (2018). The first sender's signal codes for which lower-order game ought to be played, and the second sender's signal codes for the appropriate action in that sub-game.

6.5 Conclusion

Barrett (2019) presents a model for complex NAND where agents learn to co-evolve the complex disposition $((X \text{ NAND } Y) \text{ NAND } Q)$ instead of the disposition $(P \text{ NAND } Q)$. The evolution of more complex functions out of simpler ones, as has been presented in this chapter, is complementary to Barrett (2019): this model shows how to evolve NAND (or any other logical function) in parts—e.g., evolving binary NAND out of two unary parts.

The key insight was to start with *unary* logical operations and show how these can compose to more efficiently evolve *binary* logical operations. Further, the assumption that the underlying simple disposition(s) must be pre-evolved was relaxed. Finally, this compositional extends in a natural way to learn *ternary* logical operations from two binary logical operations, or four unary logical operations, and so on *ad infinitum*.

This process is genuinely recursive to the extent that we can build arbitrarily complex games out of atomic unary simple games, or the complex games which are composed of atomic unary simple games. Thus, I believe this constitutes another step in the direction of Barrett and Skyrms (2017).

One thing that came out in the discussion of unary functions is that the ID- and NEG games are structurally more similar to one another than they are to either of the TAUT- and CONT

games and vice-versa. This highlights a subtlety that the analysis of (Barrett and Skyrms, 2017) ignores. Barrett and Skyrms (2017) suggest that '[o]nce NAND has evolved, it may be appropriated to a new context by template transfer to play the role of a different logical operator more efficiently than that operator might evolve on its own' (350-1). They report that evolving OR from NAND happens an order of magnitude faster than evolving OR from scratch.

In a suggestive footnote, they point out that '[m]ore generally, if one had five binary logical operators, one for each number of false outputs, one could get the other eleven by template transfer. And each would evolve an order of magnitude faster this way than on its own' (350). However, it would not do to have four binary logical operators. Further, template transfer fails to explain how one might be able to evolve, for example, AND from NAND, given that they have a different number of 'false' values in their truth tables. A virtue of the model presented here is that it is obvious how one would evolve each binary operator from combinations of two unary operators—it was mentioned that the unique permutations of two unary sub-games cover all 16 binary games. Finally, template transfer does not show how to evolve a ternary-input logical operator from a binary-input logical operator. Though Barrett and Skyrms (2017) do not suggest this, their process of modular composition more generally should be sufficient for this purpose, as has been shown here.

In some sense, the efficacy of template transfer on a pre-evolved disposition should come as no surprise. Since the sender dispositions in the original context are fixed, the sender in the new context is effectively playing a 'sensory-manipulation' game (Barrett and Skyrms, 2017). Furthermore, since the two senders are entirely independent, these branches constitute two independent 2×2 sensory-manipulation games—the work of composing these two games has already been achieved by the pre-evolved disposition. We have now seen, in a more general case, once the agents in a signalling context have evolved simple unary operators,

they might be able to use these previously evolved dispositions to learn new, more complex binary, ternary, etc. operations.

Though I have discussed logical operations as a concrete example, it should be clear that it is the process of modular composition itself that gives rise to efficacy in the evolution of complex dispositions. Such compositional processes (more generally) might be instantiated in nature in terms of the computational principles or neurobiological underpinnings of any adaptive decision-making process. Modular composition may arise, and aid efficacy or efficiency of such processes, in several different settings; for example, composing multiple sensory modalities, such as tactile and visual stimulation (Fazeli et al., 2019); arranging dominance relations to form a hierarchical representation of a social group (Seyfarth and Cheney, 2018); cognitive reasoning involving hierarchically organised decision-making (Sarafyazd and Jazayeri, 2019); or other such functional-demand protocols in nature, such as the availability of food, density of populations, and presence of predators in migratory species (Hopcraft et al., 2014).

When signals mediate these processes, the models provided here illustrate particular circumstances under which we might expect modular composition to be successful. When the agents in a signalling game are understood as distinct organisms, this may give rise to complex social behaviour; when they are interpreted as functional components of a single organism, this may give rise to complex cognition.

In the context of language origins, this sort of explanation is essential since an adequate description of how linguistic capacities might arise from simple communicative precursors is a diachronic story of how language gets to be complex over time via a combination of genetic and cultural evolution. Results of this sort help carve out the space of possibilities for how such dispositions may have arisen in the first place.

Concluding Remarks

The question of how language evolved is inherently challenging, for reasons discussed in the Introduction. Since language is complex and multi-faceted, it is common to narrow one's focus to salient constituents of language that are absent in animal communication systems. In this work, I likewise step back from the challenging question of how language evolved by examining simpler, related issues surrounding the prerequisites of this broader point of inquiry. The guiding question, in this simplified framework, is as follows: *How might simple signals evolve to become more complex?* In particular, what are the minimal requirements for rich degrees of complex communicative behaviour? Answering this narrower question is supposed to take steps toward answering some of the broader questions surrounding the evolution of natural language.

I assumed that the gradualist perspective is the correct approach to language origins, insofar as language is understood as a complex system, and complex systems evolve gradually. I did not argue for a gradualist position over a saltationist one (though see LaCroix (2020b)). This position is entailed by the signalling-game framework used throughout since it requires gradualist assumptions. This also assumes that linguistic communication is primarily *for* communication and that it is continuous with the simpler communication systems found in nature.

I argued that the emphasis placed on compositional syntax in language-origins research is misguided. First, there is an inherent asymmetry between the benefit that compositional syntax confers to the sender and receiver in a signalling context. Second, compositional syntax is an all-or-nothing property of language, so it cannot be given a genuinely gradualist treatment. Finally, there is no empirical evidence for any proto-compositional precursors in nature. That is to say, it is a mistake to assume that since compositional syntax provides a crucial difference between language and simple communication, research on language origins must, therefore, centre on the evolution of compositional syntax itself.

As an alternative explanatory target, I proposed that *reflexivity* provides a viable alternative to compositional syntax for explaining how complex communication systems may evolve out of simpler systems.

The main insight is that communication is a unique evolved mechanism to the extent that it can overtly influence the evolution of future communication. That is, once individuals learn to communicate, those abilities may be used to influence future communicative behaviour, leading to a feedback loop. We saw how modular compositional processes—including appropriation, transfer of learning, and analogical reasoning—which take advantage of pre-evolved dispositions can be more efficient and more efficacious for learning in novel contexts, over and above learning novel dispositions from scratch. So, individuals may learn to take advantage of prior dispositions. When individuals learn to utilise pre-evolved *communicative* dispositions to influence *future* communicative disposition, the pre-evolved signals become reflexive in the sense that they refer to the communicative context itself.

Therefore, reflexive communication systems are driven by modular compositional processes whereby entire *structures* compose to form more complex structures. This, in turn, gives rise to more complex communication. In fact, compositional *syntax* evolves in some circumstances as a consequence of reflexivity. So, this view accounts for the evolution of compositional syntax—though, as a byproduct instead of an explicit target.

The first step in the evolution of complex signalling behaviour is *functional reference*. In the simplest signalling game, the signals come to refer to states, giving rise to simple communication—i.e., signs that stand for states transfer information which reliably gives rise to appropriate actions for those states. When signals are *functionally* referential, they refer to something in the outside world and are context-dependent. For example, a leopard alarm call appears to give rise to the (proto-)concept *leopard*, to the extent that appropriate actions are elicited regardless of the (visible) presence of a leopard. When (false) alarm calls are sent repeatedly, they are ignored—they cease to 'refer' appropriately to a predator because the context has shifted.

Functional reference requires some degree of abstraction, giving rise to *concepts*. This is further elucidated by the fact that responses to a leopard alarm call followed by an eagle-cry elicits a response of surprise on the part of the receiver (due to additional information conferred by the subsequent cry). In contrast, a leopard alarm call followed by a leopard-growl does not (since no further information is conferred in this case).

Once a communication system allows individuals to refer, *abstractly*, it becomes possible for the referent of a signal to be the communication context itself. Such a signal is reflexive in a simple sense. In a simple signalling game, which is a 'flat' structure, the signals come to 'refer to' states of the world. As we saw in Chapter 5, when the output (i.e., action) of a simple signalling game is *incorrect* for a particular state, that output itself becomes the appropriate input for a pre-evolved binary signalling game, where the signals have come to mean something like 'yes'/'no'. In this case, the signals *functionally* refer to the object-level game itself. This gives rise to a *hierarchical* structure. Once such a hierarchical structure evolves, it is possible for agents to learn non-trivially compositional communication, as was shown in Chapter 6.

Therefore, reflexivity, as an alternative explanatory target to compositional syntax, depends upon a process of modular composition which, in turn, gives rise to compositional syntax.

Thus, I have demonstrated that the *reflexivity* of language provides an apt explanatory target for an evolutionary account of language.

Furthermore, I have argued that reflexivity (unlike compositionality) is a *graded* concept to the extent that modular composition comes in degrees of complexity—e.g., appropriation is simpler than transfer learning, which is simpler than analogical reasoning. Therefore, unlike compositionality, it is possible to give an account of the evolution of reflexive communication from a genuinely gradualist perspective. Furthermore, I have argued that (unlike compositionality), there is empirical evidence of evolutionary precursors—i.e., something like *proto*-reflexive communication. In particular, proto-compositional syntax appears only to be present in old-world monkeys and some species of bird. This type of syntactic communication is *analogous* rather than *homologous*.

The evolution of language is a challenging and complex subject. The purpose of this dissertation was not to provide a definitive answer to how language evolved. Instead, the aim was to highlight some apparent flaws with the current direction of language-origins research and to suggest a novel explanatory target on which an account of the evolution of language may focus. I have demonstrated that there are good reasons to take this target seriously moving forward with respect to both possibility and plausibility. Further, I have shown that it is at least possible for reflexivity to give rise to more complex communication.

Where appropriate, I have touched upon relevant connections between the evolution of communicative structures and the evolution of cognitive or social structures. A full account will need to maintain sensitivity to each of these. Furthermore, I have suggested (Chapter 4 and elsewhere (LaCroix, 2019a,b)) that the theoretical insights brought to light here may have practical implications for ongoing work in machine learning. The extent to which reflexivity also plays a role in the evolution of complex cognitive structures may come to bear on moving from narrow to general artificial intelligence.

For example, Graziano's (and colleagues') proposed *attention-schema theory* of consciousness (Graziano, 2013, 2019) appears to invoke the sort of reflexive structure that I have been describing to give a brain basis of consciousness. (Graziano, 2010; Graziano and Kastner, 2011) provide evidence that the brain has specialised machinery for computing features of awareness, and this machinery attributes awareness to other people in social contexts—i.e., the perception of other minds. This is built on the foundation of previous work that explains how the cerebral cortex monitors the space around the body and controls movement within that space (Graziano and Gross, 1993; Graziano et al., 1994, 1997b,a, 1999, 2000; Cooke et al., 2003; Cooke and Graziano, 2004b,a; Graziano et al., 2005; Graziano, 2006; Aflalo and Graziano, 2007; Alfalo and Graziano, 2008; Graziano and Aflalo, 2007). The hypothesis is that the *same* cognitive machinery applies the feature of awareness to *itself*—thus giving rise to (the perception of) consciousness.[23]

Exploring the connection between reflexive cognitive structures and reflexive communicative structures constitutes a 'next step' for this new research programme. One benefit of applications in machine learning is that these hypotheses can be explicitly tested against prior baselines for generalisation in a number of artificial-intelligence contexts.

There is much work yet to be done.

[23]To provide a bit more detail: neuroscientific studies have suggested that particular parts of the cortex are recruited during social perception—this is how the brain constructs models of other people's minds (Fletcher et al., 1995; Goel et al., 1995; Brunet et al., 2000; Gallagher et al., 2000; Vogeley et al., 2001; Saxe and Kanwisher, 2003; Saxe and Wexler, 2005; Ciaramidaro et al., 2007). Furthermore, when these areas of the cortex are damaged, individuals fail in their awareness of events and objects around them (Karnath et al., 2001; Vallar and Perani, 2001).

Bibliography

Adret, Patrice (1993). Vocal Learning Induced With Operant Techniques: An Overview. *Netherlands Journal of Zoology*, 43: 125–142.

Aflalo, T. N. and Michael S. A. Graziano (2007). Relationship between unconstrained arm movement and single neuron firing in the macaque motor cortex. *Journal of Neuroscience*, 27(11): 2760–2780.

Aitken, Peter G. and William A. Wilson Jr. (1979). Discriminative Vocal Conditioning in Rhesus Monkeys: Evidence for Volitional Control? *Brain and Language*, 8: 227–240.

Akerlof, George A. (1976). The Economics of Caste and of the Rat Race and Other Woeful Tales. *The Quarterly Journal of Economics*, 90(4): 599–617.

Aldous, David J. (1985). Exchangeability and Related Topics. In Hennequin, P. L., editor, *École d'Été de Probabilités de Saint-Flour XIII — 1983*, volume 1117 of *Lecture Notes in Mathematics*, pages 1–198. Springer, Berlin, Heidelberg.

Alexander, Jason McKenzie, Brian Skyrms, and Sandy L. Zabell (2012). Inventing New Signals. *Dynamic Games and Applications*, 2: 129–145.

Alfalo, T. N. and Michael S. A. Graziano (2008). Four dimensional spatial reasoning in humans. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5): 1066–1077.

Alston, William (1964). *Philosophy of Language*. Prentice-Hall, Englewood Cliffs, NJ.

Altmann, S. A. (1962). Social Behavior of Anthropoid Primates: Analysis of Recent Concepts. In Bliss, Eugene L., editor, *Roots of Behavior: Genetics, Instinct, and Socialization in Animal Behavior*. Hoeber-Harper, New York.

Altmann, S. A. (1967). The Structure of Primate Social Communication. In Altmann, S. A., editor, *Social Communication Among Primates*, pages 325–362. University of Chicago Press, Chicago.

Anderson, Stephen R. (2004). *Doctor Doolittle's Delusion: Animals and the Uniqueness of Human Language*. Yale University Press, New Haven & London.

Andics, Attila, Anna Gábor, Márta Gácsi, Tamás Faragó, Dora Szabó, and Ádám Miklóski (2016). Neural Mechanisms for Lexical Processing in Dogs. *Science*, 353: 6303.

Andics, Attila, Márta Gácsi, Tamás Faragó, Anna Kis, and Ádám Miklóski (2014). Voice-Sensitive Regions in the Dog and Human Brain Are Revealed by Comparative fMRI. *Current Biology*, 24(5): 574–578.

Andreas, Jacob, Marcus Rohrbach, Trevor Darrell, and Dan Klein (2016). Neural Module Networks. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Arbib, Michael A. (2005). From Monkey-Like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics. *Behavioral and Brain Sciences*, 28: 105–167.

Argiento, Raffaelle, Robin Pemantle, Brian Skyrms, and Stanislav Volkov (2009). Learning to Signal: Analysis of a Micro-Level Reinforcement Model. *Stochastic Processes and Their Applications*, 119: 373–390.

Aristotle (1995a). History of animals, book ix. In Barnes, Jonathan, editor, *The Complete Works of Aristotle, The Revised Oxford Translation*, volume 1, pages 214–256. Princeton University Press, Princeton.

Aristotle (1995b). Metaphysics. In Barnes, Jonathan, editor, *The Complete Works of Aristotle, The Revised Oxford Translation*, volume 2, pages 1552–1728. Princeton University Press, Princeton.

Armstrong, Joshua (2018). The Social Origins of the Human Language Faculty. In *Philosophy of Science Association, 50th Biennial Meeting*, Seattle.

Arnold, Kate and Klaus Zuberbühler (2006a). Language Evolution: Semantic Combinations in Primate Calls. *Nature*, 441: 303.

Arnold, Kate and Klaus Zuberbühler (2006b). The Alarm-Calling System of Adult Male Putty-Nosed Monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour*, 72: 643–653.

Arnold, Kate and Klaus Zuberbühler (2008). Meaningful Call Combinations in a Non-Human Primate. *Current Biology*, 18: R202–R203.

Arnold, Kate and Klaus Zuberbühler (2013). Female Putty-Nosed Monkeys Use Experimentally Altered Contextual Information to Disambiguate the Cause of Male Alarm Calls. *PLOS ONE*, 8(6): e65660.

Arrow, Kenneth J. (1971). A Utilitarian Approach to the Concept of Equality in Public Expenditure. *The Quarterly Journal of Economics*, 85(3): 409–415.

Arthur, Brian W. (1993). On Designing Economic Agents That Behave Like Human Agents. *Journal of Evolutionary Economics*, 3: 1–22.

Aumann, Robert (1992). Irrationality in Game Theory. In Dasgupta, Partha, Douglas Gale, Oliver Hart, and Eric Maskin, editors, *Economic Analysis of Markets and Games*, pages 214–227. The MIT Press, Cambridge, MA.

Ayasse, Manfred, Robert J. Paxton, and Jan O. Tengö (2001). Mating Behavior and Chemical Communication in the Order Hymenoptera. *Annual Review of Entomology*, 46: 31–78.

Ayer, Alfred J. (1952). Negation. *Journal of Philosophy*, 49: 797–815.

Bacharach, Michael (1992). Backward Induction and Beliefs About Oneself. *Synthese*, 91(3): 247–284.

Bahdanau, Dzmitry, Shikhar Murty, Michael Noukhovitch, Thien Huu Nguyen, Harm de Vries, and Aaron Courville (2018). Systematic generalization: What is required and can it be learned? arXiv:1881.12889v1.

Bainton, Nigel J., Barrie W. Bycroft, Siri Ram Chhabra, Paul Stead, Linden Gledhill, Philip J. Hill, Catherine E. D. Rees, Michael K. Winson, George P. C. Salmond, Gordon S. A. B. Stewart, and Paul Williams (1992). A General Role for the *lux* Autoinducer in Bacterial Cell Signaling: Control of Antibiotic Biosynthesis in *Erwinia*. *Gene*, 116: 87–91.

Barnes, Jonathan (2001). *Early Greek Philosophy*. Penguin, London, 2 edition.

Barrett, H. Clark and Robert Kurzban (2006). Modularity in Cognition: Framing the Debate. *Psychological Review*, 113(3): 628–647.

Barrett, Jeffrey (2006). Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and Evolution of Grammar. *Technical Report, Institute for Mathematical Behavioral Science*.

Barrett, Jeffrey (2007). Dynamic Partitioning and the Conventionality of Kinds. *Philosophy of Science*, 74: 527–546.

Barrett, Jeffrey (2009). The Evolution of Coding in Signaling Games. *Theory and Decision*, 67: 223–237.

Barrett, Jeffrey (2013). The Evolution of Simple Rule-Following. *Biological Theory*, 8(2): 142–150.

Barrett, Jeffrey (2016). On the Evolution of Truth. *Erkenntnis*, 81: 1323–1332.

Barrett, Jeffrey (2017). Truth and Probability in Evolutionary Games. *Journal of Experimental and Theoretical Artificial Intelligence*, 29(1): 219–225.

Barrett, Jeffrey and Kevin Zollman (2009). The Role of Forgetting in the Evolution and Learning of Language. *Journal of Experimental and Theoretical Artificial Intelligence*, 21(4): 293–309.

Barrett, Jeffrey A. (2014). Rule-Following and the Evolution of Basic Concepts. *Philosophy of Science*, 81(5): 829–839.

Barrett, Jeffrey A. (2018). The Evolution, Appropriation, and Composition of Rules. *Synthese*, 195(2): 623–636.

Barrett, Jeffrey A. (2019). Self-assembling games and the evolution of salience. Unpublished Manuscript. March, 2019. PDF File.

Barrett, Jeffrey A., Calvin T. Cochran, Simon Huttegger, and Naoki Fujiwara (2017a). Hybrid Learning in Signaling Games. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(5): 1–9.

Barrett, Jeffrey A. and Travis LaCroix (2020). Epistemology and the structure of language. *Erkenntnis*. Forthcoming.

Barrett, Jeffrey A. and Brian Skyrms (2017). Self-Assembling Games. *The British Journal for the Philosophy of Science*, 68(2): 329–353.

Barrett, Jeffrey A., Brian Skyrms, and Calvin Cochran (2018). Hierarchical Models for the Evolution of Compositional Language. Unpublished Manuscript. May, 2018. PDF File.

Barrett, Jeffrey A., Brian Skyrms, and Aydin Mohseni (2017b). Self-Assembling Networks. *British Journal for the Philosophy of Science*. Forthcoming.

Bartha, Paul (2016). Analogy and analogical reasoning. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

Bartha, Paul F. A. (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford University Press, Oxford.

Bassler, Bonnie L. (1999). How Bacteria Talk to Each Other: Regulation of Gene Expression by Quorum Sensing. *Current Opinion in Microbiology*, 2(6): 582–587.

Bauer, Wolfgang D. and Ulrike Mathesius (2004). Plant Responses to Bacterial Quorum-Sensing Signals. *Current Opinion in Plant Biology*, 7: 429–433.

