Evolutionary explanations for natural language: criteria from evolutionary biology

Zuidema, W.; de Boer, B.

Publication date
2008
Document Version
Final published version

Evolutionary Explanations for Natural Language - Criteria from Evolutionary Biology

Willem Zuidema *

Institute for Logic, Language and Computation, University of Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

Bart de Boer

Amsterdam Center for Language and Speech,University of Amsterdam
Spuistraat 210, 1012 VT Amsterdam, The Netherlands

Abstract

Theories of the evolutionary origins of language must be informed by empirical and theoretical results from a variety of different fields. Complementing recent surveys of relevant work from linguistics, animal behaviour and genetics, this paper surveys the requirements on evolutionary scenarios that derive from mathematical evolutionary biology. It presents a number of simple but fundamental models from population genetics, evolutionary game-theory and social evolution theory, and evaluates their applicability to natural language. This review yields a list of required elements of evolutionary explanations in general, and of explanations for language and communication in particular.

1 Introduction

There are two distinct ways in which the study of evolution and the study of natural language overlap. First, they overlap in the search for an evolutionary explanation for why humans, and humans alone, are capable of acquiring and using natural languages. Second, the process of evolution in biology and the historical process of language change bear many similarities, and these parallels have played a role in the development of theories in both fields since the time of Darwin. We will refer to these issues as the *biological evolution of language* (or "the language faculty") and the *cultural evolution of language(s)* respectively.

* Corresponding author. Phone: +31 20 5255360, Fax: +31 20 5255206
Email addresses: jzuidema@science.uva.nl (Willem Zuidema),
B.G.deBoer@uva.nl (Bart de Boer).

When chimpanzees, our closest living relatives, are taught human language, they acquire several hundreds of signals (R. Gardner & Gardner, 1969; Savage-Rumbaugh, McDonald, Sevcik, Hopkins, & Rubert, 1986). They fail, however, to produce speech sounds themselves, to acquire the many tens of thousands of words in natural languages, and to grasp the use of quite basic rules of grammar (Terrace, 1979; Truswell, to appear). Human infants, in contrast, acquire their native language rapidly. They produce speech sounds and comprehend simple words before the age of 1, produce their first words soon after their first birthday and the first grammatical constructions before their second birthday (Tomasello & Bates, 2001).

Why? Clearly there is something special about humans that makes them extraordinarily apt to acquire and use natural languages. Among other things, the anatomy of the vocal tract, the control mechanism in the brain for complex articulation and the cognitive ability to analyse and produce hierarchically structured sentences appear to be qualitatively different in humans than in other apes. How did this capacity for language come about? One possibility is that the human capacity for language has emerged purely as a side-effect of the many changes in anatomy and cognition that occurred in the hominid lineage. Although this possibility cannot be dismissed, from a biological point of view it does not appear very likely. Humans spend around 3 hours a day or over 20% of their awake time talking (Dunbar, 1998, and references therein), verbal abilities play a significant role in social status and, it seems, in both the reproductive success of individuals and the success of our species as a whole. Such a salient characteristic of any organism would require a Darwinian, evolutionary explanation. Hence, although the side-effect scenario is a possibility, it can only be the conclusion of an elaborate investigation, and not serve as null hypothesis. Nevertheless, although language as a whole might be considered a biological adaptation, many specifics about language (language universals) are perhaps better understood as the outcome of cultural evolution. In this view, the complex results of cultural evolution and social learning have had indirect consequences for biological evolution.

Both biological and cultural evolution of language have received a great deal of attention in recent years, leading to a plethora of theories and models (Hurford, Studdert-Kennedy, & Knight, 1998; Christiansen & Kirby, 2003) and a number of monographs each year (Dessalles, 2007; Heine & Kuteva, 2007; Hurford, 2008; MacNeilage, 2008; Mufwene, 2008). Many proposals involve a single mechanism or factor responsible for the emergence of modern natural languages. In some cases, extensive scenarios for the evolution of language are proposed. Although this enormous body of work contains a great number of interesting ideas and findings, there are also a number methodological problems. First, it is extremely difficult to relate separate proposals to each other, because of a lack of consensus on terminology and basic assumptions. Second, it is extremely difficult to evaluate the internal consistency and empirical validity of proposed theories, because of a lack of formal rigour.

In some ways this situation is reminiscent of the state of the whole field of evolutionary biology before the establishment of theoretical population genetics by Fisher, Wright, Haldane and others in the 1920s and 30s. Their mathematical models, and the subsequent informal "modern synthesis", convinced biologists of the central role of natural selection in evolution. Confusion remained about the units of selection, but with the settling of the group selection debate by Maynard Smith (1964) and Williams (1966) a consensus emerged about the minimum requirements for evolutionary explanations, as well as a common vocabulary in which disagreements can be phrased.

In the interdisciplinary field of language evolution, this clarity is yet lacking. It is our opinion that much could be gained if the progress made in evolutionary biology would be wider known in linguistics and other fields that contribute to the study of language evolution. To this end, we will review some simple mathematical models from evolutionary biology, and evaluate how they can be applied to both the biological and the cultural evolution of language. This paper thus complements important reviews of results relevant to language evolution from genetics (S. Fisher & Marcus, 2006), comparative psychology (Doupe & Kuhl, 1999; Hauser & Fitch, 2003; Pinker & Jackendoff, 2005), linguistics (Jackendoff, 1999) and neuroscience (Deacon, 2000), as well as more specific, "single-paradigm" reviews of language evolution modelling (Kirby, 2002; Nowak, Komarova, & Niyogi, 2002; Wang, Ke, & Minett, 2004; Számadó & Szathmáry, 2006). The bird's eye view of this paper makes clear, we hope, where the real challenges for language evolution research lie.