Beck, Benjamin B. (1980). *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. Garland STPM Press, New York.

Beck, Jacob (2017). Can Bootstrapping Explain Concept Learning? *Cognition*, 158: 110–121.

Beck, Sarah R., Ian A. Apperly, Jackie Chappell, Carlie Guthrie, and Nicola Cutting (2011). Making Tools Isn't Child's Play. *Cognition*, 119(2): 301–306.

Beggs, A. W. (2005). On the Convergence of Reinforcement Learning. *Journal of Economic Theory*, 122: 1–36.

Belanger, Rachelle M. and Lynda D. Corkum (2009). Review of Aquatic Sex Pheromones and Chemical Communication in Anurans. *Journal of Herpetology*, 43: 184–191.

Bellhouse, David R. and Nicolas Fillion (2015). Le Her and Other Problems in Probability Discussed by Bernoulli, Montmort and Waldegrave. *Statistical Science*, 30(1): 26–39.

Bercovitch, Fred B. (1988). Coalitions, Cooperation and Reproductive Tactics Among Adult Male Baboons. *Animal Behavavior*, 36: 1198–1209.

Bergman, Thore J., Jacinta C. Beehner, Dorothy L. Cheney, and Robert M. Seyfarth (2003). Hierarchical Classification by Rank and Kinship in Baboons. *Science*, 302: 1234–1236.

Bergstrom, Carl T. and Martin Rosvall (2011). The Transmission Sense of Information. *Biology and Philosophy*, 26: 159–176.

Bermejo, M. and A. Omedes (1999). Preliminary Vocal Repertoire and Vocal Communication of Wild Bonobos (*Pan paniscus*) at Lilungu (Democratic Republic of Congo). *Folia Primatologica*, 70: 328–357.

Bermúdez, J. L. (2003). *Thinking Without Words*. Oxford University Press, Oxford.

Berthet, Mélissa, Geoffrey Mesbahi, Aude Pajot, Cristiane Cäsar, Christof Neumann, and Klaus Zuberbühler (2018a). Titi Monkey Alarm Sequences: When Combining Creates Meaning. PSA 2018: The 26th Biennial Meeting of the Philosophy of Science Association, 1–4 November 2018.

Berthet, Mélissa, Christof Neumann, Geoffrey Mesbahi, Cristiane Cäsar, and Klaus Zuberbühler (2018b). Contextual Encoding in Titi Monkey Alarm Call Sequences. *Behavioral Ecology and Sociobiology*, 72(1): 8.

Berwick, Robert C. (1998). Language Evolution and the Minimalist Program: The Origins of Syntax. In Hurford, James R., Michael Studdert-Kennedy, and Chris Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 320–340. Cambridge University Press, Cambridge.

Berwick, Robert C. and Noam Chomsky (2011). The Biolinguistic Program: The Current State of its Development. In Sciullo, Anna Maria Di and Cedric Boeckx, editors, *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, Oxford Studies in Biolinguistics, pages 19–41. Oxford University Press, Oxford.

Berwick, Robert C. and Noam Chomsky (2016). *Why Only Us: Language and Evolution*. The MIT Press, Cambridge, MA.

Berwick, Robert C., Marc D. Hauser, and Ian Tattersall (2013). Neanderthal language? Just-So Stories Take Center Stage. *Frontiers in Psychology*, 4: 1–2.

Berwick, Robert C., Paul Pietroski, Beracah Yankama, and Noam Chomsky (2011). Poverty of the Stimulus Revisited. *Cognitive Science*, 35(7): 1207–1242.

Bicchieri, Cristina (1993). *Rationality and Coordination*. Cambridge University Press, Cambridge.

Bicchieri, Cristina and Ryan Muldoon (2014). Social norms. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2014 edition.

Bickerton, Derek (1981). *Roots of Language*. Karoma Press, Ann Arbor, MI.

Bickerton, Derek (1990). *Language and Species*. University of Chicago Press, Chicago.

Bickerton, Derek (1998). Catastrophic Evolution: The Case for a Single Step from Protolanguage to Full Human Language. In Hurford, James R., Michael Studdert-Kennedy, and Chris Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 341–358. Cambridge University Press, Cambridge.

Bickerton, Derek (1999). How to Acquire Language Without Positive Evidence: What Acquisitionists Can Learn from Creoles. In DeGraff, M., editor, *Language Creation and Language Change: Creolization, Diachrony, and Development*, pages 49–74. MIT Press, Cambridge, MA.

Bickerton, Derek (2000). How Protolanguage Became Language. In Knight, Chris, Michael Studdert-Kennedy, and James R. Hurford, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 264–284. Cambridge University Press, Cambridge.

Bickerton, Derek (2009). Recursion: Core of Complexity or Artifact of Analysis? In Givón, T. and M. Shibatani, editors, *Syntactic Complexity: Diachrony, Acquisition, Neurocognition, Evolution*, pages 531–544. John Benjamins, Philadelphia.

Biebach, H., M. Gordijn, and J. R. Krebs (1989). Time-and-place Learning by Garden warblers, *Sylvia borin*. *Animal Behaviour*, 37(3): 353–360.

Billett, Stephen (1998). Appropriation and Ontogeny: Identifying Compatibility Between Cognitive and Sociocultural Contributions to Adult Learning. *International Journal of Lifelong Education*, 17(1): 21–34.

Binmore, Kenneth (1987). Modeling Rational Players, Part I. *Economics and Philosophy*, 3(2): 179–214.

Binmore, Kenneth (1988). Modeling Rational Players, Part II. *Economics and Philosophy*, 4(1): 9–55.

Birch, Jonathan (2014). Propositional Content in Signalling Systems. *Philosophical Studies*, 171(3): 493–512.

Björnerstedt, Jonas and Jörgen W. Weibull (1993). Nash Equilibrium and Evolution by Imitation. In Arrow, Kenneth J., Enrico Colombatto, Mark Perlman, and Christian Schmidt, editors, *The Rational Foundations of Economic Behavior*, pages 155–171. Macmillan Press, London.

Blackburn, J. M. (1936). Acquisition of Skill: An Analysis of Learning Curves. *HIRB Report*, 73.

Blough, Donald S. (1959). Delayed Matching in the Pigeon. *Journal of the Experimental Analysis of Behavior*, 2(2): 151–160.

Blume, A., D. DeJong, G. Neumann, and N. E. Savin (2002). Learning and Communication in Sender–Receiver Games: An Econometric Investigation. *Journal of Applied Econometrics*, 17: 225–247.

Blumstein, Daniel T. (2007). The Evolution, Function, and Meaning of Marmot Alarm Communication. *Advances in the Study of Behavior*, 37: 371–401.

Boesch, Christophe (1991). Teaching Among Wild Chimpanzees. *Animal Behavior*, 41: 530–532.

Boesch, Christophe and Hedwige Boesch (1983). Optimization of Nut-Cracking in Wild Chimpanzees. *Behaviour*, 83: 265–286.

Boesch, Christophe and Hedwige Boesch-Achermann (2000). *The Chimpanzees of the Taï Forest*. Oxford University Press, Oxford.

Bolhuis, Johan J., Ian Tattersall, Noam Chomsky, and Robert C. Berwick (2014). How Could Language Have Evolved? *PLoS Biology*, 12(8): e1001934.

Bolt, Laura M. and Erica Tennenhouse (2017). Contact calling behaviour in the male ring-tailed lemur (Lemur catta). *Ethology*, 123(9): 614–626.

Bond, Alan B., Alan C. Kamil, and Russell P. Balda (2003). Social Complexity and Transitive Inference in Corvids. *Animal Behaviour*, 65(3): 479–487.

Bonte, Élodie, Caralyn Kemp, and Joël Fagot (2014). Age Effects on Transfer Index Performance and Executive Control in Baboons (*Papio papio*). *Frontiers in Psychology*, 5: 188.

Börgers, Tilman and Rajiv Sarin (1997). Learning through Reinforcement and the Replicator Dynamics. *Journal of Economic Theory*, 74: 235–265.

Börgers, Tilman and Rajiv Sarin (2000). Naive Reinforcement Learning with Endogenous Aspirations. *International Economic Review*, 41: 921–950.

Bradbury, Jack W. and Sandra L. Vehrencamp (2011). *Principles of Animal Communication*. Sinauer, Sunderland, MA, 2 edition.

Bradshaw, Gary (1993). Beyond Animal Language. In Roitblat, Herbert L., Louis M. Herman, and Paul E. Nachtigall, editors, *Language and Communication: Comparative Perspectives*, pages 25–44. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bratman, Michael E. (1999). *Faces of Intention*. Cambridge University Press, Cambridge.

Bredo, Eric (1994). Reconstructing Educational Psychology: Situated Cognition and Deweyian Pragmatism. *Trends in Ecology & Evolution*, 29(1): 23–35.

Brochhagen, Thomas (2015). Minimal Requirements for Productive Compositional Signaling. In *CogSci*, pages 285–290.

Brown, R. (1973). *A First Language: The Early Stages*. Harvard University Press, Cambridge, MA.

Brown, Steven (2000). The 'Musilanguage' Model of Music Evolution. In Wallin, N. L., B. Merker, and S. Brown, editors, *The Origins of Music*, pages 271–300. MIT Press, Cambridge, MA.

Bruner, Justin, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger (2018). David Lewis in the Lab: Experimental Results on the Emergence of Meaning. *Synthese*, 195(2): 603–621.

Brunet, E., Y. Sarfati, M. C. Hardy-Baylé, and J. Decety (2000). A pet investigation of the attribution of intentions with a nonverbal task. *NeuroImage*, 11(2): 157–166.

Bruschini, Claudia, Rita Cervo, and Stefano Turillazzi (2010). Pheromones in Social Wasps. *Vitamins and Hormones*, 83: 521–549.

Brusse, Carl and Justin Bruner (2017). Responsiveness and Robustness in the David Lewis Signaling Game. *Philosophy of Science*, 84(5): 1068–1079.

Bshary, Redouan and Manuela Würth (2001). Cleaner Fish *Labroides dimidiatus* Manipulate Client Reef Fish by Providing Tactile Stimulation. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 268: 1495–1501.

Bush, Robert R. and Frederick Mosteller (1955). *Stochastic Models for Learning*. John Wiley & Sons, Oxford.

Byrne, Richard W. and Andrew Whiten, editors (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Oxford. Clarendon Press.

Call, Josep, Juliane Bräuer, Juliane Kaminski, and Michael Tomasello (2003). Domestic Dogs (Canis familiaris) are Sensitive to the Attentional State of Humans. *Journal of Comparative Psychology*, 117(3): 257–263.

Call, Josep and Michael Tomasello (2007). *The Gestural Communication of Apes and Monkeys*. Lawrence Erlbaum, London.

Calvo, Francisco and Eliana Colunga (2003). The Statistical Brain: Reply to Marcus The Algebraic Mind. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25: 210–215.

Candolin, Ulrika (2003). The Use of Multiple Cues in Mate Choice. *Biological Reviews*, 78: 575–595.

Cangelosi, Angelo and Domenico Parisi (2002). Computer Simulation: A New Scientific Approach to the Study of Language Evolution. In *Simulating the Evolution of Language*, pages 3–28. Springer, London.

Carey, Susan (2004). Bootstrapping and the Origin of Concepts. *Daedalus*, 133(1): 59–68.

Carey, Susan (2009a). *The Origin of Concepts*. Oxford Series in Cognitive Development. Oxford University Press, Oxford.

Carey, Susan (2009b). Where our Number Concepts Come From. *The Journal of Philosophy*, 106(4): 220–254.

Carey, Susan (2011a). Author's Response: Concept Innateness, Concept Continuity, and Bootstrapping. *Behavioral and Brain Sciences*, 34(3): 152–162.

Carey, Susan (2011b). Precis of the Origin of Concepts. *Behavioral and Brain Sciences*, 34(3): 113–124.

Carey, Susan (2014). On Learning New Primitives in the Language of Thought: Reply to Rey. *Mind & Language*, 29(2): 133–166.

Carruthers, Peter (2002). The Cognitive Functions of Language. *Behavioral and Brain Sciences*, 25(3): 657–725.

Carruthers, Peter (2003). Monitoring Without Metacognition. *Behavioral and Brain Sciences*, 26(3): 242–243.

Carruthers, Peter (2006). *The Architecture of the Mind*. Oxford University Press, Oxford.

Carstairs-McCarthy, Andrew (1998). Synonymy Avoidance, Phonology, and the Origin of Syntax. In Hurford, James R., Michael Studdert-Kennedy, and Chris Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 279–296. Cambridge University Press, Cambridge.

Cäsar, Cristiane, Richard Byrne, Robert J. Young, and Klaus Zuberbühler (2012). The Alarm Call System of Wild Black-Fronted Titi Monkeys, *Callicebus nigrifrons*. *Behavioral Ecology and Sociobiology*, 66(5): 653–667.

Catteeuw, David and Bernard Manderick (2014). The Limits and Robustness of Reinforcement Learning in Lewis Signaling Games. *Connection Science*, 26(2): 161–177.

Cavalli-Sforza, Luigi L. (2001). *Genes, Peoples, and Languages*. University of California Press, Berkeley and Los Angeles.

Chalmers, David J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3): 200–219.

Chappell, Jackie and Alex Kacelnik (2002). Tool Selectivity in a Non-Primate, the New Caledonian Crow (Corvus moneduloides). *Animal Cognition*, 5: 71–78.

Cheney, Dorothy and Robert Seyfarth (1985). Vervet Monkey Alarm Calls: Manipulation Through Shared Information? *Behaviour*, 94(1): 150–166.

Cheney, Dorothy and Robert Seyfarth (1990). *How monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press, Chicago.

Cheney, Dorothy L. and Robert M. Seyfarth (1982). How Vervet Monkeys Perceive Their Grunts: Field Playback Experiments. *Animal Behaviour*, 30(3): 739–751.

Cheney, Dorothy L. and Robert M. Seyfarth (2007). *Baboon Metaphysics: The Evolution of a Social Mind*. University of Chicago Press, Chicago.

Chomsky, Noam (1956a). On the Limits of Finite State Description. *MIT Research Lab in Electronics Quarterly Progress Report*, 42: 64–65.

Chomsky, Noam (1956b). The Range of Adequacy of Various Types of Grammars. *MIT Research Lab in Electronics Quarterly Progress Report*, 41: 93–96.

Chomsky, Noam (1956c). Three Models for the Description of Language. *IRE Transactions on Information Theory IT-2*, 3: 113–124. Reprinted in Readings in Mathematical Psychology 2, edited by R. Luce, R. Bush, and E. Galanter, pp. 105-24. New York: Wiley and Sons, 1965.

Chomsky, Noam (1958). Some Properties of Phrase Structure Grammars. *MIT Research Lab in Electronics Quarterly Progress Report*, 49: 108–111.

Chomsky, Noam (1959a). A Note on Phrase Structure Grammars. *Information and Control*, 2: 393–395.

Chomsky, Noam (1959b). On Certain Formal Properties of Grammars. *Information and Control*, 2: 137–167. Reprinted in Readings in Mathematical Psychology 2, edited by R. Luce, R. Bush, and E. Galanter, pp. 125–55. New York: Wiley and Sons, 1965.

Chomsky, Noam (1962). Context-Free Grammars and Pushdown Storage. *MIT Research Lab in Electronics Quarterly Progress Report*, 65: 187–194.

Chomsky, Noam (1963). Formal properties of grammars. In R. Luce, R. Bush and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 2, pages 323–418. Wiley and Sons, New York.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, Noam (1979). Human Language and Other Semiotic Systems. *Semiotica*, 25: 31–44.

Chomsky, Noam (1980a). On Cognitive Structures and their Development: A reply to Piaget. In Piattelli-Palmarini, M., editor, *Language and Learning: the Debate between Jean Piaget and Noam Chomsky*, pages 175–176. Harvard University Press, Harvard.

Chomsky, Noam (1980b). *Rules and Representations*. Basil Blackwell, London.

Chomsky, Noam (1986). *Knowledge of Language*. Praeger, New York.

Chomsky, Noam (1988). *Language and Problems of Knowledge*. MIT Press, Cambridge, MA.

Chomsky, Noam (1995). *The Minimalist Program. Number 28 in Current Studies in Linguistics*. MIT Press, Cambridge, MA.

Chomsky, Noam (1999). *Derivation by Phase*. The MIT Press, Cambridge, MA.

Chomsky, Noam (2002). *On Nature and Language*. Cambridge University Press, Cambridge.

Chomsky, Noam (2002/1957). *Syntactic Structures*. Mouton de Gruyter, Berlin.

Chomsky, Noam (2005). Three Factors in Language Design. *Linguistic Inquiry*, 36: 1–22.

Chomsky, Noam (2010). Some Simple Evo Devo Theses: How True Might They Be for Language? In Larson, Richard K., Viviane Déprez, and Hiroko Yamakido, editors, *The Evolution of Human Language: Biolinguistic Perspectives*, pages 148–162. Cambridge University Press, Cambridge.

Chomsky, Noam (2017). The Language Capacity: Architecture and Evolution. *Psychonomic Bulletin & Review*, 24(1): 200–203.

Chomsky, Noam and George A. Miller (1958). Finite state languages. *Information and Control*, 1: 91–112. Reprinted in Readings in Mathematical Psychology 2, edited by R. Luce, R. Bush, and E. Galanter, pp. 156–71. New York: Wiley and Sons, 1965.

Chomsky, Noam and M. P. Schutzenberger (1963). The Algebraic Theory of Context-Free Languages. In Braffort, P. and D. Hirshberg, editors, *Computer Programming and Formal Systems: Studies in Logic*, pages 118–61. North Holland, Amsterdam.

Christiansen, Morten H. and Simon Kirby (2003). Language Evolution: The Hardest Problem in Science? In Kirby, Morten H. Christiansen Simon, editor, *Language Evolution*, pages 1–15. Oxford University Press, Oxford.

Churchward, C. Maxwell (1953). *Tongan Grammar*. Oxford University Press, London.

Ciaramidaro, A., M. Adenzato, I. Enrici, S. Erk, L. Pia, B. G. Bara, and H. Walter (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*, 45(13): 3105–3113.

Clark, Adam (1998). Magic Words: How Language Augments Human Cognition. In Carruthers, P. and J. Boucher, editors, *Language and Thought: Interdisciplinary Themes*, pages 162–83. Cambridge University Press, Cambridge.

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Clayton, Nicola S., D. P. Griffiths, Nathan J. Emery, and Anthony Dickinson (2001). Elements of Episodic-Like Memory in Animals. *Philosophical Transactions of the Royal Society B: Biologica Sciences*, 356(1413): 1483–1491.

Clayton, Nicola S., Juan C. Reboreda, and Alex Kacelnik (1997). Seasonal Changes in Hippocampus Volume in Parasitic Cowbirds. *Behavioural Processes*, 41(3): 237–243.

Clutton-Brock, T. H. and S. D. Albon (1979). The Roaring of Red Deer and the Evolution of Honest Advertisement. *Behaviour*, 69: 145–170.

Cohen, Philip R. and Hector J. Levesque (1991). Teamwork. *Nôus*, 25: 487–512.

Coleman, J. (1989). The Rational Choice Approach to Legal Rules. *Chicago-Kent Law Review*, 65: 177–191.

Collins, Chris and Edward Stabler (2016). A Formalization of Minimalist Syntax. *Syntax: A Journal of Theoretical Experimental and Interdisciplinary Research*, 19(1): 43–78.

Coltheart, Max (1999). Modularity and Cognition. *Trends in Cognitive Sciences*, 3(3): 115–120.

Compton, Richard and Christine Pittman (2010). Word-Formation by Phase in Inuit. *Lingua*, 120(9): 2167–2192.

Cook, Leslie Susan, Peter Smagorinsky, Pamela G. Fry, Bonnie Konopak, and Cynthia Moore (2002). Problems in Developing a Constructivist Approach to Teaching: One Teacher's Transition from Teacher Preparation to Teaching. *The Elementary School Journal*, 102(5): 389–413.

Cooke, D. F. and Michael S. A. Graziano (2004a). Sensorimotor integration in the precentral gyrus: Polysensory neurons and defensive movements. *Journal of Neurophysiology*, 91(4): 1648–1660.

Cooke, D. F. and Michael S. A. Graziano (2004b). Super-flinchers and nerves of steel: Defensive movements altered by chemical manipulation of a cortical motor area. *Neuron*, 43(4): 585–593.

Cooke, D. F., C. S. R. Taylor, T. Moore, and Michael S. A. Graziano (2003). Complex movements evoked by microstimulation of area vip. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10): 6163–6168.

Cooper, T. (1997). Assessing Vocabulary Size: So, What's the Problem? *Language Matters*, 26(1): 96–117.

Copi, Irving and Carl Cohen (2005). *Introduction to Logic*. Prentice-Hall, Upper Saddle River, NJ, 12 edition.

Corballis, Michael C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton University Press, Princeton, NJ.