The contribution of this paper will be mostly a methodological one. We will not focus on the exact phenomena, neither the linguistic phenomena under study, nor their genetic or neurological implemantations. We will also not focus on reconstruction of ancestral traits by comparison with our evolutionary ancestors. Rather, we will provide an overview of techniques from theoretical biology that may help to put the linguistic, genetic, neurological and biological phenomena in coherent and theoretically adequate evolutionary scenarios. As we take our inspiration from theoretical biology, the focus will be on *biological* evolution, not *cultural* evolution. However, we will from time to time indicate how results from biology might be applied to the study of cultural evolution. Most of the models we discuss are well-known in the various subdisciplines of evolutionary biology and represented in various textbooks. However, we know of no paper that reviews, as we do here, how simple mathematical models have shaped the thinking of evolution at levels from the gene to social behavior in groups, and certainly not in a form accessible to linguists and cognitive scientists.

We will start with some classical results from population genetics, about the way gene frequencies in a population change as a result of mutation and selection, and then discuss the case for viewing natural selection as optimisation, as well as the problems with this view. This optimisation view then provides a natural bridge

to evolutionary game theory, where the targets of optimisation shift because the opponents in the game evolve as well. Finally, extensions to social evolution models that deal with kin selection, will lead us to the issue of levels of selection, and clarify the relation of cultural evolution models – with the dynamics happening at the level of cultural replicators – to evolutionary biology generally.

We will propose a list of criteria that an ideal theory of the evolution of language should satisfy. We realize that in the present state of knowledge no account of the evolution of language can satisfy all criteria, but the list can nevertheless serve as a point of reference and as a reminder what in biology is expected of an evolutionary explanation.

2 Evolution as Gene Frequency Change

If we want to investigate specific hypotheses on adaptations for language, what form should such hypotheses take? The early formal models in population genetics are a useful starting point. But first, it should be clear that any statement about biological evolution is a statement about how genes mutate and spread in a population through random drift and selection. That statement in no way reflects the form of genetic determinism or naivety about "language genes" that have made some evolutionary linguists wary to talk about genes at all. But if properties of language are to be explained by some biological endowment, which in turn is to be explained as an adaptation for language, then we need to be explicit and postulate a series of altered genes that influence the ability for language. Such genes can have many additional non-linguistic effects (an illustrative example is the recently discovered FOXP2 gene, that, when mutated, causes a range of problems in language processing as well as in sequencing orofacial movements, Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). We can phrase this criterion as follows:

Criterion 1 (Heritability) *Evolutionary explanations for the origins of a trait need to postulate genetic changes required for that trait.*

Of course, one can sensibly study the evolution of traits for which the genetic component has not been identified. The point here is to emphasise that biological evolution implies genetic changes.

A formal model of evolution as gene frequency change can be built-up in the following way. Consider first that in humans, as in almost all sexually reproducing organisms, every individual inherits two versions of each gene, one from the father and one from the mother. If there is to be any change, we need to consider at least two different variants, alleles, for each gene, and monitor the increase in frequency of one allele at the expense of the other. In figure 1 the Mendelian model of inheritance of two alleles – A and a at a single locus – is depicted. Adults (top row)

have a genome that is of any of the three possible types AA, Aa or aa (Aa and aA are equivalent). These adults produce sperm and egg-cells (second row) with just a single copy of the gene under consideration. In sexual reproduction, a sperm-cell and an egg-cell fuse, and grow out to a new individual (third row). Evolution, in this simple scheme, concerns the change in frequencies of the types AA, Aa or aa, or the change in frequencies of the alleles A and a.

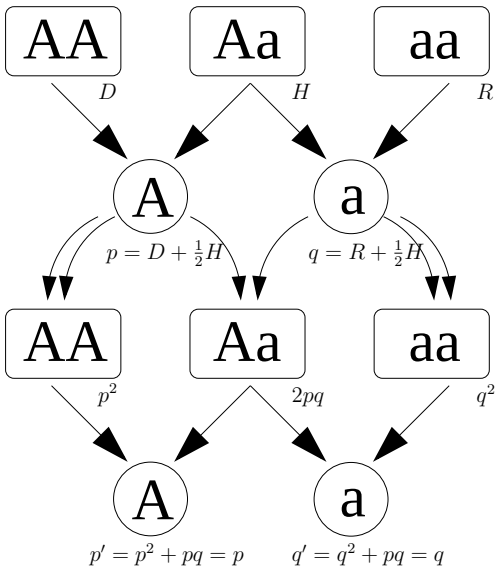

Fig. 1. Mendel's model of inheritance, and the Hardy-Weinberg model of allele and genome frequencies under Mendelian inheritance with no selection nor drift.

The Hardy-Weinberg model (developed independently by British mathematician Godfrey Harold Hardy, 1908 and German physician Wilhelm Weinberg, 1908; see Crow, 1999) describes the gene frequencies if there is no mutation or selection. Consider the frequencies of the three types (top row) at any particular point in time, and call these frequencies D, H and R. The frequencies of the alleles A and a in the sperm and egg-cells are simply:

$$\begin{aligned} \text{frequency of } A: \; p &= D + \frac{1}{2}H \\ \text{frequency of } a: \; q &= R + \frac{1}{2}H, \end{aligned} \tag{1}$$

because individuals with type AA or aa will always pass on an A or a respectively to their sperm and egg-cells, but individuals with type Aa only half of the time.

Under a number simplifying assumptions (including random mating, random meiosis, an infinite population and no sex differences at the relevant locus), the frequencies of the three types in the offspring are simply $D' = p^2$, $H' = 2pq$ and $R' = q^2$, because you need two A's or a's to make an AA or aa respectively, and you need an A from either the father or the mother and an a from the other parent to make an

Aa. When this offspring then starts producing sperm- and egg-cells, the frequencies of the alleles A and a are:

$$\begin{aligned} &\text{new frequency } p' = D' + \frac{1}{2}H' = p^2 + pq \\ &\text{new frequency } q' = R' + \frac{1}{2}H' = q^2 + pq. \end{aligned} \tag{2}$$

Hardy and Weinberg's simple but fundamental observation is that because $p+q = 1$ (the total frequency of all alleles must be 1, and thus $q = 1 - p$), it follows that p and q are constant under this model of inheritance:

$$p' = p^2 + pq = p^2 + p(1-p) = p^2 + p - p^2 = p. \tag{3}$$

This result shows that under Mendelian inheritance existing variation in gene frequencies is maintained. This is in contrast with "blending inheritance" (the assumed model of inheritance before the rediscovery of Mendel's laws around 1900), where a child's trait values are the average of the parents' and variation quickly dissipates over time. The result played a crucial role in reconciling Mendelian genetics with Darwinian evolutionary theory, because it showed variation remains stable long enough for natural selection to operate (R. A. Fisher, 1930, chapter 1).

The Hardy-Weinberg model can be extended in a straightforward manner to include the effects of selection. Natural selection, in Darwin's theory, is the consequence of differences in survival rates to the age of reproduction and the differences in reproductive success. These effects can be summarised with a fitness coefficient for each of the possible types, which gives the expected number of offspring. A high coefficient w_{AA} means that individuals of type AA reproduce successfully, such that their genes are well represented in the next generation. In terms of the equations, this just requires weighting the contributions of parents of each type with the relevant fitness coefficient:

$$p' = \frac{p^2 w_{AA} + pq w_{Aa}}{\overline{w}}, \tag{4}$$

where $\overline{w}$ is the average fitness and given by:

$$\overline{w} = p^2 w_{AA} + 2pq w_{Aa} + q^2 w_{aa} \tag{5}$$

(this term is needed to account for changes in population size due to reproduction and selection).

Equation (4) gives us a first handle on the requirements for evolutionary innovation, and, hence, evolutionary explanations. First of all, natural selection operates on

variation. Second, natural selection favours fitter genes and individuals over less fit ones. Both the variation and the fitness differences need to be made explicit:

Criterion 2 (Strategy set) *Evolutionary explanations need to postulate a set of possible types* [1] *, as well as the mutations that can move an organism from one type to another.*

Criterion 3 (Payoff function) *Evolutionary explanations need to postulate a function that relates the possible types in a given environment (that may include other evolving individuals) to fitness.*

If we are interested in a specific biological innovation – that is, a mutation – that was relevant for learning or using language, we need to consider the situations before and after that mutation. In the simplest case, a is the preexisting gene that is initially shared by the whole population, and A is the mutated version of a that has arisen in a single individual. Hence, initially $q \approx 1$ and $p \approx 0$. If A is to play a role in an evolutionary scenario, we need to establish that allele A spreads in the population (as sketched in figure 2); in other words, that p increases. We can formulate this criterion as follows:

Criterion 4 (Invasibility) *Innovations in an evolutionary scenario need to be able to* invade *a population; that is, an innovation should spread in a population where it is extremely rare.*

If we know all fitness coefficients, it is straightforward to work out what happens to the frequency of the new mutation. As it turns out, A will spread if $w_{Aa} > w_{aa}$, and it will get fixed ($p = 1$) if $w_{AA} > w_{Aa}$. In other words, the fitness of the new gene must be greater than that of the old one, and the new gene must, to some extent, be *dominant* over the old one such that its effects are noticed in individuals that inherit copies of both genes from each of the parents. In fact, the difference in fitness between the two variants must be significant, at least large enough for the new gene not to get lost by chance fluctuations (R. A. Fisher, 1922) and to get established after a reasonable number of generations (Haldane, 1932). Note that these results depend on some strong assumptions, including an infinite population with randomly interacting individuals.

[1] In biological systems, organisms have a genotype (their genes) and a phenotype (their body and behavior). Variation and inheritance take place at the level of the genotype, while selection takes place at the level of the phenotype. In cultural evolution, there is not necessarily a distinction that corresponds to the genotype and the phenotype, and we therefore use the more neutral term "type". When referring to biological organisms, we do use the terms genotype and phenotype.

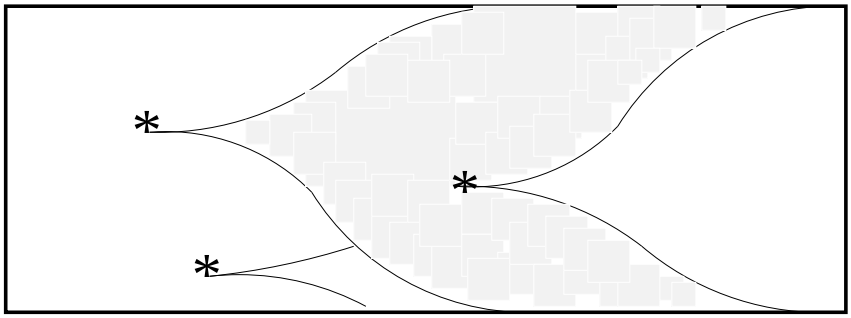

Fig. 2. Evolutionary explanations are concerned with the spread of new genes in a population. In this diagram, the horizontal axis represents time and the vertical axis represents the ordered individuals of a population. At several points in time, a new mutation arises in a single individual. A crucial question is then whether or not this new mutation is able to invade. The diagram illustrates that even if a mutation is initially successful (i.e. the second), it can be swept out by an even more successful alternative type (i.e. the first). The effects of sexual reproduction and recombination – not shown here – can be that independent beneficial mutations that arise around the same time, both get established in the population.