Costa-Leonardo, Ana Maria and Ives Haifig (2010). Pheromones and exocrine glands in Isoptera. *Vitamins and Hormones*, 83: 521–549.

Couvillon, M. J. (2012). The Dance Legacy of Karl von Frisch. *Insectes Sociaux*, 59(3): 297–306.

Couvillon, Margaret J., Hunter L. F. Phillipps, Roger Schürch, and Francis L. W. Ratnieks (2012). Working Against Gravity: Horizontal Honeybee Waggle Runs Have Greater Angular Scatter than Vertical Waggle Runs. *Biology Letters*, 8: 540–543.

Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. John Wiley & Sons, Hoboken, 2 edition.

Crane, Tim (2003). *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation*. Routledge, London, 2 edition.

Crawford, Vincent P., Uri Gneezy, and Yuval Rottenstreich (2008). The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. *American Economic Review*, 98(4): 1443–1458.

Crawford, Vincent P. and Joel Sobel (1982). Strategic Information Transmission. *Econometrica*, 50(6): 1431–1451.

Cree, Viviene E. and Cathlin Macaulay (2000). *Transfer of Learning in Professional and Vocational Education*. Routledge, London and New York.

Crick, Francis and Cristof Koch (1990). Toward a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, 2: 263–275.

Crick, Francis and Cristof Koch (1995). Are We Aware of Neural Activity in Primary Visual Cortex? *Nature*, 375: 121–123.

Crystal, David (1987). *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge.

Culicover, Peter W. and Ray Jackendoff (2005). *Simpler Syntax*. Oxford University Press, New York.

Dahl, Östen (1979). Typology of Sentence Negation. *Linguistics*, 17: 79–106.

D'Arms, Justin, Robert Batterman, and Kryzyztof Gorny (1998). Game Theoretic Explanations and the Evolution of Justice. *Philosophy of Science*, 65(1): 76–102.

Darwin, Charles R. (1868). *Variation of Plants and Animals Under Domestication*. John Murray, London.

Darwin, Charles R. (1871). *The Descent of Man, and Selection in Relation to Sex*, volume 1. John Murray, London, 1 edition.

Darwin, Charles R. (2004/1859). *On the Origin of Species*. Routledge.

Davis, Hank (1992). Transitive Inference in Rats (Rattus norvegicus). *Journal of Comparative Psychology*, 106: 342–349.

Dawkins, Marian Stamp and Tim Guilford (1996). Sensory Bias and the Adaptiveness of Female Choice. *The American Naturalist*, 148(5): 937–42.

Dawkins, Richard (1996). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. W. W. Norton and Company, New York.

Dawkins, Richard and John R. Krebs (1978). Animal signals: Information or Manipulation? In Krebs, J. R. and N. B. Davies, editors, *Behavioural Ecology*, pages 282–309. Blackwell Scientific Publications, Oxford.

de Condillac, Etienne Bonnot (1971/1747). *Essai Sur l'Origine des Connaissances Humaines*. Scholar's Facsimiles and Reprints, Gainsville.

de Swart, Henriëtte (2010). *Expression and Interpretation of Negation: An OT Typology*, volume 77 of *Studies in Natural Language and Linguistic Theory*. Springer, Dordrecht.

de Waal, Frans B. M. (1989). *Peacemaking Among Primates*. Harvard University Press, Cambridge, MA.

de Walle, Gretchen A. Van, Susan Carey, and Meredith Prevor (2001). Bases for Object Individuation in Infancy: Evidence from Manual Search. *Journal of Cognition and Development*, 1(3): 249–280.

Deacon, Terrence W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. Norton, New York.

Dediu, Dan and Stephen C. Levinson (2013). On the Antiquity of Language: The Reinterpretation of Neanderthal Linguistic Capacities and its Consequences. *Frontiers in Psychology*, 4: 397.

Dehaene, Stanislas (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press, Oxford.

Delaney, Kevin J., J. Andrew Roberts, and George W. Uetz (2007). Male Signalling Behaviour and Sexual Selection in a Wolf Spider (*Araneae: Lycosidae*): A Test for Dual Functions. *Behavioral Ecology and Sociobiology*, 62: 67–75.

Dell, G. S., L. K. Burger, and W. R. Svec (1997). Language Production and Serial Order: A Functional Analysis and a Model. *Psychological Review*, 104: 123–147.

Dennett, Daniel C. (1971). Intentional Systems. *The Journal of Philosophy*, 68(4): 87–106.

Dennett, Daniel C. (1987). *The Intentional Stance*. The MIT Press, Cambridge, MA.

Di Sciullo, Anna Maria (2011). A Biolinguistic Approach to Morphological Variation. In Sciullo, A. M. Di and C. Boeckx, editors, *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, pages 305–326. Oxford University Press, Oxford.

Di Sciullo, Anna Maria (2013). Exocentric Compounds, Language and Proto-Language. *Language and Information Society*, 20: 1–26.

Dickhaut, John W., Kevin A. McCabe, and Arijit Mukherji (1995). An Experimental Study of Strategic Information Transmission. *Economic Theory*, 6(3): 389–403.

Digweed, S. M., L. M. Fedigan, and D. Rendall (2005). Variable Specificity in the Anti-Predator Vocalizations and Behaviour of the White-Faced Capuchin, *Cebus capucinus*. *Behaviour*, 142: 997–1021.

Diller, K. C. (1978). *The Language Teaching Controversy*. Newbury House, Rowley, MA.

Dobzhansky, Theodosius (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *American Biology Teacher*, 35: 125–129.

Donald, Merlin (1991). *Origins of the Modern Mind*. Harvard University Press, Cambridge, MA.

Donaldson, M. C., M. Lachmann, and C. T. Bergstrom (2007). The Evolution of Functionally Referential Meaning in a Structured World. *Journal of Theoretical Biology*, 246: 225–233.

Dretske, Fred (1981). *Knowledge and the Flow of Information*. The MIT Press.

Dryer, Matthew S. (1988). Universal of Negative Position. In Hammond, Michael, Edith Moravcsik, and Jessica Wirth, editors, *Studies in Syntactic Typology*, pages 93–124. John Benjamins, Amsterdam.

Dryer, Matthew S. (2008). Indefinite Articles. In Haspelmath, M., M. S. Dryer, D. Gil, and B. Comrie, editors, *The World Atlas of Language Structures Online*, chapter 38. Max Planck Digital Library, Munich. (Available online at http://wals.info/feature/38).

Dummett, Michael (1975). What is a Theory of Meaning? (II). In Evans, Gareth and John McDowell, editors, *Truth and Meaning*, pages 34–93. Oxford University Press, London. Reprinted in Dummett (1993). *The Seas of Language*.

Dummett, Michael (1989). Language and Communication. In George, A., editor, *Reflections on Chomsky*, pages 166–187. Blackwell, Oxford. Reprinted in Dummett (1993). *The Seas of Language*.

Dummett, Michael (1993a). *The Origins of Analytical Philosophy*. Duckworth, London.

Dummett, Michael (1993b). *The Seas of Language*. Clarendon Press, Oxford.

Dummett, Michael (1993/1978). What Do I Know When I Know a Language? In *The Seas of Language*, pages 94–105. Clarendon Press, Oxford. First published as a paper presented at the Centenary Celebrations, Stockholm University.

Dunbar, Kevin (2001). The analogical paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In Gentner, Dedre, Keith J. Holyoak, and Boicho N. Kokinov, editors, *The Analogical Mind: Perspectives from Cognitive Science*. MIT press, Cambridge, MA.

Düsing, Carl (1884). Die Regulierung des Geschlechtsverhältnisses bei der Vermehrung der Menschen, Tiere und Pflanzen. *Jenaische Zeitschrift für Naturwissenschaft*, 17: 593–940.

Düsing, Karl (1883). Die Factoren welche die Sexualität entscheiden. *Jenaische Zeitschrift für Naturwissenschaft*, 16: 428–464.

Eberhard, Anatol (1972). Inhibition and Activation of Bacterial Luciferase Synthesis. *Journal of Bacteriology*, 109: 1101–1105.

Eberhard, Anatol, Alma L. Burlingame, C. Eberhard, G. L. Kenyon, K. H. Nealson, and Norman Oppenheimer (1981). Structural Identification of Autoinducer of *Photobacterium fischeri* Luciferase. *Biochemistry*, 20(9): 2444–2449.

Edwards, A. W. F. (2000). Carl Düsing (1884) on *T*he Regulation of the Sex-Ratio. *Theoretical Population Biology*, 58(3): 255–257.

Ehrenreich, Barbara and Janet McIntosh (1997). The New Creationism: Biology Under Attack. *The Nation*, pages 11–16.

Ehrlich, Paul (1900). Croonian Lecture. On Immunity with Special Reference to Cell Life. *Proceedings of the Royal Society of London*, 66(424–433): 424–448.

Eibl-Eibesfeldt, I. (1973). The Expressive Behaviour of the Deaf- and Blind-Born. In Cranach, M. Von. and J. Vine, editors, *Social Communication and Movement*, page 163–194. Academic, London.

Ellis, Henry Carlton (1965). *The Transfer of Learning*. The Macmillan Company, New York.

Emery, Nathan J. and Nicola S. Clayton (2004). The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science*, 306(5703): 1903–1907.

Emmorey, Karen (2002). *Language, Cognition, and Brain: Insights from Sign Language Research*. Lawrence Erlbaum, Hillsdale.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai, Victor Wiebe, Takashi Kitano, Anothony P. Monaco, and Svante Pääbo (2002). Molecular Evolution of FOXP2, A Gene Involved in Speech and Language. *Nature*, 418(6900): 869–872.

Endler, John A. (1993). Some General Comments on the Evolution and Design of Animal Communication Systems. *Philosophical Transactions: Biological Sciences*, 340(1292): 215–225.

Engebrecht, Joanne, Kenneth Nealson, and Micahel Silverman (1983). Bacterial Bioluminescence: Isolation and Genetic Analysis of Functions from *Vibrio fischeri*. *Cell*, 32: 773–781.

Engebrecht, Joanne and Micahel Silverman (1984). Identification of Genes and Gene Products Necessary for Bacterial Bioluminescence. *Proceedings of the National Academy of Sciences of the United States of America*, 81: 4154–2158.

Engebrecht, Joanne and Micahel Silverman (1986). Regulation of Expression of Bacterial Genes for Bioluminescence. In Setlow, J. K. and A. Hollaender, editors, *Genetic Engineering*, pages 31–44. Plenum, New York.

Engebrecht, Joanne and Micahel Silverman (1987). Nucleotide Sequence of the Regulatory Locus Controlling Expression of Bacterial Genes for Bioluminescence. *Nucleic Acids Research*, 15: 10455–10467.

Enns, Linda C., Masahiro M. Kanaoka, Keiko U. Torii, Luca Comai, Kiyotaka Okada, and Robert E. Cleland (2005). Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Molecular Biology*, 58(3): 333–349.

Enquist, Magnus (1985). Communication During Aggressive Interactions with Particular Reference to Variation in Choice of Behaviour. *Animal Behavior*, 33: 1152–1161.

Erev, Ido and Alvin E. Roth (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *The American Economic Review*, 88(4): 848–881.

Evans, Christopher S. and Linda Evans (2007). Representational Signalling in Birds. *Biology Letters*, 3: 8–11.

Evans, Christopher S., Linda Evans, and Peter Marler (1993). On the Meaning of Alarm Calls: Functional Reference in an Avian Vocal System. *Animal Behaviour*, 46(1): 23–38.

Evans, Christopher S. and Peter Marler (1994). Food-Calling and Audience Effects in Male Chickens, Gallus gallus: Their Relationships to Food Availability, Courtship and Social Facilitation. *Animal Behaviour*, 47(5): 1159–1170.

Everett, Daniel L. (2005). Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language. *Current Anthropology*, 46(4): 621–646.

Fagot, Jöel, Edward A. Wasserman, and Michael E. Young (2001). Discriminating the Relation Between Relations: The Role of Entropy in Abstract Conceptualization by Baboons (*Papio papio*) and Humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, 27(4): 316–328.

Falk, Dean (2004). Prelinguistic Evolution in Early Hominins: Whence Motherese? *Behavioral and Brain Sciences*, 27: 491–450.

Falkenhainer, B., K. Forbus, and D. Gentner (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41: 1–63.

Farrell, Joseph and Matthew Rabin (1996). Cheap Talk. *The Journal of Economic Perspectives*, 10(3): 103–118.

Fazeli, N., M. Oller, J. Wu, Z. Wu, J. B. Tenenbaum, and A. Rodriguez (2019). See, Feel, Act: Hierarchical Learning for Complex Manipulation Skills with Multisensory Fusion. *Science Robotics*, 4(26): eaav3123.

Feigenson, Lisa, Susan Carey, and Elizabeth S. Spelke (2002). Infants' Discrimination of Number vs. Continuous Extent. *Cognitive Psychology*, 44(1): 33–36.

Fennell, B. A. (2001). *A History of English: A Sociolinguistic Approach*. Blackwell, Oxford.

Ferster, Charles Bohris (1960). Intermittent Reinforcment of Matching to Sample in the Pigeon. *Journal of the Experimental Analysis of Behavior*, 3(3): 259–272.

Fillion, Nicolas (2015). The Eighteenth-Century Origins of the Concept of Mixed-Strategy Equilibrium in Game Theory. In Zack, M. and E. Landry, editors, *Research in History and Philosophy of Mathematics*, pages 63–78. Birkhauser.

Fisher, R. A. (1958/1930). *The Genetical Theory of Natural Selection*. Dover, New York, 2 edition.

Fitch, W. Tecumseh (1997). Vocal Tract Length and Formant Frequency Dispersion Correlate with Body Size in Rhesus Macaques. *Journal of the Acoustical Society of America*, 102: 1213–1222.

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

Fitch, W. Tecumseh (2006). The Biology and Evolution of Music: A Comparative Perspective. *Cognition*, 100: 173–215.

Fitch, W. Tecumseh (2008). Kin Selection and 'Mother Tongues': A Neglected Component in Language Evolution. In Oller, D. K. and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 275–296. MIT Press, Cambridge, MA.

Fitch, W. Tecumseh (2010). *The Evolution of Language*. Cambridge University Press, Cambridge.

Fitch, W. Tecumseh (2017). Empirical Approaches to the Study of Language Evolution. *Psychonomic Bulletin & Review*, 24(1): 3–33.

Flache, A. and M. Macy (2002). Stochastic Collusion and the Power Law of Learning. *Journal of Conflict Resolution*, 46: 629–653.

Flemming, Timothy M., Roger K. R. Thompson, Michael J. Beran, and David A. Washburn (2011). Analogical Reasoning and the Differential Outcome Effect: Transitory Bridging of the Conceptual Gap for Rhesus Monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, 37(3): 353–360.

Fletcher, P. C., F. Happé, U. Frith, S. C. Baker, R. J. Dolan, R. S. Frackowiak, and C. D. Frith (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition*, 57(2): 109–128.

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

Fodor, Jerry (1984a). Observation Reconsidered. *Philosophy of Science*, 51: 23–43.

Fodor, Jerry (1984b). Semantics, Wisconsin Style. *Synthese*, 59: 231–250.

Fodor, Jerry (1987). *Psychosemantics*. MIT Press, Cambridge MA.

Fodor, Jerry (2000). *The Mind Doesn't Work That Way*. MIT Press, Cambridge MA.

Fodor, Jerry (2010). Woof, Woof: Review of *The Origin of Concepts*, by S. Carey. *Times Literary Supplement*, pages 7–8.

Fodor, Jerry A. and Zenon W. Pylyshyn (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28(1): 3–71.

Forbus, Kenneth (2001). Exploring analogy in the large. In Gentner, Dedre, Keith J. Holyoak, and Boicho N. Kokinov, editors, *The Analogical Mind: Perspectives from Cognitive Science*. MIT press, Cambridge, MA.

Forbus, Kenneth, Ronald W. Ferguson, and Dedre Gentner (1994). Incremental structure-mapping. In Ram, A. and K. Eiselt, editors, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 313–318. Lawrence Erlbaum, Hillsdale, NJ.

Forbus, Kenneth, Dedre Gentner, and Keith Law (1995). MAC/FAC: A Model of Similarity-based Retrieval. *Cognitive Science*, 19: 141–205.

Foster, Dean and Peyton Young (1990). Stochastic Evolutionary Game Dynamics. *Theoretical Population Biology*, 38: 219–232.

Frank, Michael J., Jerry W. Rudy, William B. Levy, and Randall C. O'Reily (2005). When Logic Fails: Implicit Transitive Inference in Humans. *Memory & Cognition*, 33(4): 742–750.

Franke, Michael (2013). Compositionality from reinforcement learning. In *Proceedings of Games, Interactive Rationality, Learning (G.I.R.L.) 2013*, volume i.

Franke, Michael (2014). Creative Compositionality from Reinforcement Learning in Signaling Games. In Cartmill, Erica A., Seén Roberts, Heidi Lyn, and Hannah Cornish, editors, *The Evolution of Language: Proceedings of the 10th International Conference*, volume 10 of *Evolang*, pages 82–89. World Scientific, Singapore.

Franke, Michael (2016). The Evolution of Compositionality in Signaling Games. *Journal of Logic, Language and Information*, 25(3): 355–377.

Frege, Friedrich Ludwig Gottlob (1923). Logische Untersuchungen. Dritter Teil: Gedankengefüge. *Beiträge zur Philosophie des deutschen Idealismus*, III: 36–51. Reprinted in Stoothoff, R. H. (Trans.) (1963). "Compound Thoughts" *Mind* 72(285): 1–17.

Fried, Mirjam and Jan-Ola Östman (2004). Construction Grammar: A Thumbnail Sketch. In Fried, Mirjam and Jan-Ola Östman, editors, *Construction Grammar in a Cross-Language Perspective*, pages 11–86. John Benjamins, Amsterdam.

Fu, Feng, Corina E. Tarnita, Nicholas A. Christakis, Long Wang, David G. Rand, and Martin A. Nowak (2012). Evolution of In-Group Favoritism. *Scientific Reports*, 2: 460.

Fugère, Vincent, Hernán Ortega, and Rüdiger Krahe (2011). Electrical Signalling of Dominance in a Wild Population of Electric Fish. *Biology Letters*, 7: 197–200.

Fuqua, W. Claiborne, Stephen C. Winans, and E. Peter Greenberg (1994). Quorum Sensing in Bacteria: the LuxR-LuxI Family of Cell Density-Responsive Transcriptional Regulators. *Journal of Bacteriology*, 176(2): 269–275.

Gallagher, H. L., F. Happé, N. Brunswick, P. C. Fletcher, U. Frith, and C. D. Frith (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, 38(1): 11–21.

Gallie, Walter Bryce (1955). Essentially Contested Concepts. *Proceedings of the Aristotelian Society*, 56: 167–198.

Gallistel, C. Randy (1990). *The Organization of Learning*. MIT Press, Cambridge, MA.

Gallistel, C. Randy and Rochel Gelman (1992). Preverbal and Verbal Counting and Computation. *Cognition*, 44(1–2): 43–74.

Gambello, M. J. and B. H. Iglewski (1991). Cloning and Characterisation of the *Pseudomonas aeruginosa lasR* Gene, a Transcriptional Activator of Elastase Expression. *Journal of Bacteriology*, 173: 3000–3009.

Gardner, Andy and Stuart A. West (2010). Greenbeards. *Evolution*, 64(1): 25–38.

Garner, Richard Lynch (1892). *The Speech of Monkeys*. William Heinemann, London.

Garrett, Merrill F. (1975). The Analysis of Sentence Production. In Bower, G., editor, *Psychology of Learning and Motivation*, volume 9, pages 505–529. Academic Press, New York.

Garrett, Merrill F. (1982). Production of Speech: Observations from Normal and Pathological Language Use. In Ellis, A. W., editor, *Normality and Pathology in Cognitive Functions*, pages 19–76. Academic Press, London.

Gaunt, Alexander L., Marc Brockschmidt, Nate Kushman, and Daniel Tarlow (2016). Differentiable Programs with Neural Libraries. In *Proceedings of the 34th International Conference on Machine Learning*. arXiv:1611.02109.

Gauvain, Mary (1993). The Development of Spatial Thinking in Everyday Activity. *Development Review*, 13(1): 92–121.

Gelman, Rochel (2009). Learning in Core and Noncore Domains. In Tommasi, L., M. A. Peterson, and L. Nadel, editors, *Cognitive biology*, pages 247–260. MIT Press, Cambridge.

Gentner, Dedre (1983). Structure Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7: 155–170.

Gentner, Dedre, Brian F. Bowdle, Phillip Wolff, and Consuelo Boronat (2001). Metaphor is like analogy. In Gentner, Dedre, Keith J. Holyoak, and Boicho N. Kokinov, editors, *The Analogical Mind: Perspectives from Cognitive Science*. MIT press, Cambridge, MA.