3 Evolution as Optimisation

Since Darwin (1859), the notion of "adaptation" has played a major role in evolutionary thinking. His work offered a coherent framework to study the traits of organisms in terms of their *function* for survival and reproduction. Even before the mechanisms of genetic inheritance were unravelled, Darwin thus transformed biology from a descriptive to an explanatory science. In the early 1920s the "founding fathers" of population genetics – Fisher, Wright and Haldane – worked out what happens to a single new gene when it appears in a population. But do the dynamics described by equation (4) constitute "adaptation"? In other words, does the predicted change in gene frequencies also mean the population will get better adapted to its environment, i.e. improve its average fitness?

Both Fisher and Wright set out to work out a more general result. I will discuss Fisher's "fundamental theorem of natural selection" (R. A. Fisher, 1930) in section 6. Here I will follow Wright's analysis of the average fitness in a population, in particular Roughgarden's (1979) version of these equations. It is useful to look at a couple of Wright's equations. First, it is convenient to look at the *change* in the frequency p at every time step. This is, using equation (4), given by:

$$\begin{aligned}\Delta p &= p' - p \\ &= \frac{p^2 w_{AA} + pq w_{Aa}}{\overline{w}} - p \end{aligned} \tag{6}$$

This equation can be rewritten as follows:

$$\Delta p = \frac{pq}{\overline{w}} \left(p \left(w_{AA} - w_{Aa} \right) - q \left(w_{aa} - w_{Aa} \right) \right) \tag{7}$$

This equation tells us nothing new; it is essentially equation (4) in a different form. However, the new form will prove useful when we have worked out the next equation. We are interested in what happens to the average fitness when the frequency (p) of the innovation changes. Mathematically, this is represented by the derivative of $\overline{w}$ with respect to p. The expression for average fitness is given in equation (5). Its derivative, if we assume the fitness coefficients are independent of p and q (that is, no frequency-dependence) turns out to be:

$$\frac{\mathrm{d}\overline{w}}{\mathrm{d}p} = 2 \left(p \left(w_{AA} - w_{Aa} \right) - q \left(w_{aa} - w_{Aa} \right) \right) \tag{8}$$

When we note that equations (7) and (8) are very similar, it is clear that we can replace a large part of (7) with (8), and get:

$$\Delta p = \frac{pq}{\overline{w}} \left(\frac{1}{2} \right) \frac{\mathrm{d}\overline{w}}{\mathrm{d}p}. \tag{9}$$

This is a fundamental result for evolutionary biology. The equation says that the change in the frequency of a new gene, will be *in the direction* of the derivative of fitness with respect to that gene's frequency. That means that only if the average fitness increases with increasing p, will the new gene spread. Moreover, the spread will be fastest at intermediate frequencies (high variance) and low average fitness. In other words, evolution – under the assumption mentioned – will act to optimise the average fitness in the population: it will lead to adaptation.

However, the mathematical derivation of this intuitive result also tells us about its limitations. First of all, evolution is shortsighted. We saw a simple example at the end of the previous section: if $w_{Aa} < w_{aa}$ (there is "heterozygous disadvantage"), then the new allele A will not spread in the population, even though at fixation it might improve the mean fitness in the population. Second, evolution needs (heritable) variation. If $p = 0$ or $p = 0$, nothing will change. Thirdly, the equation is only valid if the fitness coefficients are *independent* of p and q. That is, whatever the traits are that allele A influences, the usefulness of the innovation should not depend on how many others in the population share it. This condition is obviously violated in the evolution of communication, because the usefulness of a signal will always depend on the presence of others that can perceive and understand it. Fourthly, the original Hardy-Weinberg model brought quite a lot of assumptions, including the independence of the single locus we looked at from other loci, random mating, discrete generations and infinite populations. Some of the consequences of relaxing dependence between genes, of relaxing the frequency independence and of relaxing the random mating assumptions will be evaluated in sections 4.2 and 5 and 7, respectively.

Finally, as R. A. Fisher (1930) emphasised, these calculations deal only with the direct effects of natural selection. They predict the direction of change, but it is unwarranted to conclude that the average fitness in a population will increase. Environmental conditions might have changed in the mean time and, even if the environment is constant, all individuals in the population are better adapted to it such that competition is fiercer. These effects – not modelled by Wright and Fisher's equations – were collectively labelled "deterioration of the environment" by Fisher.

In addition to these quantitative results, Wright made a much more qualitative contribution relating evolution and optimisation. In a paper without any mathematics (Wright, 1932) he introduced an extremely influential metaphor: the **adaptive landscape**. The adaptive landscape is a landscape of 3 or more dimensions, with the plane (or hyperplane) representing the space of possible types, and the height of every point representing fitness (see figure 3). On such a landscape, a population is a collection of points. Mutations correspond to steps in the landscape; selection corresponds to the selective removal of individuals that are lower down. The process of evolution involves the population to climb up-hill, following a local gradient to a local peak.