Gentner, Dedre and Donald R. Gentner (1983). Flowing waters or teeming crowds: Mental models of electricity. In Gentner, Dedre and Albert L. Stevens, editors, *Mental Models*. Lawrence Erlbaum Associates Inc., Hillsdale, NJ.

Gentner, Dedre and Cecile Toupin (1986). Systematicity and Surface Similarity in the Development of Analogy. *Cognitive Science*, 10(3): 277–300.

Gera, Charu and S. Srivastava (2006). Quorum-Sensing: the Phenomenon of Microbial Communication. *Current Science*, 90: 666–677.

Gick, Mary L. and Keith J. Holyoak (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology*, 15: 1–38.

Gil, David (2012). Where Does Predication Come From? *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 57(2): 303–333.

Gilbert, Margaret (1989). *On Social Facts*. Princeton University Press, Princeton.

Gillan, Douglas J. (1981). Reasoning in the Chimpanzee II: Transitive Inference. *Journal of Experimental Psychology: Animal Behavior Processes*, 7: 150–164.

Givón, Talmy (1978). Negation in Language: Pragmatics, Function, Ontology. In Cole, P., editor, *Syntax and Semantics 9: Pragmatics*, volume 53, pages 69–112. Academic Press, New York.

Givón, Talmy (1979). *On understanding grammar*. Academic Press, New York.

Givón, Talmy (1995). *Functionalism and Grammar*. John Benjamins Publishing Company, Amsterdam.

Givón, Talmy (2002a). *Bio-Linguistics: The Santa Barbara Lectures*. John Benjamins Publishing Company, Amsterdam.

Givón, Talmy (2002b). The Visual Information-Processing System as an Evolutionary Precursor to Human Language. In Givón, Talmy and Bertram F. Malle, editors, *The Evolution of Language Out of Pre-Language*, volume 53 of *Typological Studies in Language*, pages 3–50. John Benjamins, Amsterdam.

Givón, Talmy (2009). *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neurocognition, Evolution*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Glaser, Robert (1990). Reemergence of Learning Theory Within Instructional Research. *American Psychologist*, 45(1): 29–39.

Godfrey-Smith, Peter (1989). Misinformation. *Canadian Journal of Philosophy*, 19: 533–550.

Godfrey-Smith, Peter (2011). Signals: Evolution, Learning, and Information by Brian Skyrms (Review). *Mind*, 120(480): 1288–1297.

Godfrey-Smith, Peter (2018). Primates, Cephalopods, and the Evolution of Communication. In Seyfarth, Robert M., Dorothy L. Cheney, and Michael L. Platt, editors, *The Social Origins of Language*, pages 102–120. Princeton University Press, Princeton.

Godfrey-Smith, Peter and Manolo Martínez (2013). Communication and Common Interest. *PLoS Computational Biology*, 9(11): 1–6.

Goel, V., J. Grafman, N. Sadato, and M. Hallett (1995). Modeling other minds. *NeuroReport*, 6(13): 1741–1746.

Goffman, Erving (1974). *Frame Analysis: An Essay on the Organization of Experience*. Harvard University Press, Cambridge, MA.

Goldberg, Yoav (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, 10(1): 1–309.

Gong, Yichen, Heng Luo, and Jian Zhang (2017). Natural Language Inference over Interaction Space. In *Proceedings of the 2018 International Conference on Learning Representations*.

Goodall, Jane (1968). The Behaviour of Free-Living Chimpanzees in the Gombe Stream Reserve, Tanzania. *Animal Behaviour Monographs*, 1: 161–311.

Goodall, Jane (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Harvard University Press, Cambridge, MA.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Goodman, Nelson (1965). *Fact, Fiction, and Forecast*. The Bobs-Merrill Company, Inc., London.

Gordon, Shira D. and George W. Uetz (2011). Multimodal Communication of Wolf Spiders on Different Substrates: Evidence for Behavioural Plasticity. *Animal Behaviour*, 81: 367–375.

Gould, James L. (1975). Honey Bee Recruitment: The Dance-Language Controversy. *Science*, 189(4204): 685–693.

Gould, Stephen Jay and Richard C. Lewontin (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 205(1161): 581–598.

Goulden, R., P. Nation, and J. Read (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics*, 11(4): 341–363.

Grafen, Alan (1990). Biological Signals as Handicaps. *Journal of Theoretical Biology*, 144: 517–546.

Graziano, Michael S. A. (2006). The organization of behavioral repertoire in motor cortex. *Annual Review of Neuroscience*, 29(1): 105–134.

Graziano, Michael S. A. (2010). *God, Soul, Mind, Brain: A Neuroscientist's Reflections on the Spirit World*. Leapfrog Press, Teaticket MA.

Graziano, Michael S. A. (2013). *Consciousness and the Social Brain*. Oxford University Press, Oxford.

Graziano, Michael S. A. (2019). *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. W. W. Norton & Co., New York.

Graziano, Michael S. A. and T. N. Aflalo (2007). Mapping behavioral repertoire onto the cortex. *Neuron*, 56(2): 239–251.

Graziano, Michael S. A., T. N. Aflalo, and D. F. Cooke (2005). Arm movements evoked by electrical stimulation in the motor cortex of monkeys. *Journal of Neurophysiology*, 94(6): 4209–4223.

Graziano, Michael S. A., D. F. Cooke, and C. S. R. Taylor (2000). Coding the location of the arm by sight. *Science*, 290(5497): 1782–1786.

Graziano, Michael S. A. and C. G. Gross (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, 97(1): 96–109.

Graziano, Michael S. A., X. T. Hu, and C. G. Gross (1997a). Coding the locations of objects in the dark. *Science*, 277(5323): 239–241.

Graziano, Michael S. A., X. T. Hu, and C. G. Gross (1997b). Visuo-spatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77(5): 2268–2292.

Graziano, Michael S. A. and S. Kastner (2011). Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cognitive Neuroscience*, 2(2): 98–113.

Graziano, Michael S. A., L. A. J. Reiss, and C. G. Gross (1999). A neuronal representation of the location of nearby sounds. *Nature*, 397(6718): 428–430.

Graziano, Michael S. A., G. S. Yap, and C. G. Gross (1994). Coding of visual space by pre-motor neurons. *Science*, 266(5187): 1054–1057.

Green, Steven and Peter Marler (1979). The analysis of animal communication. In Marler, P. and G. Vandenbergh, editors, *Handbook of Behavioral Neurobiology*, volume 3: Social Behavior and Communication, page 73–158. Plenum, New York.

Greenberg, E. P., J. W. Hastings, and S. Ulitzer (1979). Induction of Luciferase Synthesis in *Beneckea harveyi* by Other Marine Bacteria. *Archives of Microbiology*, 120: 87–91.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H., editor, *Universals of Language*, pages 73–113. MIT Press, London.

Greeno, James G. (1989a). A Perspective on Thinking. *American Psychologist*, 44(2): 134–141.

Greeno, James G. (1989b). Situations, mental models, and generative knowledge. In Klahr, D. and K. Kotovsky, editors, *Complex Information Processing: The Impact of Herbert A. Simon*, pages 285–318. Erlbaum Associates, Hillsdale, NJ.

Grice, H. Paul (1975). Logic and conversation. In Cole, Peter and Jerry L. Morgan, editors, *Syntax and Semantics, Vol. 3: Speech Acts*, pages 41–58. Academic Press, New York.

Griffin, Donald (1992). *Animal Minds: Beyond Cognition to Consciousness*. University of Chicago Press, Chicago.

Griffiths, D. P., A. Dickinson, and N. S. Clayton (1999). Episodic Memory: What Can Animals Remember About Their Past. *Trends in Cognitive Sciences*, 3(2): 74–80.

Grossman, Pamela L., Peter Smagorinsky, and Sheila Valencia (1999). Appropriating Tools for Teaching English: A Theoretical Framework for Research on Learning to Teach. *American Journal of Education*, 108(1): 1–29.

Haddock, Steven H. D., Mark A. Moline, and James F. Case (2010). Bioluminescence in the Sea. *Annual Review of Marine Science*, 2: 443–493.

Hadeler, K. P. (1981). Stable Polymorphisms in a Selection Model with Mutation. *SIAM Journal of Applied Mathematics*, 41: 1–7.

Hailman, J., M. Ficken, and R. Ficken (1985). The 'Chick-a-dee' calls of *Parus atricapillus*. *Semiotica*, 56: 191–224.

Hailman, Jack P. (1977). *Optical Signals: Animal Communication and Light*. Indiana University Press, Oxford.

Haldane, John Burdon Sanderson (1927). A Mathematical Theory of Natural and Artificial Selection. Part V. Selection and Mutation. *Proceedings of the Cambridge Philosophical Society*, 23: 838–844.

Haldane, John Burdon Sanderson (1955). Aristotle's Account of Bees' 'Dances'. *The Journal of Hellenic Studies*, 75: 24–25.

Haldane, J. B. S. and H. Spurway (1954). A statistical analysis of communication in 'Apis mellifera' and a comparison with communication in other animals. *Insectes Sociaux*, 1(3): 247–283.

Hall, K. and G. B. Schaller (1964). Tool Using Behavior of the California Sea Otter. *Journal of Mammalogy*, 45: 287–298.

Hall Jr., Robert A. (1966). *Pidgin and Creole Languages*. Cornell University Press, Ithaca.

Halliday, Tim (1983). Information and communication. In Halliday, T. and P. J. B. Slater, editors, *Communication: Animal Behaviour, Vol. 2*, page 43–81. Blackwell Scientific, Oxford.

Hamilton, William D. (1967). Extraordinary Sex Ratios. *Science*, 156(3774): 477–488.

Harder, John D. and Leslie M. Jackson (2010). Chemical communication and reproduction in the gray short-tailed opossum (*Monodelphis domestica*). *Vitamins and Hormones*, 83: 373–399.

Harms, William F. (2004a). *Information and Meaning in Evolutionary Processes*. Cambridge University Press, Cambridge.

Harms, William F. (2004b). Primitive content, translation, and the emergence of meaning in animal communication. In Oller, D. Kimbrough and Ulrike Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 31–48. MIT Press, Cambridge, MA.

Hauser, Marc D. (1996). *The Evolution of Communication*. MIT Press, Cambridge, MA.

Hauser, Marc D. (1998). Functional Referents and Acoustic Similarity: Field Playback Experiments with Rhesus Monkeys. *Animal Behaviour*, 55: 1647–1658.

Hauser, Marc D. (2000). *Wild Minds: What Animals Really Think*. Henry Holt, New York.

Hauser, Marc D., Susan Carey, and Lilan B. Hauser (2000). Spontaneous Number Representation in Semi-Free-Ranging Rhesus Monkeys. *Proceedings of the Royal Society B: Biological Sciences*, 267(1445): 829–833.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298: 1569–1579.

Hauser, Marc D. and W. Tecumseh Fitch (2003). What are the uniquely human components of the language faculty? In Christiansen, M. H. and S. Kirby, editors, *Language Evolution*, pages 158–181. Oxford University Press, Oxford.

Hauser, Marc D., Pogen MacNeilage, and Molly Ware (1996). Numerical Representations in Primates. *Proceedings of the National Academy of Sciences*, 93: 1514–1517.

Hauser, Marc D. and Peter Marler (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*). I. Socioecological factors influencing call production. *Behavioral Ecology*, 4: 194–205.

Healy, S. D. and J. Suhonen (1996). Memory for Location of Stored Food in Willow Tits and Marsh Tits. *Behaviour*, 133(1–2): 71–80.

Hebets, Eileen A. and Daniel R. Papaj (2005). Complex Signal Function: Developing a Framework of Testable Hypotheses. *Behavioral Ecology and Sociobiology*, 57: 197–214.

Heine, Bernd and Tania Kuteva (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press, Oxford.

Herder, Johann Gottfried (1966/1772). *Essay on the Origin of Language [Über den Ursprung der Sprache]*. Verlag Freies Geistesleben, Stuttgart.

Herman, Louis M. and Paul H. Forestell (1985). Reporting Presence or Absence of Named Objects by a Language-Trained Dolphin. *Neuroscience and Biobehavioral Reviews*, 9: 667–681.

Herrmann, Esther and Josep Call (2012). Are There Geniuses Among the Apes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603): 2753–2761.

Herrnstein, Richard J. (1970). On the Law of Effect. *Journal of Experimental Analysis of Behavior*, 13: 243–266.

Hesse, Mary B. (1966). *Models and Analogies in Science*. University of Notre Dame Press, Notre Dame, IN.

Hewes, Gordon W. (1973). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, 14: 5–24.

Hinde, Robert A. (1981). Animal Signals: Ethological and Games-Theory Approaches are Not Incompatible. *Animal Behavior*, 29: 535–542.

Hines, W. G. S. (1987). Evolutionary Stable Strategies: A Review of Basic Theory. *Theoretical Population Biology*, 31: 195–272.

Hockett, Charles F. (1958). *A Course in Modern Linguistics*. Macmillan, New York.

Hockett, Charles F. (1959). Animal Languages and Human Language. *Americal Institute of Biological Sciences*, 31(1): 32–39.

Hockett, Charles F. (1960a). Logical Considerations in the Study of Animal Communication. *Americal Institute of Biological Sciences*, pages 392–430.

Hockett, Charles F. (1960b). The Origin of Speech. *Scientific American*, 203: 88–111.

Hockett, Charles F. (1963). The problem of universals in language. In Greenberg, J., editor, *Universals of Language*, pages 1–29. MIT Press, Cambridge, MA.

Hockett, C. F. and S. A. Altmann (1968). A note on design features. In Sebeok, T. A., editor, *Animal Communication*, pages 574–575. Indiana University Press, Bloomington.

Hockett, Charles F. and Robert Ascher (1964). The Human Revolution. *Current Anthropology*, 5: 135–147.

Hodges, Wilfrid (2012). Formalizing the Relationship Between Meaning and Syntax. In Hinzen, Wolfram, Edouard Machery, and Markus Werning, editors, *The Oxford Handbook of Compositionality*, pages 245–261. Oxford University Press, Oxford.

Hodges, Wilfrid and Thomas Scanlon (2018). First-order model theory. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Hofbauer, Josef (1985). The Selection-Mutation Equation. *Journal of Mathematical Biology*, 23: 41–53.

Hofbauer, Josef and Simon Huttegger (2008). The Feasibility of Communication in Binary Signaling Games. *Journal of Theoretical Biology*, 254: 843–849.

Hofbauer, Josef, Peter Schuster, and Karl Sigmund (1979). A Note on Evolutionary Stable Strategies and Game Dynamics. *Journal of Theoretical Biology*, 81: 609–612.

Hofbauer, Josef and Karl Sigmund (1988). *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection*. Cambridge University Press, Cambridge.

Hofbauer, Josef and Karl Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.

Holm, Håkan (2000). Gender-Based Focal Points. *Games and Economic Behavior*, 32(2): 292–314.

Holyoak, Keith J. and Paul Thagard (1989). Analogical Mapping by Constraint Satisfaction. *Cognitive Science*, 13: 295–355.

Holyoak, Keith J. and Paul Thagard (1995). *Mental Leaps: Analogy in Creative Thought*. MIT Press, Cambridge, MA.

Homans, George C. (1961). *Social Behavior: Its Elementary Forms*. Harcourt Brace and World, New York.

Hopcraft, J. Grant C., Juan Manuel Morales, H. L. Beyer, Markus Borner, Ephraim Mwangomo, A. R. E. Sinclair, Han Olff, and Daniel T. Haydon (2014). Competition, Predation, and Migration: Individual Choice Patterns of Serengeti Migrants Captured by Hierarchical Models. *Ecological Monographs*, 84(3): 355–372.

Hopkins, Ed and Martin Posch (2005). Attainability of Boundary Points Under Reinforcement Learning. *Games and Economic Behavior*, 53: 110–125.

Hoppe, F. M (1984). Pólya-Like Urns and the Ewans Sampling Formula. *Journal of Mathematical Biology*, 20: 91–94.

Hopper, Paul J. (1991). On some principles of grammaticalization. In Traugott, E. C. and B. Heine, editors, *Approaches to Grammaticalization*, pages 17–35. John Benjamins, Amsterdam.

Hopper, Paul J. and Elizabeth Closs Traugott (2003). *Grammaticalization*. Cambridge University Press, Cambridge, 2 edition.

Horn, Andrew G. and Peter K. McGregor (2013). Influence and Information in Communication Networks. In Stegmann, Ulrich E., editor, *Animal Communication Theory: Information and Influence*, pages 43–61. Cambridge University Press, Cambridge.

Horn, Larry (1989). *A Natural History of Negation*. University of Chicago Press, Chicago.

Horn, Laurence R. and Heinrich Wansing (2017). Negation. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition.

Hornstein, Norbert (2008). *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge.

Houck, Lynne D. (2009). Pheromone Communication in Amphibians and Reptiles. *Annual Review of Physiology*, 71: 161–176.

Hrdy, Sarah Blaffer (2005). Comes the Child Before Man: How Cooperative Breeding and Prolonged Postweaning Dependence Shaped Human Potentials. In Hewlett, B. and M. Lamb, editors, *Hunter-Gatherer Childhoods*, pages 65–91. Aldine Transaction, London.

Hu, Ronghang, Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Kate Saenko (2017). End-to-End Module Networks for Visual Question Answering. In *Proceedings of 2017 IEEE International Conference on Computer Vision*. arXiv:1704.05526.

Hu, Yilei, Brian Skyrms, and Pierre Tarrès (2011). Reinforcement Learning in Signaling Game. *arXiv preprint arXiv:1103.5818*.

Humboldt, Wilhelm Von (1836). *Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Druckerei der Königlichen Akademie der Wissenschaften, Berlin.

Hume, David (1739). *A Treatise of Human Nature*. John Noon, London.

Hume, David (1777/1748). An Enquiry Concerning Human Understanding. In Selby-Bigge, L. A., editor, *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, pages 5–165. Clarendon Press, London.

Humphries, Mark D. and Kevin Gurney (2008). Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLOS ONE*, 3(4): e0002051.

Hung, Woei (2013). Problem-Based Learning: A Learning Environment for Enhancing Learning Transfer. *New Directions for Adult and Continuing Education*, 137: 27–38.

Hunt, Gavin R. and Russell D. Gray (2003). Diversification and Cumulative Evolution in New Caledonian Crow Tool Manufacture. *Proceedings of the Royal Society London, B*, 270: 867–874.

Hunt, Gavin R. and Russell D. Gray (2004a). Direct Observations of Pandanus-Tool Manufacture and Use by a New Caledonian Crow (Corvus moneduloides). *Animal Cognition*, 7: 114–120.

Hunt, Gavin R. and Russell D. Gray (2004b). The Crafting of Hook Tools by Wild New Caledonian Crows. *Proceedings of the Royal Society London, B*, 271(Suppl. 3): S88–90.

Hunter III, Maxwell W. and Alan C. Kamil (1971). Object-discrimination learning set and hypothesis behavior in the northern bluejay (*Cynaocitta cristata*). *Psychonomic Science*, 22(5): 271–273.

Hurford, James R. (1987). *Language and Number: The Emergence of a Cognitive System*. Basil Blackwell, Oxford.

Hurford, James R. (2007). *Language in the Light of Evolution I: The Origins of Meaning*. Oxford University Press, Oxford.

Hurford, James R. (2009). Universals and the Diachronic Life Cycle of Languages. In Christiansen, M., C. Collins, and S. Edelman, editors, *Language Universals*, pages 40–53. Oxford University Press, Oxford.

Hurford, James R. (2012). *Language in the Light of Evolution II: The Origins of Grammar*. Oxford University Press, Oxford.

Huttegger, Simon M. (2007a). Evolution and the Explanation of Meaning. *Philosophy of Science*, 74: 1–27.

Huttegger, Simon M. (2007b). Evolutionary Explanations of Indicatives and Imperatives. *Erkenntnis*, 66: 409–436.

Huttegger, Simon M., Brian Skyrms, Rory Smead, and Kevin J. S. Zollman (2010). Evolutionary Dynamics of Lewis Signaling Games: Signaling Systems vs. Partial Pooling. *Synthese*, 172(1): 177–191.

Huttegger, Simon M. and Kevin Zollman (2010). Dynamic Stability and Basins of Attraction in the Sir Philip Sidney Game. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1689): 1915–1922.

Huxley, Julian (1966). A Discussion on Ritualization of Behaviour in Animals and Man: Introduction. *Philosophical Transactions of the Royal Society of London*, 251: 249–271.

Isoni, Andrea, Anders Poulsen, Robert Sugden, and Kei Tsutsui (2013). Focal Points in Tacit Bargaining Problems: Experimental Evidence. *European Economic Review*, 59: 167–188.

Jackendoff, Ray S. (1999). Possible Stages in the Evolution of the Language Capacity. *Trends in Cognitive Sciences*, 3: 272–279.

Jackendoff, Ray S. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press, Oxford.

Jackendoff, Ray S. (2007). *Language, Consciousness, Culture: Essays on Mental Structure.* MIT Press, Cambridge, MA.

Jackendoff, Ray S. (2010). Your Theory of Language Evolution Depends Upon Your Theory of Language. In Larson, Richard K., Viviane Déprez, and Hiroko Yamakido, editors, *The Evolution of Human Language: Biolinguistic Perspectives*, pages 63–72. Cambridge University Press, Cambridge.

Jäger, Gerhard (2008). Applications of Game Theory in Linguistics. *Language and Linguistics Compass*, 2(3): 406–421.

Janssen, Theo M. V. (2012). Compositionality: Its Historic Context. In Hinzen, Wolfram, Edouard Machery, and Markus Werning, editors, *The Oxford Handbook of Compositionality*, pages 19–46. Oxford University Press, Oxford.

Jennions, Michael, Tamás Székely, Steven R. Beissinger, and Peter M. Kappeler (2017). Sex Ratios. *Current Biology*, 27(16): R790–R792.

Jespersen, Otto (1917). *Negation in English and Other Languages.* Høst, Copenhagen.

Jespersen, Otto (1922). *Language: Its Nature, Development and Origin.* W. W. Norton & Co., New York.

Jiang, Yu, Vivek Natarajan, Xinlei Chen, Marcus Rohrbach, Dhruv Batra, and Devi Parikh (2018). Pythia v0.1: The winning entry to the vqa challenge 2018. https://github.com/facebookresearch/pythia.

Johnson, Thompson, Smagorinsky, and Fry (2003). Learning to Teach the Five-Paragraph Theme. *Research in the Teaching of English*, 38(2): 136–176.

Johnson, Justin, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick (2016). Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* arXiv:1612.06890.

Johnson, Justin, Bharath Hariharan, Laurens van der Maaten, Judy Hoffman, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick (2017). Inferring and executing programs for visual reasoning. In *Proceedings of 2017 IEEE International Conference on Computer Vision.*

Johnstone, Rufus A. (1995). Sexual Selection, Honest Advertisement and the Handicap Principle: Reviewing the Evidence. *Biological Reviews*, 7: 1–65.

Jolly, Alison (1966). Lemur Social Behavior and Primate Intelligence. *Science*, 153: 501–506.

Jones, Lynette A. (2000). Kinesthetic sensing. In *Proceedings of Workshop on Human and Machine Haptics*, pages 1–10. MIT Press, Cambridge, MA.

Jürgens, Uwe (2002). Neural Pathways Underlying Vocal Control. *Neuroscience & Biobehavioral Reviews*, 26: 235–258.

Just, Marcel Adam and Patricia Ann Carpenter (1971). Comprehension of Negation with Quantification. *Journal of Verbal Learning and Verbal Behavior*, 12: 21–31.

Kafri, Ran, Melissa Levy, and Yitzhak Pilpel (2006). The regulatory utilization of genetic redundancy through responsive backup circuits. *Proceedings of the National Academy of Sciences, USA*, 103(31): 11653–11658.

Kaminski, Juliane, Josep Call, and Julia Fischer (2004). Word Learning in a Domestic Dog: Evidence for 'Fast Mapping'. *Science*, 304: 1682–1683.

Kamp, Hans and Barbara Partee (1995). Prototype Theory and Compositionality. *Cognition*, 57: 129–191.

Kannan, Anjuli, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, and Vivek Ramavajjala (2016). Smart Reply: Automated Response Suggestion for Email. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 955–964, New York. ACM.

Kappeler, Peter M. (1998). To Whom it May Concern: The Transmission and Function of Chemical Signals in Lemur catta. *Behavioral Ecology and Sociobiology*, 42: 411–421.

Karakashian, S. J., M. Gyger, and P. Marler (1988). Audience Effects on Alarm Calling in Chickens (*Gallus gallus*). *Journal of Comparative Psychology*, 102(2): 129–135.

Karnath, H. O., S. Ferber, and M. Himmelbach (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, 411(6840): 950–953.

Katahira, Kentaro, Kazuo Okanoya, and Masato Okada (2007). A Neural Network Model for Generating Complex Birdsong Syntax. *Biological Cybernetics*, 97: 441–448.

Katz, J. S., A. A. Wright, and J. Bachevalier (2002). Mechanisms of Same/Different Abstract-Concept Learning by Rhesus Monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, 28(4): 358–368.

Kavanagh, Michael (1978). Monkeys' New Life in the Forest. *New Scientist*, 77(109): 515–517.

Kegl, Judy (2002). Language Emergence in a Language-Ready Brain: Acquisition numbers. In Morgan, G. and B. Woll, editors, *Language Acquisition in Signed Languages*, pages 207–254. Cambridge University Press, Cambridge.

Keller, Laurent and Kenneth G. Ross (1998). Selfish Genes: A Green Beard in the Red Fire Ant. *Nature*, 394: 573–575.

Kempner, E. S. and F. E. Hanson (1968). Aspects of light production by *Photobacterium fischeri*. *Journal of Bacteriology*, 95: 975–979.

Kenward, Ben, Alex A. S. Weir, Christian Rutz, and Alex Kacelnik (2005). Tool Manufacture by Naive Juvenile Crows. *Nature*, 433: 121.

Keynes, John Maynard (1921). *A Treatise on Probability*. Macmillan, London.

Kirby, Simon, Hannah Cornish, and Kenny Smith (2008). Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language. *PNAS*, 105(31): 10681–10686.

Klima, Edward S. and Ursula Bellugi (1979). *The Signs of Language*. Harvard University Press, Cambridge, MA.

Koechlin, Etienne, Stanislas Dehaene, and Jacques Mehler (1998). Numerical Transformations in Five-Month-Old Human Infants. *Mathematical Cognition*, 3(2): 89–104.

Köhler, Wolfgang (1927). *The Mentality of Apes*. Kegan Paul, Trench, Trubner and Co., London.

Kottur, Satwik, José M. F. Moura, Stefan Lee, and Dhruv Batra (2017). Natural language does not emerge 'naturally' in multi-agent dialog. *CoRR*, abs/1706.08502.

Krebs, John R. and Nicholas B. Davies (1993). *An Introduction to Behavioural Ecology*. Blackwell Scientific Publications, Oxford.

Krebs, John R. and Richard Dawkins (1984). Animal Signals: Mind Reading and Manipulation. In Krebs, J. R. and N. B. Davies, editors, *Behavioural Ecology*, pages 380–402. Sinauer Associates, Sunderland, MA.

Kulahci, Ipek G., Anna Dornhaus, and Daniel R. Papaj (2008). Multimodal Signals Enhance Decision Making in Foraging Bumble-Bees. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 275: 797–802.

Kuran, Timur (1988). The Tenacious Past: Theories of Personal and Collective Conservatism. *Journal of Economic Behavior and Organization*, 10: 143–171.

Kutsukake, Suetsugu, and Hasegawa (2006). Pattern, Distribution, and Function of Greeting Behavior Among Black-and-White Colobus. *International Journal of Primatology*, 27(5): 1271–1291.

Lachmann, Michael and Carl T. Bergstrom (2004). The Disadvantage of Combinatorial Communication. *Proceedings of the Royal Society of London B*, 271: 2337–2343.

LaCroix, Travis (2018). On Salience and Signaling in Sender-Receiver Games: Partial Pooling, Learning, and Focal Points. *Synthese*. Forthcoming.

LaCroix, Travis (2019a). Biology and compositionality: Empirical considerations for emergent-communication protocols. *Emergent Communication Workshop at NeurIPS 2019. (2019 Conference on Neural Information Processing Systems)*. `arXiv.org/abs/1911.11668`.

LaCroix, Travis (2019b). Evolutionary Explanations of Simple Communication: Signalling Games and Their Models. *Journal for General Philosophy of Science / Zeitschrift für allgemeine Wissenschaftstheorie*. Forthcoming.

LaCroix, Travis (2019c). Using logic to evolve more logic: Composing logical operators via self-assembly. *British Journal for the Philosophy of Science*. Forthcoming.

LaCroix, Travis (2020a). Communicative bottlenecks lead to maximal information transfer. *Journal of Experimental and Theoretical Artificial Intelligence*. Forthcoming.

LaCroix, Travis (2020b). Saltationist versus gradualist approaches to language origins: A critical discussion. Unpublished Manuscript. March, 2020. PDF File.

LaCroix, Travis, Michael Noukhovitch, and Aaron Courville (2020). Cooperative communication under conflict of interest. Unpublished Manuscript. February, 2020. PDF File.

Ladd, D. Robert (2012). What is duality of patterning, anyway? *Language and Cognition*, 4(4): 261–273.

Laitman, Jeffrey Todd and Joy S. Reidenberg (1988). Advances in Understanding the Relationship Between the Skull Base and Larynx with Comments on the Origins of Speech. *Journal of Human Evolution*, 3: 99–109.

Lake, Brenden M. and Marco Baroni (2017). Generalization Without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks. arXiv:1711.00350.

Landau, Barbara, Elizabeth Spelke, and Henry Gleitman (1984). Spatial Knowledge in a Young Blind Child. *Cognition*, 16(3): 225–260.

Latifi, Amel, Michael K. Winson, Maryline Foglino, Barrie W. Bycroft, Gordon S. A. B. Stewart, and Paul Williams (1995). Multiple Homologues of LuxR and LuxI Control Expression of Virulence Determinants and Secondary Metabolites Through Quorum Sensing in *Pseudomonas aeruginosa* PAO1. *Molecular Microbiology*, 17(2): 333–343.

Lave, Jean (1990). The Culture of Acquisition and the Practice of Understanding. In Stigler, J. W., R. A. Shweder, and G. Herdt, editors, *Cultural psychology*, pages 259–286. Cambridge University Press, Cambridge.

Lazareva, Olga F. and Edward A. Wasserman (2010). Nonverbal Transitive Inference: Effects of Task and Awareness on Human Performance. *Behavioral Processes*, 83(1): 99–112.

Lea, Amanda J., June P. Barrera, Lauren M. Tom, and Daniel T. Blumstein (2008). Heterospecific Eavesdropping in a Nonsocial Species. *Behavioral Ecology*, 19(5): 1041–1046.

Lea, Stephen E.G., Alan M. Slater, and Catriona M. E. Ryan (1996). Perception of Object Unity in Chicks: A Comparison with the Human Infant. *Infant Behavior and Development*, 19(4): 501–504.

Leonard, R. (1994). Reading Cournot, Reading Nash: The Creation and Stabilisation of the Nash Equilibrium. *Economic Journal*, 104: 492–511.

Leontyev, Aleksei Nikolaevich (1981). *Problems of the Development of the Mind*. Progress Publishers, Moscow.

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

Levelt, W. J. M. (1992). Accessing Words in Speech Production: Stages, Processes and Representations. *Cognition*, 42: 1–22.

Lewis, David (1967). *Conventions of Language*. PhD thesis, Harvard University.

Lewis, David (1970). General Semantics. *Synthese*, 22(1–2): 18–67.

Lewis, David (1975). Languages and Language. In Gunderson, Keith, editor, *Language, Mind, and Knowledge*, pages 3–35. University of Minnesota Press, Minneapolis.

Lewis, David (2002/1969). *Convention*. Blackwell, Oxford.

Lewontin, Richard C. (1961). Evolution and the Theory of Games. *Journal of Theoretical Biology*, 1: 382–403.

Liberman, Alvin M. (1996). *Speech: A Special Code*. MIT Press, Cambridge, MA.

Lieberman, Philip (1984). *The Biology and Evolution of Language*. Harvard University Press, Cambridge, MA.

Lieberman, Philip (2000). *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax and Thought*. Harvard University Press, Cambridge, MA.

Lieberman, Philip, Edmund S. Crelin, and Dennis H. Klatt (1972). Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. *American Anthropologist*, 74(3): 287–307.

Lightfoot, David (1991). Subjacency and Sex. *Language & Communication*, 11(1–2): 67–69.

Limber, J. (1977). Language in Child and Chimp? *American Psychologist*, 32: 280–295.

Lin, Francis Y. (1999). Chomsky on the 'Ordinary Language' View of Language. *Synthese*, 120(2): 151–192.

Lineburg, Bruce (1924). Communication by Scent in the Honeybee: A theory. *American Naturalist*, 58: 530–537.

Lishak, Robert S. (1984). Alarm Vocalizations of Adult Gray Squirrels. *Journal of Mammalogy*, 65: 681–684.

Livingstone, Frank B. (1973). Did the Australopithecines Sing? *Current Anthropology*, 14(1/2): 25–29.

Livingstone, Margaret S, Warren W. Pettine, Krishna Srihasam, Brandon Moore, Istvan A. Morocz, and Daeyeol Lee (2014). Symbol Addition by Monkeys Provides Evidence for Normalized Quantity Coding. *Proceedings of the National Academy of Science*, 111: 6822–6827.

Locke, John L. and Barry Bogin (2006). Language and Life History: A New Perspective on the Development and Evolution of Human Language. *Behavioral & Brain Sciences*, 29: 259–280.

Loewenstein, J. and D. Gentner (2005). Relational Language and the Development of Relational Mapping. *Cognitive Psychology*, 50: 315–353.

Lorenz, Konrad (1965). *Evolution and Modification of Behavior*. University of Chicago Press, Chicago.

Lorenz, Konrad (1966). Evolution of Ritualization in the Biological and Cultural Spheres. *Philosophical Transactions of the Royal Society of London B*, 251: 273–84.

Lovejoy, C. Owen (1981). The Origin of Man. *Science*, 211: 341–350.

Lubbock, John (1874). *Ants, Bees and Wasps*. Kegan Paul, Trench, London.

Luce, R. Duncan and Howard Raiffa (1957). *Games and Decisions*. John Wiley & Sons, Inc., New York.

Lugli, Marco, Hong Y. Yan, and Michael L. Fine (2003). Acoustic Communication in Two Freshwater Gobies: the Relationship Between Ambient Noise, Hearing Thresholds and Sound Spectrum. *Journal of Comparative Physiology A*, 189: 309–320.

Lusseau, David, Ben Wilson, Philip S. Hammond, Kate Grellier, John W. Durban, Kim M. Parsons, Tim R. Barton, and Paul M. Thompson (2006). Quantifying the influence of sociality on population structure in bottlenose dolphins. *Journal of Animal Ecology*, 75: 14–24.

MacArthur, Robert H. (1965). Ecological Consequences of Natural Selection. In Waterman, T. and H. Horowitz, editors, *Theoretical and Mathematical Biology*, pages 388–397. Blaisdell, New York.

Macedonia, Joseph M. (1993). The Vocal Repertoire of the Ringtailed Lemur (*Lemur catta*). *Folia Primatologica*, 61(4): 186–217.

Macedonia, Joseph M. and Christopher S. Evans (1993). Variation Among Mammalian Alarm Call Systems and the Problem of Meaning in Animal Signals. *Ethology*, 93: 177–197.

Macy, M. (1991). Learning to Cooperate: Stochastic and Tacit Collusion in Financial Exchange. *American Journal of Sociology*, 97: 808–843.

Macy, M. and A. Flache (2002). Learning Dynamics in Social Dilemmas. *Proceedings of the National Academy of Sciences of the USA*, 99: 7229–7236.

Madden, Joah R., Julian A. Drewe, Gareth P. Pearce, and Tim H. Clutton-Brock (2011). The social network structure of a wild meerkat population: 3. Position of individuals within networks. *Behavioral Ecology and Sociobiology*, 65(10): 1857–1871.

Maeterlink, Maurice (1901). *La Vie des Abeilles*. Charpentier, Paris.

Magrath, Robert D., Benjamin J. Pitcher, and Janet L. Gardner (2007). A Mutual Understanding? Interspecific Responses by Birds to Each Other's Aerial Alarm Calls. *Behavioral Ecology*, 18(5): 944–951.

Majeed, Raamy (2016). The Hard Problem & Its Explanatory Targets. *Ratio*, 29(3): 298–311.

Mandeville, Bernard (1997/1723). *The Fable of the Bees and Other Writings*. Hackett, Cambridge.

March, James G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, 10(1): 299–316.

Marcus, Gary F. (1998). Rethinking Eliminative Connectionism. *Cognitive Psychology*, 37(3): 243–282.

Marcus, Gary F. (2003). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. MIT press, Cambridge, MA.

Margolis, Eric and Stephen Laurence (2008). How to Learn the Natural Numbers: Inductive Inference and the Acquisition of Number Concepts. *Cognition*, 106(2): 924–939.

Margolis, Eric and Stephen Laurence (2011). Beyond the Building Blocks Model. *Behavioral and Brain Sciences*, 34(3): 139–140.

Markl, Hubert (1985). Manipulation, Modulation, Information, Cognition: Some of the Riddles of Communication. In Hölldobler, B. and M. Lindauer, editors, *Experimental Behavioral Ecology and Sociobiology*, pages 163–194. G. Fischer Verlag, Stuttgart.

Marler, Peter (1961). The Logical Analysis of Animal Communication. *Journal of Theoretical Biology*, 7: 295–317.

Marler, Peter (1991). The Instinct to Learn. In Carey, S. and R. Gelman, editors, *The Epigenesis of Mind: Essays on Biology and Cognition*, pages 37–66. Lawrence Erlbaum Associates, Hillsdale.

Marler, Peter (1998). Animal Communication and Human Language. In Jablonski, N. G. and L. C. Aiello, editors, *The Origin and Diversification of Language*, pages 1–19. California Academy of Sciences, San Francisco.

Marler, Peter, Christopher S. Evans, and Marc D. Hauser (1992). Animal Signals? Motivational, Referential, or Both? In Papoucek, H., U. Jürgens, and M. Papoucek, editors, *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pages 66–86. Cambridge University Press, Cambridge.

Marler, Peter, Stephen Karakashian, and Marcel Gyger (1991). Do Animals Have the Option of Withholding Signals When Communication is Inappropriate? The Audience Effect. In Ristau, C., editor, *Cognitive Ethology: The Minds of Other Animals*, pages 135–186. Lawrence Erlbaum Associates, Hillsdale, NJ.

Marler, Peter and Hans Slabbekoorn (2004). *Nature's Music*. Elsevier Academic Press, Amsterdam.

Martínez, Manolo (2015). Deception in Sender-Receiver Games. *Erkenntnis*, 80(1): 215–227.

Martínez, Manolo and Peter Godfrey-Smith (2016). Common Interest and Signaling Games: A Dynamic Analysis. *Philosophy of Science*, 83(3): 371–392.

Masuda, Naoki and Feng Fu (2015). Evolutionary Models of In-Group Favoritism. *F1000Prime Reports*, 3: 7.

Mäthger, Lydia M., Eric J. Denton, N. Justin Marshall, and ROger T. Hanlon (2009). Mechanisms and Behavioural Functions of Structural Coloration in Cephalopods. *Journal of the Royal Society Interface*, 6: S149–S163.

Matsuzawa, Tetsuro (2009). Symbolic Representation of Number in Chimpanzees. *Current Opinion Neurobiology*, 19: 92–98.

Matzel, Louis D. and Stefan Kolata (2010). Selective Attention, Working Memory, and Animal Intelligence. *Neuroscience & Biobehavioral Reviews*, 34: 23–30.

Maurer, Daphne (1993). Neonatal Synesthesia: Implications for the Processing of Speech and Faces. In de Boysson-Bardies, B., S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton, editors, *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, pages 109–124. Kluwer, Dordrecht.

Maurer, Daphne and Catherine J. Mondlach (2005). Neonatal Synaesthesia: A Reevaluation. In Robertson, L. C. and S. Sagiv, editors, *Synaesthesia: Perspectives from Cognitive Neuroscience*, pages 193–213. Oxford University Press, New York.

Maynard Smith, John (1978). Optimization Theory in Evolution. *Annual Review of Ecology & Systematics*, 9: 31–56.

Maynard Smith, John (1979). Game Theory and the Evolution of Behaviour. *Proceedings of the Royal Society, London, B*, 205: 475–488.

Maynard Smith, John (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Maynard Smith, John (1991). Honest Signalling: The Philip Sidney Game. *Animal Behaviour*, 42: 1034–1035.

Maynard Smith, John and David Harper (2003). *Animal Signals*. Oxford University Press, Oxford.

Maynard Smith, John and George M. Price (1973). The Logic of Animal Conflict. *Nature*, 246: 15–18.

Maynard Smith, John and Eörs Szathmáry (1995). *The Major Transitions in Evolution*. Oxford University Press, New York.

McBrearty, Sally and Alison S. Brooks (2000). The Revolution that Wasn't: A New Interpretation of the Origin of Modern Human Behaviour. *Journal of Human Evolution*, 39(5): 453–563.

McComb, Karen and Stuart Semple (2005). Coevolution of Vocal Communication and Sociality in Primates. *Biology Letters*, 1(4): 381–385.

McCowan, Brenda, Laurence R. Doyle, and Sean F. Hanser (2002). Using Information Theory to Assess the Diversity, Complexity, and Development of Communicative Repertoires. *Journal of Comparative Psychology*, 116: 166–172.