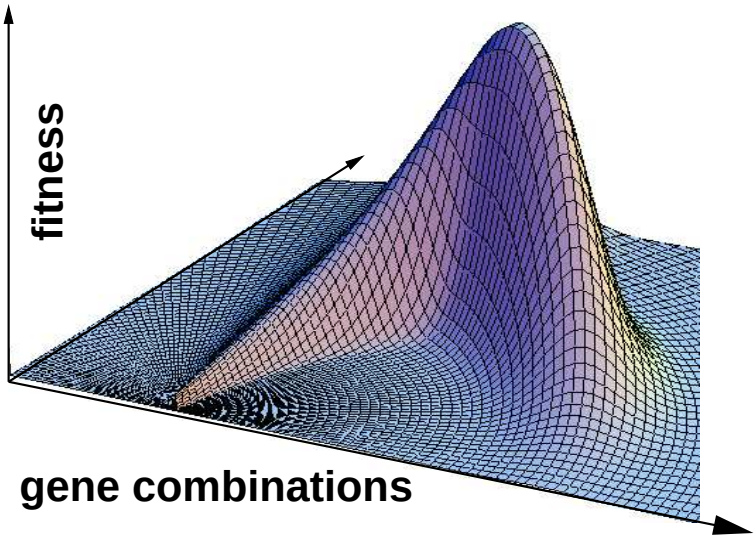

Fig. 3. The adaptive landscape of fitness as a function of type. The graph illustrates an hypothetical examples in which two genes have a continuous range of effects. Real organisms have, in contrast, a discrete set of possible genotypes involving many more than two genes. Thus, mutations can take them in very many directions. This high dimensionality makes it more likely that there is some path uphill to the "adaptive peak" (see Provine (1986), chapter 9).

We will discuss some problems with the concept below. However, the adaptive landscape representation in this form does illustrate Darwin's (1859) insight that for a process of continuing evolution, we need a path of ever increasing fitness from the hypothesised initial point in type space to the end result. (In finite populations, stochastic drift can bridge fitness barriers in the adaptive landscape, but only if they are relatively shallow.) For complex traits, such as language, it seems reasonable to

postulate a series of many genetic changes. Wright's metaphor highlights the fact that each of these changes needs to confer an adaptive advantage:

Criterion 5 (Fit intermediates) *Explanations for complex traits, that involve a series of genetic changes, need to show a path of fit intermediates, from the hypothesised initial state to the desired end state.*

This criterion is important, but it might not be as problematic as it looks at first sight. First, although evolution will generally lead uphill, there is some room for random processes as well. Wright used the adaptive landscape metaphor to explain the effects of increases or decreases of the rate of mutation and the strength of selection. He also discussed at some length the effects of small population sizes, where inbreeding will lead to the non-selective process of genetic drift: random deviations from the locally optimal type due to accumulation of mutations and a lack of variation for selection to operate on. Wright's shifting balance theory (or at least one version of it) argues that the additional variation inherent in subdivided and inbreeding populations could help the population as a whole bridge fitness barriers. Although the shifting balance theory has little empirical support (Coyne, Barton, & Turelli, 2000), the basic idea that, under some conditions, genetic drift could help bridge a fitness barrier remains.

Second, one of the basic tenets of evolutionary biology is that all life originates from the same source. If that is true, all complex traits of all organisms are connected through paths of fit intermediates. Thus, if we wonder if there is a path on the adaptive landscape through which humans could evolve wings, the answer must be yes. Humans, bats and birds have a common ancestor, so there must be at least one series of environments (including other species) that would yield a path that leads from humans back to the common ancestor with bats, and again forward to modern bats (ignoring some difficulties such as frequency-dependent fitness).

Third, intuitions about getting stuck in local peaks based on the three-dimensional representation as in figure 3 must be treated with care. There are, in fact, a great number of problems with the concept (Provine, 1986, in his biography of Wright, gives a thoughtful critique). First of all, as Wright indicated, an actual genome consists of many (tens of) thousands of genes. Hence, the adaptive landscape has tens of thousands of dimensions, rather than just 3. That makes a big difference, because whereas local peaks seem extremely likely in 3 dimensions, they are in fact increasingly less likely with more and more dimensions. But, perhaps more importantly, the type space in Wright's graph is continuous, whereas the genotypes of actual organisms are discrete. Wright's landscapes, as drawn here, can in fact never be constructed for a real example.

Wright and others have looked at other versions of the adaptive landscape that are, in contrast, rigorously defined. One approach is to choose the gene frequencies and population average fitness as axes. A population, in this representation, is then a

single point in the landscape. The advantage of this representation is that it ties in nicely with the mathematical model of equation (9). However, the disadvantage is that in such a landscape one cannot visualise the effects of selection, mutation, genetic drift and subdivision of the population, which was the whole point of introducing the metaphor.

Alternatively, one can choose to use phenotypic, continuous traits against individual fitness as the axes of the landscape. The disadvantage of this approach is that mutations, which define what a type's "neighbours" are, are of course defined genotypically. Therefore, the random variation that builds up by mutation, will not generally be centred around a single population mean in phenotypic space. In cases where very little is known about the genetics anyway, such as language, that might not really matter, but, as we will see, there the landscape cannot be constructed anyway because of frequency dependence.

Nevertheless, the view of evolution as optimisation yields a powerful approach for deriving predictions about an evolving system, or for understanding an evolved system as adapted for a specific purpose. Parker and Maynard Smith (1990) present a methodology for evolutionary reasoning based on this view which they call "optimality theory" [2] . They first emphasise that every evolutionary study must start with identifying a clear biological question. Step 2 is to identify a set of strategies that are available for evolution to choose from. Step 3 is to identify a pay-off function, which evolution is supposed to optimise, and to show that the observed biological phenomenon tends towards the optimum. Step 4 is to relate pay-off, which is an indirect measure for fitness, to actual fitness. Finally, step 5 is to derive predictions and test them empirically.