McGonigle, Brendan O. and Margaret Chalmers (1977). Are Monkeys Logical? *Nature*, 267: 694–696.

McGrew, William Clement (1992). *Chimpanzee Material Culture*. Cambridge University Press, Cambridge.

McNamara, John M. and Sasha R. X. Dall (2010). Information is a Fitness Enhancing Resource. *Oikos*, 119(2): 231–236.

McNeill, David (2005). *Gesture and Thought*. The University of Chicago Press, Chicago.

Mehler, Jacques, Marina Nespor, Mohinish Shukla, and Marcela Peña (2006). Why is Language Unique to Humans? *Novartis Foundation Symposium*, 270: 251–280.

Mehta, Judith, Chris Starmer, and Robert Sugden (1984a). Focal Points in Pure Coordination Games: An Experimental Investigation. *Theory and Decision*, 36: 163–185.

Mehta, Judith, Chris Starmer, and Robert Sugden (1984b). The Nature of Salience: An Experimental Investigation of Pure Coordination Games. *American Economic Review*, 84: 658–673.

Mellars, Paul (2006). Going East: New Genetic and Archaeological Perspectives on the Modern Human Colonization of Eurasia. *Science*, 313: 796–800.

Merker, Björn (2000). Synchronous Chorusing and Human Origins. In Wallin, N. L., B. Merker, and S. Brown, editors, *The Origins of Music*, pages 315–327. MIT Press, Cambridge, MA.

Miestamo, Matti (2007). Negation: An Overview of Typological Research. *Language and Linguistics Compass*, 1(5): 552–570.

Mill, John Stewart (1930/1843). *A System of Logic*. Longmans-Green, London.

Miller, Geoffrey (2001). *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. Doubleday, New York.

Miller, Melissa B. and Bonnie L. Bassler (2001). Quorum Sensing in Bacteria. *Annual Review of Microbiology*, 55(1): 165–199.

Millikan, Ruth Garrett (1995). Pushmi-Pullyu Representations. *Philosophical Perspectives*, 9: 185–200.

Millikan, Ruth Garrett (2005). *Language: A Biological Model*. Oxford University Press, Oxford.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Weidenfeld and Nicolson, London.

Miyagawa, Shigeru (2017). Integration Hypothesis: A Parallel Model of Language Development in Evolution. In Watanabe, S., M. Hofman, and T. Shimizu, editors, *Evolution of the Brain, Cognition, and Emotion in Vertebrates*, Brain Science, pages 225–247. Springer, Tokyo.

Miyagawa, Shigeru, Shiro Ojima, Robert C. Berwick, and Kazuo Okanoya (2015). The Integration Hypothesis of Human Language Evolution and the Nature of Contemporary Languages. *Frontiers in psychology*, 39: 564.

Moore, Brooke Noel and Richard Parker (1998). *Critical Thinking*. Mayfield, Mountain View, CA, 5 edition.

Morgan, Morris Hicky (1967). *Vitruvius: The Ten Books on Architecture, Book II*. Harvard University Press, Cambridge, MA.

Moro, Andrea (2008). *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. The MIT Press, Cambridge, MA.

Mufwene, Salikoko S. (2001). *The Ecology of Language Evolution*. Cambridge University Press, New York.

Mühlenbernd, Ronald (2011). Learning with neighbours: Emergence of Convention in a Society of Learning Agents. *Synthese*, 183: 87–109.

Mühlhäusler, Peter (1997). *Pidgin and Creole Linguistics*. University of Westminster Press, London.

Müller, Friedrich Max (1864). Lecture ix: The theoretical stage in the science of language, and origin of language. In *Lectures on the Science of Language*, pages 356–410. Longman, Green, Longman, Roberts, & Green, London, 4 edition.

Müller, Friedrich Max (1873). Lectures on Mr Darwin's Philosophy of Language. *Fraser's Magazine*, 7–8: 147–233.

Muszynski, Eric (2015). Récursion, Saltation, mais sans Communication? Critique de la Théorie Chomskyenne de l'Évolution du Langage. Master's thesis, Université du Québec à Montréal, Montréal, Québec.

Myers, Ronald E. (1976). Comparative Neurology of Vocalization and Speech: Proof of a Dichotomy. *Annals of the New York Academy of Science*, 280: 745–757.

Myers, Shirley A., James A. Horel, and Henry S. Pennypacker (1965). Operant Control of Vocal Behavior in the Monkey Cebus albifrons. *Psychonomic Science*, 3: 389–390.

Nagel, Rosemarie (1995). Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, 85(5): 1313–1326.

Nash, John (1950a). Equilibrium Points in n-Person Games. *Econometrica*, 18: 155–162.

Nash, John (1950b). *Non-Cooperative Games*. PhD thesis, Princeton University.

Nash, John (1950c). The Bargaining Problem. *Proceedings of the National Academy of Sciences USA*, 36: 48–49.

Nash, John (1951). Non-Cooperative Games. *The Annals of Mathematics*, 54(2): 286–295.

Nash, John F. (1996). *Essays on Game Theory*. Edward Elgar Publishing Ltd., Cheltenham, UK.

Nealson, Kenneth H., Terry Platt, and J. Woodland Hastings (1970). Cellular Control of the Synthesis and Activity of the Bacterial Luminescent System. *Journal of Bacteriology*, 104(1): 313–322.

Neumann, John Von and Oskar Morgenstern (2007/1944). *Theory of Games and Economic Behavior*. Princeton university press.

Nevins, A., D. M. Pesetsky, and C. Rodrigues (2009). Evidence and argumentation: A Reply to Everett. *Language*, 85(3): 671–681.

Newmeyer, Frederick J. (1991). Functional Explanation in Linguistics and the Origin of Language. *Language and Communication*, 11: 1–28.

Newmeyer, Frederick J. (1998). On the supposed 'counterfunctionality' of Universal Grammar: Some evolutionary implications. In Hurford, James R., Michael Studdert-Kennedy, and Chris Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 305–319. Cambridge University Press, Cambridge.

Newmeyer, Frederick J. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford University Press, Oxford.

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. University of Chicago Press, Chicago.

Noukhovitch, Michael, Travis LaCroix, and Aaron Courville (2020). Emergent communication under conflict of interest. Unpublished Manuscript. February, 2020. PDF File.

Nowak, Martin A. and David C. Krakauer (1999). The Evolution of Language. *Proceedings of the National Academy of Sciences*, 96: 8028–8033.

Nowak, Martin A., Plotkin, and Jansen (2000). The Evolution of Syntactic Communication. *Nature*, 404: 495–498.

Ochsner, Urs A., Andreas K. Koch, Armin Fiechter, and Jakob Reiser (1994). Isolation and Characterization of a Regulatory Gene Affecting Rhamnolipid Biosurfactant Synthesis in *Pseudomonas aeruginosa*. *Journal of bacteriology*, 176: 2044–2054.

O'Connor, Cailin (2014). The Evolution of Vagueness. *Erkenntnis*, 79(4): 707–727.

Ogilvie, Ryan and Peter Carruthers (2016). Opening Up Vision: The Case Against Encapsulation. *Review of Philosophy and Psychology*, 7(4): 721–742.

Orenstein, Ronald I. (1972). Tool-Use by the New Caledonian Crow (Corvus moneduloides). *Auk*, 89: 674–676.

Otte, Daniel (1974). Effects and Functions in the Evolution of Signaling Systems. *Annual Reviews of Ecology and Systematics*, 5: 385–417.

Ouattara, Karim, Alban Lemasson, and Klaus Zuberbühler (2009). Campbell's Monkeys Concatenate Vocalizations into Context-Specific Call Sequences. *Proceeding of the National Acadaemy of Sciences USA*, 106: 22026–22031.

Owings, Donald H. and Eugene S. Morton (1998). *Animal Vocal Communication: A New Approach*. Cambridge University Press, Cambridge.

Owren, Michael J., Jacquelyn A. Dieter, Robert M. Seyfarth, and Dorothy L. Cheney (1993). Vocalizations of Rhesus (Macaca mulatta) and Japanese (M. fuscata) Macaques Cross-Fostered Between Species Show Evidence of Only Limited Modification. *Developmental Psychobiology*, 26: 389–406.

Owren, Michael J., Drew Rendall, and Michael J. Ryan (2010). Redefining Animal Signaling: Influence Versus Information in Communication. *Biology and Philosophy*, 25: 755–780.

Pagin, P. and D. Westerståhl (2010a). Compositionality I: Definitions and Variants. *Philosophy Compass*, 5(3): 250–264.

Pagin, P. and D. Westerståhl (2010b). Compositionality II: Arguments and Problems. *Philosophy Compass*, 5(3): 265–282.

Pan, Sinno Jialin and Qiang Yang (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22: 1345–1359.

Parravano, Melanie and Odile Poulsen (2015). Stake Size and the Power of Focal Points in Coordination Games: Experimental Evidence. *Games and Economic Behavior*, 94: 191–199.

Partee, Barbara Hall (1984). Compositionality. In Landman, F. and F. Veltman, editors, *Varieties of Formal Semantics*, page 281–311. Foris, Dordrecht.

Patterson, Francine (1978). The Gestures of a Gorilla: Language Acquisition in Another Pongid. *Brain and Language*, 5: 72–97.

Pawlowitsch, Christina (2008). Why Evolution Does Not Always Lead to an Optimal Signaling System. *Games and Economic Behavior*, 63(1): 203–226.

Payne, John (1985). Negation. In Shopen, T., editor, *Language Typology and Syntactic Description I: Clause Structure*, pages 197–242. Cambridge University Press, Cambridge.

Paz-y-Miño, Guillermo, Alan B. Bond, and Russell P. Balda (2004). Pinyon Jays Use Transitive Inference to Predict Social Dominance. *Nature*, 430: 778–781.

Pearce, Andrew C., Yotis A. Senis, Daniel D. Billadeau, Martin Turner, Steve P. Watson, and Elena Vigorito (2004). Vav1 and Vav3 Have Critical but Redundant Roles in Mediating Platelet Activation by Collagen. *The Journal of Biological Chemistry*, 279(52): 53955–53962.

Pearson, James P., Kendall M. Gray, Luciano Passador, Kenneth D. Tucker, Anatol Eberhard, Barbara H. Iglewski, and E. P. Greenberg (1994). Structure of the Autoinducer Required for Expression of *Pseudomonas aeruginosa* Virulence Genes. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 197–201.

Pelissier, Catherine (1991). The Anthropology of Teaching and Learning. *Annual Review of Anthropology*, 20: 75–95.

Pemantle, Robin (2007). A Survey of Random Processes With Reinforcement. *Probability Surveys*, 4: 1–79.

Pemantle, Robin and Brian Skyrms (2004). Network Formation by Reinforcement Learning: the Long and the Medium Run. *Mathematical Social Sciences*, 48: 315–327.

Penn, Derek C., Keith J. Holyoak, and Daniel J. Povinelli (2008). Darwin's Mistake: Explaining the Discontinuity Between Human and Nonhuman Minds. *Behavioral and Brain Sciences*, 31(2): 109–130. Discussion, 130–178.

Penn, Derek C. and Daniel J. Povinelli (2007a). Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. *Annual Review of Psychology*, 58: 97–118.

Penn, Derek C. and Daniel J. Povinelli (2007b). On the Lack of Evidence that Non-Human Animals Possess Anything Remotely Resembling a 'Theory of Mind'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480): 731–744.

Pepperberg, Irene Maxine (1990). Conceptual Abilities of Some Nonprimate Species, with an Emphasis on an African Grey Parrot. In Parker, S. T. and K. R. Gibson, editors, *"Language" and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives*, pages 469–507. Cambridge University Press, New York.

Pepperberg, Irene Maxine (1999). *The Alex studies*. Harvard University Press, Cambridge, MA.

Piaget, Jean (1950). *The Psychology of Intelligence*. Routledge and Kegan Paul, London.

Piaget, Jean (1968). *Structuralism*. Routledge and Kegan Paul, London.

Piaget, Jean (1976). *Biology and Knowledge*. University of Chicago Press, Chicago.

Piaget, Jean (1980). *Adaptation and Intelligence*. University of Chicago Press, Chicago.

Piantadosi, Steven T., Joshua B. Tenenbaum, and Noah D. Goodman (2012). Bootstrapping in a Language of Thought: A Formal Model of Numerical Concept Learning. *Cognition*, 123(2): 199–217.

Piattelli-Palmarini, Massimo (1989). Evolution, Selection, and Cognition: From "Learning" to Parameter Setting in Biology and in the Study of Language. *Cognition*, 31: 1–44.

Piattelli-Palmarini, Massimo (2010). What Is Language, That It May Have Evolved, and What Is Evolution, That It May Apply to Language? In Larson, Richard K., Viviane Déprez, and Hiroko Yamakido, editors, *The Evolution of Human Language: Biolinguistic Perspectives*, pages 148–162. Cambridge University Press, Cambridge.

Piattelli-Palmarini, Massimo and Juan Uriagereka (2004). The Immune Syntax: The Evolution of the Language Virus. In Jenkins, Lyle, editor, *Variation and Universals in Biolinguistics*, volume 62 of *Linguistic Variations*, chapter 14, pages 341–377. Elsevier, Oxford.

Piattelli-Palmarini, Massimo and Juan Uriagereka (2011). A Geneticist's Dream, a Linguist's Nightmare: The Case of FOXP2. In Sciullo, Anna Maria Di and Cedric Boeckx, editors, *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, Oxford Studies in Biolinguistics, chapter 5, pages 100–125. Oxford University Press, Oxford.

Piccinini and Scarantino (2011). Information Processing, Computation and Cognition. *Journal of Biological Physics*, 37: 1–38.

Pilley, John W. and Alliston K. Reid (2011). Border Collie Comprehends Object Names as Verbal Referents. *Behavioural Processes*, 86(2): 184–195.

Pinker, Steven (1994). *The Language Instinct*. William Morrow and Company, New York.

Pinker, Steven (2002). *The Blank Slate: The Modern Denial of Human Nature*. Viking, New York.

Pinker, Steven and Paul Bloom (1990). Natural Language and Natural Selection. *Behavioral and Brain Sciences*, 13: 707–726.

Pinker, Steven and Ray Jackendoff (2005). The Faculty of Language: What's Special About It? *Cognition*, 95: 201–236.

Pitman, Jim (1995). Exchangeable and Partially Exchangeable Random Partitions. *Probability Theory Related Fields*, 102(2): 145–158.

Plato (1921a). Cratylus. In *Plato in Twelve Volumes*, volume 12. Harvard University Press, Cambridge, MA.

Plato (1921b). Sophist. In *Plato in Twelve Volumes*, volume 7. Harvard University Press, Cambridge, MA.

Platt, Michael L. (2018). Introduction. In Seyfarth, Robert M., Dorothy L. Cheney, and Michael L. Platt, editors, *The Social Origins of Language*, pages 1–6. Princeton University Press, Princeton.

Ploeger, Annemie and Frietson Galis (2011). Evo Devo and Cognitive Science. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(4): 429–440.

Poleman, Joseph L. (2006). Mastery and Appropriation as Means to Understand the Interplay of History Learning and Identity Trajectories. *Proceedings of the Royal Society London, B*, 15(2): 221–259.

Pomiankowski, Andrew (1987). Sexual Selection: The Handicap Principle Does Work – Sometimes. *Proceedings of the Royal Society London, B*, 231: 123–145.

Pope, Denise S. and Brian R. Haney (2008). Interspecific Signalling Competition Between Two Hood-Building Fiddler Crab Species, *Uca latimanus* and *U. musica musica*. *Animal Behaviour*, 76: 2037–2048.

Post, Emil (1943). Formal Reductions of the General Combinatorial Decision Problem. *American Journal of Mathematics*, 65: 197–215.

Povinelli, Daniel J. (2000). *Folk Physics for Apes*. Oxford University Press, Oxford.

Prasnikar, Vesna and Alvin Roth (1992). Considerations of Fairness and Strategy: Experimental Data from Sequential Games. *The Quarterly Journal of Economics*, 107(3): 865–888.

Pratt, Lorien Y. (1993). Discriminability-Based Transfer Between Neural Networks. In *NIPS Conference: Advances in Neural Information Processing Systems 5*, pages 204–211.

Pratt, Stephen C. (2005). Quorum Sensing by Encounter Rates in the Ant *Temnothorax albipennis*. *Behavioral Ecology*, 16(2): 488–496.

Prawat, Richard S. and Robert E. Floden (1994). Philosophic Perspectives on Constructivist Views of Learning. *Educational Psychologist*, 29(1): 37–48.

Premack, David (1971). Language in Chimpanzee? *Science*, 172: 808–822.

Premack, David (1986). *'Gavagai!' or the Future History of the Animal Language Controversy*. MIT Press, Cambridge, MA.

Premack, David and Ann James Premack (1983). *The Mind of the Ape*. Norton, New York.

Progovac, Ljiljana (2006). The Syntax of Nonsententials: Small Clauses and Phrases at the Root. In Progovac, Ljiljana, Kate Paesani, Eugenia Casielles-Suarez, and Ellen Barton, editors, *The Syntax of Nonsententials: Multidisciplinary Perspectives*, volume 93 of *Linguistik Aktuell/Linguistics Today*, pages 33–71. John Benjamins, Amsterdam.

Progovac, Ljiljana (2009a). Layering of Grammar: Vestiges of Protosyntax in Present-Day Languages. In Sampsom, Geoffrey, David Gil, and Peter Trudgill, editors, *Language Complexity as an Evolving Variable*, Studies in the Evolution of Language. Oxford University Press, New York.

Progovac, Ljiljana (2009b). Sex and Syntax: Subjacency Revisited. *Biolinguistics*, 3(2–3): 305–336.

Progovac, Ljiljana (2013). Rigid Syntax, Rigid Sense: Absolutives/Unaccusatives as Evolutionary Precursors. In Franks, Steven, Markus Dickinson, George Fowler, Melissa Witcombe, and Ksenia Zanon, editors, *Proceedings of Formal Approaches to Slavic Linguistics (FASL), The Third Indiana Meeting, Bloomington, IN.*, pages 246–259. Michigan Slavic Publications, Ann Arbor.

Progovac, Ljiljana (2015). *Evolutionary Syntax*. Oxford University Press, Oxford.

Progovac, Ljiljana (2019). *A Critical Introduction to Language Evolution: Current Controversies and Future Prospects*. Springer, Berlin.

Pugh, Kevin J. and David A. Bergin (2006). Motivational Influences on Transfer. *Educational Psychologist*, 41(3): 147–160.

Pullum, Geoffrey K. and Gerald Gazdar (1982). Natural Languages and Context-Free Languages. *Linguistics and Philosophy*, 4(4): 471–504.

Pylyshyn, Zenon (1984). *Computation and Cognition*. MIT Press, Cambridge, MA.

Pylyshyn, Zenon (1999). Is Vision Continuous with Cognition? The Case for Cognitive Penetrability of Vision. *Behavioral and Brain Sciences*, 22(3): 341–423.

Quastler, Henry (1956). A primer on information theory. In Yockey, H. P., R. L. Platzman, and H. Quastler, editors, *Symposium on Information Theory in Biology. Gatlinburg, TN*, pages 3–49, London. Pergamon Press.

Quine, Willard van Orman (1960). *Word and Object*. MIT Press, Cambridge, MA.

Quine, Willard van Orman (1967). Truth by Convention. In *Philosophical Essays for Alfred North Whitehead*, pages 90–124. Russell & Russell, New York.

Rabin, Lawrence A., Brenda McCowan, Stacie L. Hooper, and Donald H. Owings (2003). Anthropogenic Noise and its Effect on Animal Communication: An Interface Between Comparative Psychology and Conservation Biology. *International Journal of Comparative Psychology*, 16: 172–192.

Raguso, Robert A. (2008). Wake Up and Smell the Roses: The Ecology and Evolution of Floral Scent. *Annual Review of Ecology, Evolution, and Systematics*, 39: 549–569.

Randolph, M. C. and B. B. Brooks (1967). Conditioning of a Vocal Response in a Chimpanzee Through Social Reinforcement. *Folia Primatologica*, 5: 70–79.

Reader, Simon M., Yfke Hager, and Kevin N. Laland (2011). The Evolution of Primate General and Cultural Intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567): 1017–1027.

Reby, David and Karen McComb (2003). Anatomical Constraints Generate Honesty: Acoustic Cues to Age and Weight in the Roars of Red Deer Stags. *Animal Behavior*, 65: 519–530.

Reby, David, Karen McComb, Bruno Cargnelutti, Chris Darwin, W. Tecumseh Fitch, and Tim Clutton-Brock (2005). Red Deer Stags Use Formants as Assessment Cues During Intrasexual Agonistic Interactions. *Procedings of the Royal Society London, B*, 272: 941–947.

Regolin, Lucia and Giorgio Vallortigara (1995). Perception of Partly Occluded Objects by Young Chicks. *Perception and Psychophysics*, 57: 971–976.