This scheme provides a coherent framework for thinking about the evolution of language, and it is essentially the approach we champion in this paper. Note however, that the mathematical models discussed so-far concerned changes in gene frequencies, whereas Optimality Theory and language evolution research are concerned with phenotypic traits that typically involve many, often unknown genes. We will first discuss some limitations of the optimality view that apply even when we look at traits controlled by a single gene, and then discuss the more difficult issue of going from single-gene models to the evolution of complex phenotypic traits such as language.

[2] Parker & Maynard Smith's (1990) Optimality Theory is completely unrelated to Optimality Theory (Prince & Smolensky, 2004) in linguistics.

4 Limits to Optimality

> *"Natural selection tends only to make each organic being as perfect as, or slightly more perfect than, the other inhabitants of the same country with which it comes into competition. And we see that this is the standard of perfection attained under nature" (Darwin, 1872, p 163; quoted in Provine, 1986, p209).*

As Darwin was well aware, the fact that evolution can be thought of as optimisation does not imply that the features of organisms are optimal or perfectly adapted to their environment. The most obvious evidence for the existence of limits to optimality, are the many examples of indigenous species that are rapidly driven to extinction after humans introduced a foreign, competing species. There is a whole tradition of listing the limitations of natural selection (e.g. Dawkins, 1982; N. Barton & Partridge, 2000). These can be classified in five types: (i) biophysical constraints, (ii) genetic constraints, (iii) the speed of evolution, (iv) mutational load and (v) fluctuating fitness.

4.1 Biophysical constraints

With regard to **biophysical constraints**, it is clear that all of the complexities of biological organisms need to grow out of a single cell. Throughout its development, an organism needs to maintain its metabolism, to selectively take up chemicals from its environment and to autonomously build-up all of its complex features. That process of biological pattern formation is constrained by what is possible at all with the materials available in a biotic environment, by what can be coded for by genes, and by which possibilities are reachable for evolution. It is obvious that these constraints are at work, given for instance the limitations in speed of both a prey and a predator trying to outrun each other. It is also obvious, however, that these limitations have not prevented evolution from building exquisitely complex and well-adapted organs such as, for instance, the human ear.

The constraints and trade-offs are all crucial elements of an adaptive explanation. As evolution operates in the physical world, it is impossible to optimize all aspects of an organism simultaneously. All traits of an organism have costs and benefits, and the payoff function (as mentioned in criterion 3) represents the difference between the benefit of a trait and its cost. As evolution optimizes the payoff, rather than maximizing the benefit or minimizing the cost, it is expected to move towards a solution that is in some sense a compromise between cost and benefit. The more precise we can be about constraints and trade-offs,the more convincing demonstrations of optimality within these constraints are as evolutionary explanations.

The best examples of trade-offs in language are probably in the physical properties of speech. Liljencrants and Lindblom's (1972) demonstration that the vowel

systems in human language appear to be optimised for reliable recognition under noisy conditions and under constraints on perception and articulation, is suggestive. Lieberman (1984) has argued that the human larynx has descended deeper down the throat in order to allow more flexibility of the articulatory organs. This allows us to make many different speech sounds, at the expense of an increased propensity to choke. Although controversial (Hauser & Fitch, 2003), this theory on the evolution of language does illustrate the role of evolutionary trade-offs that result from the physiological constraints in speech production.

For other components of human language, such as its semantics or syntax, it is extremely difficult to derive biophysical constraints. What sort of grammars can or cannot be encoded by genes and implemented in neuronal tissue? The only solid results relevant to this question, suggest that quite a variety of networks of interacting cells are *Turing equivalent*. That is, they can – if sufficiently large, given sufficient time and properly initialised and interpreted – compute any computable function (Siegelmann & Sontag, 1991; Wolfram, 2002). This is not to say that any grammar can be easily encoded by genes or acquired by a neural net; but without better models of the neural implementation of language, we cannot start to make sensible assumptions about the actual architectural constraints on natural language syntax that were at work during human evolution.

However, it would be overly pessimistic to conclude that we can therefore not say anything sensible about how language evolved. There are two categories of constraints in language evolution that can be made precise. First of all, we have good "mentalist" models of syntax that describe its fundamental computational properties, and the **computational constraints** that any implementation will face. For instance, we know there exist constructions in natural languages that cannot be modelled by weaker formalisms (in terms of the extended Chomsky Hierarchy) than (mildly) context-sensitive rewriting grammars (Joshi, Vijay-Shanker, & Weir, 1991); we know that the whole class of context-sensitive rewriting grammars is not *identifiable in the limit* from positive samples alone (Gold, 1967); and we know that grammars of that type have a worst-case time-complexity of $O(n^5)$ in parsing (G. E. Barton & Berwick, 1987). Such computational constraints on representation, learning and processing, and the formalisms they are expressed in, allow us to at least make a start with testing the internal consistency of an evolutionary scenario, and with formulating a sensible strategy set for evolution.

4.2 Genetic constraints

Population and molecular genetics make some specific predictions on **genetic constraints**. Natural selection can often not optimise all different phenotypic traits independently from each other, because of the following features of genes:

- A single gene typically has an effect on many different phenotypic traits (pleiotropy);
- The effect of a gene on a trait depends on the presence or absence of other genes (epistasis);
- Genes are physically linked to each other in a chromosome (linkage).