Regolin, Lucia, Giorgio Vallortigara, and Mario Zanforlin (1995). Object and Spatial Representations in Detour Problems by Chicks. *Animal Behaviour*, 49: 195–199.

Rendall, Drew, John R. Vokey, and Christina Ney (2005). Reliable but Weak Voice-Formant Cues to Body Size in Men but Not Women. *Journal of the Acoustical Society of America*, 117: 2372.

Rey, Georges (2014). Innate and Learned: Carey, Mad Dog Nativism, and the Poverty of Stimuli and Analogies (Yet Again). *Mind & Language*, 29(2): 109–132.

Richardson, Michael K. and Gerhard Keuck (2002). Haeckel's ABC of evolution and development. *Biological Reviews*, 77: 495–528.

Richman, Bruce (1993). On the Evolution of Speech: Singing as the Middle Term. *Current Anthropology*, 34: 721–722.

Ridley, Mark (1993). *Evolution*. Blackwell Scientific, Oxford.

Riede, Tobias and W. Tecumseh Fitch (1999). Vocal Tract Length and Acoustics of Vocalization in the Domestic Dog Canis familiaris. *Journal of Experimental Biology*, 202: 2859–2867.

Rips, Lance J., Jennifer Asmuth, and Amber Bloomfield (2006). Giving the Boot to the Bootstrap: How Not to Learn the Natural Numbers. *Cognition*, 101(3): B51–B60.

Rips, Lance J., Jennifer Asmuth, and Amber Bloomfield (2008). Do Children Learn the Integers by Induction? *Cognition*, 106(2): 940–951.

Rips, Lance J., Jennifer Asmuth, and Amber Bloomfield (2013). Can Statistical Learning Bootstrap the Integers? *Cognition*, 128(3): 320–330.

Rips, Lance J. and Susan J. Hespos (2011). Rebooting the Bootstrap Argument: Two Puzzles for Bootstrap Theories of Concept Development. *Behavioral and Brain Sciences*, 34(3): 145–146.

Rissanen, Jorma (1978). Modeling by Shortest Data Description. *Automatica*, 14: 465–471.

Rissanen, Jorma (1989). *Stochastic Complexity in Statistical Enquiry*. World Scientific Publishing, Singapore.

Ristau, C. A. and D. Robbins (1982). Language in the Great Apes: A Critical Review. *Advances in the Study of Behavior*, 12: 141–255.

Rizzolatti, Giacomo and Michael A. Arbib (1998). Language Within our Grasp. *Trends in Neuroscience*, 21: 188–194.

Rizzolatti, Giacomo, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi (1996). Premotor Cortex and the Recognition of Motor Actions. *Cognitive Brain Research*, 3(2): 131–141.

Robbins, Herbert (1952). Some Aspects of the Sequential Design of Experiments. *Bulletin of the American Mathematical Society*, 58: 527–535.

Robbins, Philip (2017). Modularity of mind. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition.

Robinson, Daniel-Sommer (1930). *The Principles of Reasoning*. D. Appleton, New York, 2 edition.

Rogoff, Barbara (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. Oxford University Press, New York.

Rogoff, Barbara (1993). Children's Guided Participation and Participatory Appropriation in Sociocultural Activity. In Woxniak, R. and K. Fischer, editors, *Development in Context: Acting and Thinking in Specific Environments*, pages 121–153. Erlbaum, Hillsdale, NJ.

Rogoff, Barbara (1995). Observing Sociocultural Activity on Three Planes: Participatory Appropriation, Guided Participation, and Apprenticeship. In Wertsch, J. V., P. del Rio, and A. Alvarez, editors, *Sociocultural Studies of Mind*, pages 139–164. Cambridge University Press, Cambridge. Reprinted (2008) in K. Hall & P. Murphy (Eds.), Pedagogy and practice: Culture and identities. London: Sage.

Rose, Gary J., Franz Goller, Howard J. Gritton, Stephanie L. Plamondon, Alexander T. Baugh, and Brenton G. Cooper (2004). Species-Typical Songs in White-Crowned Sparrows Tutored with Only Phrase Pairs. *Nature*, 432: 753–758.

Rose, Tania A., Adam J. Munn, Daniel Ramp, and Peter B. Banks (2006). Foot-Thumping as an Alarm Signal in Macropodoid Marsupials: Prevalence and Hypotheses of Function. *Mammal Review*, 36: 281–298.

Rosenthal, Gil G. (2007). Spatiotemporal Dimensions of Visual Signals in Animal Communication. *Annual Review of Ecology, Evolution, and Systematics*, 38: 155–178.

Roth, Alvin and Ido Erev (1995). Learning in Extensive Form Games: Experimental Data and Simple Dynamical Models in the Intermediate Term. *Games and Economic Behavior*, 8: 164–212.

Roth, Alvin and Ido Erev (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88(4): 848–881.

Roth, Alvin, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study. *American Economic Review*, 81: 1068–1095.

Rousseau, Jean-Jacques (1984/1755). *A Discourse on Inequality*. Penguin.

Rubin, Hannah, Justin Bruner, Cailin O'Connor, and Simon M. Huttegger (2016). Communication Without the Cooperative Principle: A Signaling Experiment. Unpublished Manuscript. August, 2018. PDF File.

Rumbaugh, Duane M. (1970). Learning Skills of Anthropoids. In Rosenblum, L., editor, *Primate Behavior: Developments in Field and Laboratory Research*, pages 2–70. Aldine, New York.

Rumbaugh, Duane M. (1971). Evidence of Qualitative Differences in Learning Processes Among Primates. *Journal of Comparative and Physiological Psychology*, 76(2): 250–255.

Rumbaugh, Duane M. (1995). Primate Language and Cognition: Common Ground. *Social Research*, 62(3): 711–730.

Rumbaugh, Duane M. and James L. Pate (1984a). Primates' Learning by Levels. In Greenberg, G. and E. Tobach, editors, *Behavioral Evolution and Integrative Levels*, pages 221–240. Lawrence Erlbaum Associates, Hillsdale, NJ.

Rumbaugh, Duane M. and James L. Pate (1984b). The Evolution of Cognition in Primates: A Comparative Perspective. In Roitblat, H., T. G. Bever, and H. S. Terrace, editors, *Animal Cognition*, pages 569–587. Lawrence Erlbaum Associates, Hillsdale, NJ.

Russell, Bertrand (1922). *Analysis of Mind*. The MacMillan Company, New York.

Ryan, Michael J. and A. Stanley Rand (1993). Sexual Selection and Signal Evolution: The Ghosts of Biases Past. *Philosophical Transactions of the Royal Society of London B*, 340(1292): 187–295.

Sandholm, W. H., E. Dokumaci, and F. Franchetti (2012). Dynamo: Diagrams for evolutionary game dynamics. http://www.ssc.wisc.edu/ whs/dynamo.

Sapir, Edward (1921). *Language: An Introduction to the Study of Speech*. Harcourt, Brace & Co., New York.

Sarafyazd, Morteza and Mehrdad Jazayeri (2019). Hierarchical Reasoning by Neural Circuits in the Frontal Cortex. *Science*, 364(6441): eaav8911.

Savage-Rumbaugh, E. Sue, Jeannine Murphy, Rose A. Sevcik, Karen E. Brakke, Shelly L. Williams, and Duane M. Rumbaugh (1993). Language Comprehension in Ape and Child. *Monographs of the Society for Research in Child Development*, 58: 1–221.

Savage-Rumbaugh, Sue (1986). *Ape Language: From Conditioned Response to Symbol*. Columbia University Press, New York.

Saxe, R. and N. Kanwisher (2003). People thinking about thinking people: fMRI investigations of theory of mind. *NeuroImage*, 19(4): 1835–1842.

Saxe, R. and A. Wexler (2005). Making sense of another mind: the role of the right temporoparietal junction. *Neuropsychologia*, 43(10): 1391–1399.

Scarantino, Andrea (2013). Animal Communication as Information-Mediated Influence. In Stegmann, Ulrich E., editor, *Animal Communication Theory: Information and Influence*, pages 63–87. Cambridge University Press, Cambridge.

Scarantino, Andrea and Gualtiero Piccinini (1993). Information Without Truth. *Metaphilosophy*, 41: 313–330.

Schaefer, H. Martin, Veronika Schaefer, and Douglas J. Levey (2004). How Plant–Animal Interactions Signal New Insights in Communication. *Trends in Ecology and Evolution*, 19: 577–584.

Schauder, Stephan and Bonnie L. Bassler (2001). The Languages of Bacteria. *Genes and Development*, 15: 1468–1480.

Schelling, Thomas C. (1980/1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.

Schleicher, August (1869/1863). *Darwinism Tested by the Science of Language*. John Camden Hotten, London.

Schlenker, Philippe, Emmanuel Chemla, Kate Arnold, Alban Lemasson, Karim Ouattara, Sumir Keenan, Claudia Stephan, Robin Ryder, and Klaus Zuberbühler (2014). Monkey Semantics: Two 'Dialects' of Campbell's Monkey Alarm Calls. *Linguistics and Philosophy*, 37: 439–501.

Schlimm, Dirk (2008). Two Ways of Analogy: Extending the Study of Analogies to Mathematical Domains. *Philosophy of Science*, 75: 178–200.

Schumpeter, Joseph A. (1934). *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.

Schunk, Dale H. (2004). *Learning Theories: An Educational Perspective*. Pearson, Upper Saddle River, NJ, 4 edition.

Schürch, Roger and Margaret J. Couvillon (2013). Too Much Noise on the Dance Floor: Intra- and Inter-Dance Angular Error in Honey Bee Waggle Dances. *Communicative and Integrative Biology*, 6(1): 1–3.

Schürch, Roger, Margaret J. Couvillon, Dominic D. R. Burns, Kiah Tasman, David Waxman, and Francis L. W. Ratnieks (2013). Incorporating Variability in Honey Bee Waggle Dance Decoding Improves the Mapping of Communicated Resource Locations. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 199(12): 1143–1152.

Schürch, Roger and Francis L. W. Ratnieks (2015). The Spatial Information Content of the Honey Bee Waggle Dance. *Frontiers in Ecology and Evolution*, 3(22): 1–7.

Schuster, Peter and Karl Sigmund (1983). Replicator Dynamics. *Journal of Theoretical Biology*, 100: 535–538.

Schuster, Peter and Karl Sigmund (1986). Evolutionary Game Dynamics. *Mondes en Développement*, 54–55: 229–236.

Scott-Phillips, Thomas C. and Richard A. Blythe (2013). Why is Combinatorial Communication Rare in the Natural World, and Why is Language an Exception to this Trend? *Journal of the Royal Society Interface*, 10(88): 1–7.

Searle, John (1976). The Rules of the Language Game. *Times Literary Supplement*.

Searle, John (1980). Rules and Causation. *The Behavioral and Brain Sciences*, 3: 37–39.

Searle, John (1995). *The Construction of Social Reality*. Free Press, New York.

Seddon, N., J. A. Tobias, and A. Alvarez (2002). Vocal Communication in the Pale-Winged Trumpeter (*Psophia leucoptera*): Repertoire, Context and Functional Reference. *Behaviour*, 139: 1331–1359.

Seed, Amanda and Richard Byrne (2010). Animal Tool-Use. *Current Biology*, 20(23): R1032–R1039.

Seed, Amanda, Eleanor Seddon, Bláthnaid Greene, and Josep Call (2012). Chimpanzee 'Folk Physics': Bringing Failures into Focus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603): 2743–2752.

Sellen, A. J. and D. A. Norman (1992). The Psychology of Slips. In Baars, B., editor, *Experimental Slips and Human Error: Exploring the Architecture of Volition*, pages 317–339. Plenum Press, New York.

Selten, Reinhard (1991). Anticipatory Learning in Two-Person Games. In Selten, R., editor, *Game Equilibrium Models I*, pages 98–154. Springer Verlag, Berlin.

Senghas, Ann and Marie Coppola (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science*, 12: 323–328.

Senghas, Ann, Sotaro Kita, and Asli Özyürek (2005). Children Creating Core Properties of Language: Evidence From an Emerging Sign Language in Nicaragua. *Science*, 305: 1779–1782.

Seyfarth, Robert M. and Dorothy L. Cheney (2003). Signalers and Receivers in Animal Communication. *Annual Review of Psychology*, 54: 145–173.

Seyfarth, Robert M. and Dorothy L. Cheney (2005). Constraints and Preadaptations in the Earliest Stages of Language Evolution. *Linguistic Review*, 22: 135–159.

Seyfarth, Robert M. and Dorothy L. Cheney (2018). The Social Origins of Language. In Seyfarth, Robert M., Dorothy L. Cheney, and Michael L. Platt, editors, *The Social Origins of Language*, pages 9–33. Princeton University Press, Princeton.

Seyfarth, Robert M., Dorothy L. Cheney, Thore Bergman, Julia Fischer, Klaus Zuberbühler, and Kurt Hammerschmidt (2010). The Central Importance of Information in Studies of Animal Communication. *Animal Behaviour*, 80(1): 3–8.

Seyfarth, Robert M., Dorothy L. Cheney, and Thore J. Bergman (2005). Primate Social Cognition and the Origins of Language. *Trends in Cognitive Science*, 9: 264–266.

Seyfarth, Robert M., Dorothy L. Cheney, and Peter Marler (1980a). Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science*, 210: 801–803.

Seyfarth, Robert M., Dorothy L. Cheney, and Peter Marler (1980b). Vervet Monkey Alarm Calls: Semantic Communication in a Free-Ranging Primate. *Animal Behaviour*, 28(4): 1070–1094.

Shabani, Shkelzen, Michiya Kamio, and Charles D. Derby (2009). Spiny Lobsters Use Urine-Borne Olfactory Signaling and Physical Aggressive Behaviors to Influence Social Status of Conspecifics. *Journal of Experimental Biology*, 212: 2464–2474.

Shanahan, Murray (2012). The Brain's Connective Core and its Role in Animal Cognition. *Philosophical Transactions of the Royal Society B: Biological Science*, 367(1603): 2704–2714.

Shannon, Claude (1948). A Mathematical Theory of Communication. *The Bell System Mathematical Journal*, 27: 379–423.

Shannon, Claude and Warren Weaver (1949). *The Mathematical Theory of Communication.* University of Illinois Press, Urbana and Chicago.

Shea, Nicholas (2009). New Concepts Can Be Learned. *Biology and Philosophy*, 26(1): 129–139.

Shea, Nicholas, Peter Gofrey-Smith, and Rosa Cao (2018). Content in Simple Signalling Systems. *The British Journal for the Philosophy of Science*, 69(4): 1009–1035.

Sherman, Paul W. (1977). Nepotism and the Evolution of Alarm Calls. *Science*, 197: 1246–1253.

Sherry, David F. (2006). Neuroecology. *Annual Review of Psychology*, 57: 167–197.

Sherry, David F. and Daniel L. Schacter (1987). The Evolution of Multiple Memory Systems. *Psychological Review*, 94(4): 439–454.

Shettleworth, Sara J. (2010). Clever Animals and Killjoy Explanations in Comparative Psychology. *Trends in Cognitive Sciences*, 14(11): 477–481.

Shettleworth, Sara J. (2010/1998). *Cognition, Evolution, and Behavior*. Oxford University Press, New York, 2 edition.

Shettleworth, Sara J. (2012). Modularity, Comparative Cognition and Human Uniqueness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603): 2794–2802.

Shi, R. (1996). *Perceptual Correlates of Content Words and Function Words in Early Language Input*. PhD thesis, Brown University.

Shi, R., J. Morgan, and P. Allopenna (1998). Phonological and Acoustic Bases for Earliest Grammatical Category Assignment: A Crosslinguistic Perspective. *Journal of Child Language*, 25: 169–201.

Siegel, Jeff (2008). *The Emergence of Pidgin and Creole Languages*. Oxford University Press, Oxford.

Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106: 467–482.

Simon, H. A. (1972). Complexity and the Representation of Patterned Sequences of Symbols. *Psychological Review*, 79: 369–382.

Simon, Tony J., Susan J. Hespos, and Philippe Rochat (1995). Do Infants Understand Simple Arithmetic? A Replication of Wynn (1992). *Cognitive Development*, 10(2): 253–269.

Sinervo, Barry, Alexis Chaine, Jean Clobert, Ryan Calsbeek, Lisa Hazard, Lesley Lancaster, Andrew G. McAdam, Suzanne Alonzo, Gwynne Corrigan, and Michael E. Hochberg (2006). Self-Recognition, Color Signals, and Cycles of Greenbeard Mutualism and Altruism. *Proceedings of the National Academy of Sciences of the United States of America*, 103(19): 7372–7377.

Singh, Ishtla (2000). *Pidgins and Creoles: An Introduction*. Arnold.

Skinner, Burrhus F. (1948). 'Superstition' in the Pigeon. *Journal of Experimental Psychology*, 38: 168–172.

Skinner, Burrhus F. (1950). Are Theories of Learning Necessary? *Psychological Review*, 57: 193–216.

Skyrms, Brian (2000a). Evolution of Inference. In Kohler, Tim and George Gumerman, editors, *Dynamics of Human and Primate Societies*, pages 77–88. Oxford University Press, New York.

Skyrms, Brian (2000b). Stability and Explanatory Significance of Some Simple Evolutionary Models. *Philosophy of Science*, 67(1): 94–113.

Skyrms, Brian (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge.

Skyrms, Brian (2006). Signals Presidential Address. In *Philosophy of Science Association Biennial Meeting*.

Skyrms, Brian (2010a). *Signals: Evolution, Learning, & Information*. Oxford University Press, Oxford.

Skyrms, Brian (2010b). The Flow of Information in Signaling Games. *Philosophical Studies*, 147(1): 155–165.

Skyrms, Brian (2014/1996). *Evolution of the Social Contract*. Cambridge University Press, Cambridge.

Skyrms, Brian and Jeffrey A. Barrett (2018). Propositional Content in Signals. *Studies in the History and Philosophy of Science C*. Forthcoming.

Slocombe, Katie E. and Klaus Zuberbühler (2005). Functionally referential communication in a chimpanzee. *Current Biology*, 15(19): 1779–1784.

Smith, Adam (1983/1761). Considerations concerning the first formation of languages. In Bryce, J. C., editor, *Lectures on Rhetoric and Belles Lettres*. Oxford University Press, Oxford.

Smith, M. and L. Wheeldon (1999). High level processing scope in spoken sentence production. *Cognition*, 17(3): 205–246.

Smith, W. John (1965). Message, meaning, and context in ethology. *American Naturalist*, 99: 405–409.

Smith, W. John (1997). The behavior of communicating, after twenty years. In Owings, D. H., M. D. Beecher, and N. S. Thompson, editors, *Perspectives in Ethology*, volume 12: Communication, pages 7–54. Plenum, New York.

Smolensky, Paul (1987). The constituent structure of connectionist mental states: A reply to fodor and pylyshyn. *Southern Journal of Philosophy*, 26(Supplement): 137–161.

Snowdon, Charles T. (1990). Language capacities of nonhuman animals. *Yearbook of Physical Anthropology*, 33: 215–243.

Spelke, Elizabeth S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, 21(2): 181–200.

Spelke, Elizabeth S. (2003). What Makes Us Smart? Core Knowledge and Natural Language. In Gentner, Dedre and Susan Goldin-Meadow, editors, *Language in Mind: Advances in the Investigation of Language and Thought*, pages 277–311. MIT Press, Cambridge, MA.

Spelke, Elizabeth S. and Katherine D. Kinzler (2007). Core knowledge. *Developmental Science*, 10(1): 89–96.

Spelke, Elizabeth S. and Sang Ah Lee (2012). Core systems of geometry in animal minds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603): 2784–2793.

Sperber, Dan (1994). The Modularity of Thought and the Epidemiology of Representations. In Hirschfeld, L. A. and S. A. Gelman, editors, *Mapping the Mind*, pages 39–67. Cambridge University Press, Cambridge, MA.

Sperber, Dan (2002). In Defense of Massive Modularity. In Dupoux, I., editor, *Language, Brain, and Cognitive Development*, pages 47–57. MIT Press, Cambridge, MA.

Spitzner, M. J. E. (1788). Ausführliche beschreibung der korbbienenzucht im sächsischen churkreise, ihrer dauer und ihres nutzens, ohne künstliche vermehrung nach den gründen der naturgeschichte und nach eigener langer erfahrung. Leipzig.

Stahl, Dale O. (1993). The evolution of smart$_n$ players. *Games and Economic Behavior*, 5(4): 604–617.

Stahl, Dale O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, 16(2): 303–330.

Stahl, Dale O. and Paul W. Wilson (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, 25(3): 309–327.

Stampe, Dennis W. (1977). Toward a causal theory of linguistic representation. *Midwest Studies in Philosophy*, 2: 42–63.

Stanger-Hall, Kathrin F., James E. Lloyd, and David M. Hillis (2007). Phylogeny of north american lightning bugs (coleoptera: Lampyridae): Implications for the evolution of light signals. *Molecular Phylogenetics and Evolution*, 45(1): 33–49.

Stebbing, L. Susan (1933). *A Modern Introduction to Logic*. Methuen, London, 2 edition.

Steedman, Mark (2009). Foundations of Universal Grammar in Planned Action. In Christiansen, Morten H., Chris Collins, and Shimon Edelman, editors, *Language Universals*, pages 174–199. Oxford University Press, Oxford.

Steels, Luc (1995). A Self-Organizing Spatial Vocabulary. *Artificial Life*, 2(3): 319–332.

Steels, Luc (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4): 339–356.

Stegmann, Ulrich E. (2013). Introduction: A Primer on Information and Influence in Animal Communication. In Stegmann, Ulrich E., editor, *Animal Communication Theory: Information and Influence*, pages 43–39. Cambridge University Press, Cambridge.

Steinert-Threlkeld, Shane (2014). Learning to use function words in signaling games. In Lorini, Emiliano and Laurent Perrussel, editors, *Proceedings of Information Dynamics in Artificial Societies*, (IDAS-14).

Steinert-Threlkeld, Shane (2016). Compositional signaling in a complex world. *Journal of Logic, Language, and Information*, 25(3–4): 379–397.

Steinert-Threlkeld, Shane (2017). *Communication and Computation: New Questions About Compositionality*. PhD thesis, Stanford University.

Steinert-Threlkeld, Shane (2018). Function Words and Context Variability. PSA 2018: The 26th Biennial Meeting of the Philosophy of Science Association, 1–4 November 2018.

Steinert-Threlkeld, Shane (2020). Towards the emergence of non-trivial compositionality. *Philosophy of Science*. Forthcoming.

Stevenson, J. C. (1994). Vocational Expertise. In Stevenson, J., editor, *Cognition at Work: The Development of Vocational Expertise*, pages 7–35. National Centre for Vocational Education Research, Adelaide.

Stokoe, William C. (1960). *Sign Language Structure: An Outline of the Communicative Systems of the American Deaf*. Linstock Press, Silver Spring, MD.

Stone, Linda and Paul F. Lurquin (2007). *Genes, Culture, and Human Evolution: A Synthesis*. Blackwell Publishing, Oxford.

Strawson, P. F. (1952). *Introduction to Logical Theory*. Methuen, London.

Strawson, P. F. (1970). *Meaning and Truth*. Oxford University Press, Oxford.

Strawson, P. F. (1974). *Subject and Predicate in Logic and Grammar*. Methuen, London.

Stringer, Christopher B. and Peter Andrews (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, 239: 1263–1268.

Sugiyama, Y. and J. Koman (1979). Tool-using and tool-making behavior in wild chimpanzees at bossou, guinea. *Primates*, 20: 513–524.

Suppes, Patrick and Richard Atkinson (1960). *Markov Learning Models for Multi-person Interactions*. Stanford University Press, Stanford.

Sutton, Dennis, Charles R. Larson, Emily M. Taylor, and Robert Lindeman (1973). Vocalization in rhesus monkeys: Conditionability. *Brain Research*, 52: 225–231.

Sutton, Richard S. and Andrew G. Barto (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.

Suzuki, Toshitaka N., David Wheatcroft, and Michael Griesser (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, 7: 10986.

Szabó, Zoltán Gendler (2012). The Case for Compositionality. In Hinzen, Wolfram, Edouard Machery, and Markus Werning, editors, *The Oxford Handbook of Compositionality*, pages 64–80. Oxford University Press, Oxford.

Szabó, Zoltán Gendler (2017). Compositionality. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition.

Taga, Michiko E. and Bonnie L. Bassler (2003). Chemical Communication Among Bacteria. *Proceedings of the National Academy of Sciences of the USA*, 100(2): 14549–14554.

Tallerman, Maggie (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3): 579–604.

Tallerman, Maggie (2013a). Join the Dots: A Musical Interlude in the Evolution of Language? *Journal of Linguistics*, 49: 455–487.

Tallerman, Maggie (2013b). Kin Selection, Pedagogy, and Linguistic Complexity: Whence Protolanguage? In Botha, Rudolf and Martin Everaert, editors, *The Evolutionary Emergence of Language*, page 77–96. Oxford University Press, Oxford.

Tallerman, Maggie (2014a). Is the Syntax Rubicon More of a Mirage? A Defence of Pre-Syntactic Protolanguage. In Cartmill, Erica A., Seén Roberts, Heidi Lyn, and Hannah Cornish, editors, *The Evolution of Language: Proceedings of the 10th International Conference*, volume 10 of *Evolang*, pages 318–325. World Scientific, Vienna.

Tallerman, Maggie (2014b). The Evolutionary Origins of Syntax. In Carnie, Andrew, Yosuke Sato, and Daniel Siddiqi, editors, *The Routledge Handbook of Syntax*, pages 446–462. Routledge, London.

Taylor, Matthew E. and Peter Stone (2009). Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research*, 10: 1633–1685.

Taylor, P. and L. Jonker (1978). Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences*, 40: 145–156.

Telesford, Qawi K., Karen E. Joyce, Satoru Hayasaka, Jonathan H. Burdette, and Paul J. Laurienti (2011). The Ubiquity of Small-World Networks. *Brain Connectivity*, 1(5): 367–375.

Tempelton, C., E. Greene, and K. Davis (2005). Allometry of Alarm Calls: Black-Capped Chickadees Encode Information about Predator Size. *Science*, 308: 1934–1937.

Terman, C. Richard (1980). Social Factors Influencing Delayed Reproductive Maturation in Prairie Deermice (*Peromyscus maniculatus bairdii*) in Laboratory Populations. *Journal of Mammalogy*, 61: 219–223.

Thibaut, John W. and Harold H. Kelley (1959). *The Social Psychology of Groups.* Wiley, New York.

Thiel, Martin and Thomas Breithaupt (2011). Chemical Communication in Crustaceans: Research Challenges for the Twenty-First Century. In Breithaupt, T. and M. Thiel, editors, *Chemical Communication in Crustaceans*, pages 3–22. Springer, New York.

Thompson, R. K. R. and D. L. Oden (2000). Categorical Perception and Conceptual Judgments by Non-Human Primates: the Paleological Monkey and the Analogical Ape. *Cognitive Science: A Multidisciplinary Journal*, 24(3): 363–396.

Thorndike, Edward L. (1905). *The Elements of Psychology.* The Mason Press, Syracuse.

Thorndike, Edward L. (1911). *Animal Intelligence: Experimental Studies.* The Macmillan Company, New York.

Thorndike, Edward L. (1927). The Law of Effect. *American Journal of Psychology*, 39: 212–222.

Thorndike, Edward L. and Robert S. Woodworth (1901a). The Influence of Improvement in one Mental Function Upon the Efficiency of Other Functions (I). *Psychological Review*, 8(3): 247–261.

Thorndike, Edward L. and Robert S. Woodworth (1901b). The Influence of Improvement in one Mental Function Upon the Efficiency of Other Functions (II) The Estimation of Magnitudes. *Psychological Review*, 8(4): 384–395.

Thorndike, Edward L. and Robert S. Woodworth (1901c). The Influence of Improvement in one Mental Function Upon the Efficiency of Other Functions (III) Functions Involving Attention, Observation and Discrimination. *Psychological Review*, 8(6): 553–564.

Tinbergen, Niko (1952). 'Derived' Activities: Their Causation, Biological Significance, Origin, and Emancipation during Evolution. *Quarterly Review of Biology*, 27: 1–23.

Tinbergen, Niko (1963). On Aims and Methods of Ethology. *Zeitschrift für Tierpsychologie*, 20: 410–433.

Tobias, Phillip V. (1987). The Brain of Homo habilis: A New Level of Organization in Cerebral Evolution. *Journal of Human Evolution*, 16: 741–761.

Tomasello, Michael (1999). *The Cultural Origins of Human Cognition.* Harvard University Press, Cambridge, MA.

Tomasello, Michael (2000). Primate Cognition: Introduction to the Issue. *Cognitive Science*, 24(3): 351–361.

Tomasello, Michael and Josep Call (2007). Ape gestures and the origins of language. In Call, J. and M. Tomasello, editors, *The Gestural Communication of Apes and Monkeys*, pages 221–239. Lawrence Erlbaum, London.

Tooby, John and Leda Cosmides (1992). The Psychological Foundations of Culture. In Barkow, Jerome, Leda Cosmides, and John Tooby, editors, *The Adapted Mind*, pages 19–136. Oxford University Press, New York.

Torrey, Lisa and Jude Shavlik (2009). Transfer Learning. In Soria, E., J. Martin, R. Magdalena, M. Martinez, and A. Serrano, editors, *Handbook of Research on Machine Learning Applications*. IGI Global.

Touhara, Kazushige (2008). Sexual Communication via Peptide and Protein Pheromones. *Current Opinion in Pharmacology*, 8(6): 759–764.

Townsend, Simon W., Tobias Deschner, and Klaus Zuberbühler (2008). Female Chimpanzees Use Copulation Calls Flexibly to Prevent Social Competition. *PLOS One*, 3: e2431.

Toya, Genta and Takashi Hashimoto (2015). Computational study on evolution and adaptability of recursive operations. In *The 20th (AROB) International Symposium on Artificial Life and Robotics*, pages 68–73, Beppu, Japan.

Truppa, Valentina, Eva Piano Mortari, Duilio Garofoli, Sara Privitera, and Elisabetta Visalberghi (2011). Same/Different Concept Learning by Capuchin Monkeys in Matching-to-Sample Tasks. *PLOS ONE*, 6(8): e23809.

Tulving, E. and H. J. Markowitsch (1998). Episodic and Declarative Memory: Role of the Hippocampus. *Hippocampus*, 8: 198–204.

Turovskiy, Yevgeniy, Dimitri Kashtanov, Boris Paskhover, and Michael L. Chikindas (2007). Quorum Sensing: Fact, Fiction, and Everything in Between. *Advances in Applied Microbiology*, 62: 191–234.

Ullmann-Margalit, Edna (1977). *The Emergence of Norms*. Clarendon Press, Oxford.

Unhoch, Nikolaus (1823). Anleitung zur wahren kenntniß und zweckmäßigsten behandlung der bienen. Munich.

Vallar, G. and D. Perani (2001). The anatomy of unilateral neglect after right-hemisphere stroke lesions. a clinical/ct-scan correlation study in man. *Neuropsychologia*, 24(5): 609–622.

van der Vaart, Elske, Rineke Verbrugge, and Charlotte K. Hemelrijk (2012). Corvid Re-Caching Without 'Theory of Mind'. *PLoS ONE*, 7: e32904.

Vanderschraaf, Peter (1995). Convention as Correlated Equilibrium. *Erkenntnis*, 42: 65–87.

Vanderschraaf, Peter (1998). Knowledge, Equilibrium and Convention. *Erkenntnis*, 49: 337–369.

Vanpé, Cécile, Jean-Michel Gaillard, Petter Kjellander, Atle Mysterud, Pauline Magnien, Daniel Delorme, Guy Van Laere, François Klein, Olof Liberg, and A. J. Mark Hewison (2007). Antler Size Provides an Honest Signal of Male Phenotypic Quality in Roe Deer. *The American Naturalist*, 169(4): 481–493.

Ventura, Rafael (2017). Ambiguous Signals, Partial Beliefs, and Propositional Content. *Synthese*, pages 1–18. Forthcoming.

Virányia, Zsófia, József Topálb, Márta Gácsib, Ádá m Miklósia, and Vilmos Csányia (2004). Dogs Respond Appropriately to Cues of Humans' Attentional Focus. *Behavioural Processes*, 66(2): 161–172.

Vogeley, K., P. Bussfeld, A. Newen, S. Herrmann, F. Happé, P. Falkai, W. Maier, N.J. Shah, G.R. Fink, and K. Zilles (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage*, 14(1): 170–181.

von Frisch, Karl (1946). Die Tänze der Bienen. *Österreichische Zoologische Zeitschrift*, 1: 1–148.

von Frisch, Karl (1967). *The Dance Language and Orientation of the Bees*. Harvard University Press, Cambridge, MA.

Vygotsky, Lev S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA.

Vygotsky, Lev S. (1987). *Thought and Language*. MIT Press, Cambridge, MA.

Wacewicz, Sławomir and Przemysław Żywiczyński (2015). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemiotics*, 8: 29–46.

Wagers, M. W. and C. Phillips (2009). Multiple Dependencies and the Role of the Grammar in Real-Time Comprehension. *Journal of Linguistics*, 45(2): 395–433.

Wagner, Elliott O. (2009). Communication and Structured Correlation. *Erkenntnis*, 71(3): 377–393.

Wagner, Elliott O. (2012). Deterministic Chaos and the Evolution of Meaning. *British Journal for the Philosophy of Science*, 63: 547–575.

Wagner, Elliott O. (2014). Conventional Semantic Meaning in Signalling Games with Conflicting Interests. *British Journal for the Philosophy of Science*, 66(4): 751–773.

Wang, Wei, Ming Yan, and Chen Wu (2018). Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, volume 1: Long Papers, pages 1705–1714, Melbourne, Australia. Association for Computational Linguistics.

Wärneryd, Karl (1993). Cheap Talk, Coordination and Evolutionary Stability. *Games and Economic Behavior*, 5(4): 532–546.

Wasserman, Edward A., Michael E. Young, and Jöel Fagot (2001). Effects of Number of Items on the Baboon's Discrimination of Same from Different Visual Displays. *Animal Cognition*, 4(3-4): 163–170.

Watts, Duncan J. and Steven H. Strogatz (1998). Collective Dynamics of 'Small-World' Networks. *Nature*, 393: 440–442.

Wehner, Rüdiger and Mandyam V. Srinivasan (1981). Searching Behavior of Desert Ants, Genus *Cataglyphis* (Formicidae, Hymenoptera). *Journal of Comparative Physiology*, 142(3): 315–338.

Wei, L. J. and S. Durham (1978). The Randomized Play-the-Winner Rule in Medical Trials. *Journal of the American Statistical Association*, 73(364): 840–843.

Weibull, Jörgen W. (1994). The 'As If' Approach to Game Theory: Three Positive Results and Four Obstacles. *European Economic Review*, 38: 868–882.

Weibull, Jörgen W. (1995). The Mass-Action Interpretation of Nash Equilibrium. Working Paper Series 427, Research Institute of Industrial Economics.

Weir, Alex A. S., Jackie Chappell, and Alex Kacelnik (2004). Shaping of hooks in New Caledonian crows. *Science*, 297: 981.

West-Eberhard, Mary Jane (2003). *Developmental Plasticity and Evolution*. Oxford University Press, New York.

Wheeler, Brandon C., William A. Searcy, Morten H. Christiansen, Micahel C. Corballis, Julia Fischer, Christoph Grüter, Daniel Margoliash, Michael J. Owren, Tabitha Price, Robert Seyfarth, and Markus Wild (2011). Communication. In Menzel, Randolf and Julia Fischer, editors, *Animal Thinking: Contemporary numbers in Comparative Cognition*, pages 187–205. MIT Press, Cambridge, MA.

Whiteley, Marvin, Stephen P. Diggle, and E. Peter Greenberg (2017). Progress in and Promise of Bacterial Quorum Sensing Research. *Nature*, 551: 313–320.

Wiggins, David (1997). Languages as Social Objects. *Philosophy*, 72(282): 499–524.

Wiley, R. Haven (1983). The evolution of communication: Information and manipulation. In Halliday, T. and P. J. B. Slater, editors, *Communication: Animal Behaviour, Vol. 2*, pages 156–189. Blackwell Scientific, Oxford.

Wilkins, H. Dawn and Gary Ritchison (1999). Drumming and Tapping by Red-Bellied Woodpeckers: Description and Possible Causation. *Journal of Field Ornithology*, 70: 578–586.

Wilkins, Wendy K. and Jennie Wakefield (1995). Brain Evolution and Neurolinguistic Preconditions. *Behavioral and Brain Sciences*, 18(1): 161–226.

Wilkinson, Gerald S. and Gary N. Dodson (1997). Function and Evolution of Antlers and Eye Stalks in Flies. In Choe, J. C. and B. J. Crespi, editors, *The Evolution of Mating Systems in Insects and Arachnids*, pages 310–327. Cambridge University Press, Cambridge.

Wilson, Bundy, Nicholas John Mackintosh, and Robert A. Boakes (1985). Transfer of Relational Rules in Matching and Oddity Learning by Pigeons and Corvids. *The Quarterly Journal of Experimental Psychology*, 37(4): 313–332.

Wilson Jr., William A. (1975). Discriminative Conditioning of Vocalizations in Lemur Catta. *Behaviour*, 23: 432–436.

Winter, Peter, Patricia Handley, Detlev Ploog, and Ditmar Schott (1973). Ontogeny of Squirrel Monkey Calls Under Normal Conditions and Under Acoustic Isolation. *Behaviour*, 47: 230–239.

Wittgenstein, Ludwig (2009/1953). *Philosophical Investigations*. Wiley-Blackwell, Oxford, 4 edition.

Wolff, Phillip and Jason Shepard (2013). Causation, touch, and the perception of force. In Ross, Brian H., editor, *Psychology of Learning and Motivation*, volume 58, chapter 5, pages 167–202. Academic Press, Cambridge, MA.

Woods, John, Andrew Irvine, and Doug Walton (2004). *Argument: Critical Thinking, Logic and the Fallacies*. Prentice-Hall, Toronto, 2 edition.

Wray, Alison (1998). Protolanguage as a Holistic System for Social Interaction. *Language & Communication*, 18: 47–67.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikunand Yuan Cao, Qin Gao, Klaus Macherey, et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144.

Wyatt, Tristram D. (2010). Pheromones and Signature Mixtures: Defining Species-Wide Signals and Variable Cues for Identity in Both Invertebrates and Vertebrates. *Journal of Comparative Physiology A*, 196: 685–700.

Wynn, Karen (1992). Addition and Subtraction by Human Infants. *Nature*, 358: 749–750.

Wynn, Karen (1998). Psychological Foundations of Number: Numerical Competence in Human Infants. *Trends in Cognitive Sciences*, 2(8): 296–303.

Xu, Fei and Elizabeth S. Spelke (2000). Large Number Discrimination in 6-Month-Old Infants. *Cognition*, 74: B1–B11.

Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson (2014). How Transferable are Features in Deep Neural Networks? *Advances in Neural Information Processing Systems*, 27: 3320–3328.

Young, H. Peyton (2001/1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton & Oxford.

Young, H. Peyton (2011). Commentary: John Nash and Evolutionary Game Theory. *Games and Economic Behavior*, 71: 12–13.

Zahavi, Amotz (1975). Mate Selection: A Selection for a Handicap. *Journal of Theoretical Biology*, 53(1): 205–214.

Zahavi, Amotz (1987). The theory of signal selection and some of its implications. In Delfino, U. P., editor, *International Symposium on Biology and Evolution*, pages 294–327. Adriatica Editrice, Bari.

Zahavi, Amotz and Avishag Zahavi (1997). *The Handicap Principle*. Oxford University Press, New York.

Zeeman, E. C. (1980). Population Dynamics From Game Theory. In *Global Theory of Dynamical Systems*, volume 819 of *Lecture Notes in Mathematics*. Springer.

Zentall, Thomas R., Edward A. Wasserman, Olga F. Lazareva, Roger K. R. Thompson, and Mary Jo Rattermann (2008). Concept Learning in Animals. *Comparative Cognition & Behavior Reviews*, 3: 13–45.

Zhang, Dong, John A. Terschak, Maggie A. Harley, Junda Lin, and Jörg D. Hardege (2011). Simultaneously Hermaphroditic Shrimp use Lipophilic Cuticular Hydrocarbons as Contact Sex Pheromones. *PLOS ONE*, 6(4): e17720.

Zollman, Kevin J. S. (2005). Talking to Neighbors: The Evolution of Regional Meaning. *Philosophy of Science*, 72(1): 69–85.

Zollman, Kevin J. S. (2011). Separating Directives and Assertions Using Simple Signaling Games. *The Journal of Philosophy*, 108(3): 158–169.

Zollman, Kevin J. S., Carl T. Bergstrom, and Simon M. Huttegger (2012). Between Cheap and Costly Signals: The Evolution of Partially Honest Communication. *Proceedings of the Royal Society B: Biological Sciences*, 280(1750): 20121878.

Zuberbühler, Klaus (2000). Referential Labeling in Diana Monkeys. *Animal Behaviour*, 59: 917–927.

Zuberbühler, Klaus (2001). Predator-Specific Alarm Calls in Campbell's Monkeys, *Cercopithecus Campbelli*. *Behavioural Ecology and Sociobiology*, 50: 414–422.

Zuberbühler, Klaus (2002). A Syntactic Rule in Forest Monkey Communication. *Animal Behaviour*, 63(2): 293–299.

Zuberbühler, Klaus, Dorothy L. Cheney, and Robert M. Seyfarth (1999). Conceptual Semantics in a Non-Human Primate. *Journal of Comparative Psychology*, 113(1): 33–42.