The little that is known about human genetics relevant for language (e.g. Lai et al., 2001) suggests, unsurprisingly, that all these general observations hold for language as well. Most of these biophysical and genetic constraints are reflected in the choice of the strategy set, which contains all strategies/trait values that are available to evolution, and excludes those that cannot be instantiated. The physical linkage between genes, however, is – in the long term – not one of these hard constraints on what can evolve, because recombination will eventually break the linkage such that one gene can occur without the other. Linkage does constrain how fast things can evolve, which is also crucial for the course of evolution.

4.3 The speed of evolution

More generally, the **speed of evolution** is constrained by the available genetic variation at every step (including effects from linkage) and the strength of selection. Considerations about evolutionary time should be included in evolutionary explanations:

Criterion 6 (Sufficient time) *Evolutionary explanations need to establish that there has been enough time for favourable alleles to get established in the population.*

4.4 mutational load

Evolution needs variation to operate on, and mutation is the source of this variation. However, because mutation is indiscriminate and random, it will also constantly create individuals that are worse than average, or even inviable. This is called **mutational load**. In the adaptive landscape metaphor, whereas selection will push a population to the top of an adaptive peak, mutation will pull the population downhill. The dynamic equilibrium is called *mutation–selection balance*.

A series of formal models of the cultural transmission of language have been proposed (Nowak, Komarova, & Niyogi, 2001; Komarova, Niyogi, & Nowak, 2001; Mitchener & Nowak, 2002) based on the concept of mutational load. It is therefore worth looking at how this concept has been formalised. However, rather than discussing the general population genetics analysis of mutational load, we will here only focus on Eigen's (1971) model, on which the studies of language evolution are based. See Wilke (2005) for a discussion of its relation with population genetics.

Using notation loosely based on Maynard Smith and Szathmáry (1995) and Nowak et al. (2001), we can write Eigen's equation as follows:

$$\Delta x_i = \sum_{j=1}^{M} x_j w_j \mathbf{Q}_{ji} - x_i, \tag{10}$$

where i and j are indices for all the M distinct possible genotypes. Δx_i stands for the changes of the frequencies of all genotypes i (hence, the expression (10) defines a coupled system of equations with one equation for each possible i). x_i is the frequency of genotype i and w_i its fitness. $\mathbf{Q}_{ji}$ is the probability that a given child will have genotype i if its parent has genotype j. Hence, $\mathbf{Q}$ is a matrix of size $M \times M$ that describes the effects of mutation.

Eigen looked at a very specific choice of parameters. Suppose that there is a single genotype with a high fitness, and all other genotypes have the same, low fitness. That is, the adaptive landscape is flat, except for a single high peak. Now suppose there is a constant probability μ of mutation per gene, and no cross-over. The probability q that an individual consisting of l genes produces identical offspring is now:

$$q = (1 - \mu)^l, \tag{11}$$

where l is the genome length. q is called the "copying fidelity". One can work out where the mutation–selection balance is for different mutation probabilities. Eigen's result is that there is a precise value of q where the mutation–selection balance suddenly drops to vanishingly small quantities of each possible genotype. That is, if the mutation probability is above a threshold value – the *error threshold* – selection ceases to play any role, and individuals have essentially random genotypes:

Criterion 7 (Mutational load) *Evolutionary explanations need to postulate a mutation rate high enough to generate the variation needed, but low enough to not suffer from an extreme mutational load (to cross the error threshold).*

Komarova et al. (2001) worked out a version of Eigen's equations under frequency-dependent selection (see below), and applied it to the learning and evolution of language. They showed that there is a precise threshold for the accuracy of language learning, for the selective evolution of alternative Universal Grammars to be possible.

4.5 Fluctuating fitness

A final category of limits on optimality comes from **fluctuating fitness**, that is, from the fact that the fitness regime of organisms is constantly changing. First of all, there are temporal fluctuations in the environmental conditions on many different timescales, both regular and irregular: from the day and night cycle to climate changes. Similarly, there are geographic differences, such that migrating organisms might find themselves in very different habitats. Organisms adapted to one set of conditions, are not necessarily adapted to other conditions.

But perhaps more interesting is the situation where the fitness regime of a particular species changes due to evolutionary changes of the species itself (**frequency dependent selection**) or of any of the other species it interacts with (**co-evolution**). The evolution of language and communication is frequency-dependent, because linguistic innovations are unlikely to pay off if there is no one to talk to. The fitness coefficients in language evolution are therefore not constants, as in equation (8), but will depend on the frequencies of the different alleles in the population. Evolutionary game theory is the general framework for addressing frequency-dependent selection, and will be discussed in the next section. Because natural languages are transmitted culturally, there can also be a process of cultural evolution, such that we can perhaps sensibly speak about the *coevolution of language and the brain* (Deacon, 1997; Christiansen, 1994; Kirby, 1994). This is explored a bit further in section 8 in general terms.

A related phenomenon is **sexual selection**, where selection is not on the ability to survive to reproductive age or the ability to reproduce per se, but on the ability to beat rivals of the same sex in the competition for a mate, or on the ability to persuade potential sexual partners to choose one as a mate (Darwin, 1859, p.94). Here, the fitness of a given genotype (defining e.g. a male trait) is not fixed, but also dependent on the frequency of all the possible genotypes (regulating e.g. female preferences) in the population. Traits that are the result of sexual selection generally appear at first sight useless or maladaptive, but help to attract mates or deter competitors for mates. In the evolution of speech, sexual selection seems to have played a role in shaping the secondary sexual traits, such as the lower pitch in human male voices, which results from larger larynx and vocal folds, and a change in formant frequencies at puberty, which makes males appear larger and results from a second descent of the larynx.

5 Evolutionary Game Theory

The formal framework to describe the consequences of multiple agents optimising their own payoff in a social context is the **Theory of Games**. Game theory concep-

tualises the interaction between agents, the "players", as a game where all players choose from a set of available strategies. Crucially, the outcome of a game for each player, its payoff, depends on the strategies of other players.

The following example is derived from Maynard Smith and Price (1973). Imagine a conflict between two birds competing for a single food source, each with the choice between three strategies: "dove" (retreat immediately if the other player is aggressive), "hawk" (always be aggressive) and "prober" (start off aggressive, but share the food source peacefully if the other player does not give up, but does not escalate either, and continue aggressively if the other player does give up). We assume that the value of the food source is $b = 10$ and the expected cost of an escalated fight $c = 100$. For 2 players and a small number of discrete strategies, the payoffs can be conveniently summarised with a *payoff matrix*, as in figure 4.

	player 2's strategy		
player 1's strategy ↓	Dove	Hawk	Prober
Dove	$\frac{1}{2}b = 5$ $\frac{1}{2}b = 5$	$b = 10$ 0	$b = 10$ 0
Hawk	0 $b = 10$	$-c = -100$ $-c = -100$	$-c = -100$ $-c = -100$
Prober	0 $b = 10$	$-c = -100$ $-c = -100$	$\frac{1}{2}b = 5$ $\frac{1}{2}b = 5$

Fig. 4. Extensive and matrix representations of games, showing the payoff of players 1 (F_1) and 2 (F_2) for different strategies played by each player.

We can postulate a decision mechanism for each player, and study how the outcome of the game changes with players adapting their strategies based on what the other players do. The dynamics of such games, with all players making their own decisions, are often extremely difficult to describe. Often, however, it is possible to derive the conditions under which a game is stable (i.e. where no player will change strategy). In non-cooperative game-theory – where "selfish" players each try to optimise their own payoff – the crucial concept is that of a **Nash equilibrium** (Nash, 1950). This equilibrium is defined as the situation where no player can increase its payoff by unilaterally changing its strategy. As *rational* players are assumed to maximise their payoff, games will therefore typically evolve toward a Nash equilibrium.

In evolutionary biology (after some pioneering work by R.C. Lewontin and W.D. Hamilton, as is discussed in Maynard Smith, 1982) the use of game theory took off with the work of Maynard Smith and Price (1973) and Maynard Smith (1982). Maynard Smith & Price introduced the concept of **Evolutionarily Stable Strategy** (ESS) in an analysis of the evolutionary advantages of "limited war" strategies in

animal conflicts, such as the prober strategy introduced above. An ESS is a strategy that cannot be *invaded* by any other strategy, because all other strategies get either a lower payoff when playing against the ESS, or if their payoff is equal, they get a lower payoff when playing against themselves. That is, if $F(i, j)$ gives the payoff for a player playing strategy i against an opponent playing strategy j, then i is an ESS if for every strategy j either $F(i, i) > F(j, i)$ or $F(i, i) = F(j, i) > F(j, j)$. Every ESS also defines a Nash Equilibrium (if the game is symmetric) but the stability criterion is stricter, because it implies that every alternative strategy will be selected against if it occurs at small frequency in the population.

In the example of figure 4, we can see that the dove-strategy is not an ESS, because the hawk-strategy has a higher payoff when playing against it. In a populations of doves, the hawk strategy thus enjoys an initial selective advantage and will increase in frequency. The hawk-strategy is not an ESS either. A population consisting of just hawks can in turn be invaded by the dove-strategy, which has a higher payoff in a population of hawks, or by the prober-strategy, which has equal payoff against hawk but a higher payoff against itself. Only the prober strategy, in the present simple model, is an ESS: both doves and hawks fare worse than the prober in a population of probers [3] .

If we exclude the prober-strategy from the strategy set, the resulting hawk-dove game has no ESS, i.e. a population of individuals all playing one pure strategy, can be invaded by the other strategy. However, as a small group of doves will increase in number in a population of mostly hawks and vice versa, the population will stablize in a state in which there are both hawks and doves. This is called an **Evolutionarily Stable State**. In such a situation, there are distinct, genetically different players in the population, and this variation is maintained by selection. This implies that strategies can remain in a population, even though they are not evolutionary stable strategies.

The techniques and formalisms from evolutionary game theory immediately lead to some fundamental observations on the evolution of communication. Consider the evolution of an alarm call system similar to the calls that, for instance, ground squirrels (Sherman, 1977) or vervet-monkeys (R. Seyfarth, Cheney, & Marler, 1980) use to inform conspecifics of the presence of predators. If we focus on just two signals, 1 and 2, and just two types of predators, aerial (E, e.g. eagles) and terrestrial predators (L, e.g. leopards), we can postulate the following strategy set:

[3] In the original paper (Maynard Smith & Price, 1973), this game was introduced with "dove" labelled "mouse" and "prober" labelled "prober-retaliator". Incidentally, an unfortunate choice of parameters resulted in there being no ESS at all, even though a fourth strategy "retaliator" was erroneously identified as such.

Sender strategies

A: send 1 when observing E; send 2 when observing L.

B: send 2 when observing E; send 1 when observing L.

C: never send anything.

Receiver strategies

A': act as if observing E when hearing 1; act as if observing L when hearing 2.

B': act as if observing E when hearing 2; act as if observing L when hearing 1.

C': ignore all received calls.

In the case of alarm calls, the payoffs for senders and receivers are very different. Senders will suffer a cost, because by calling they alert the predator of their presence and location. Evidence of the existence of a real cost in nature comes from the fact that alarm calls typically have very high pitch, which makes it more difficult for predators to locate the caller (Maynard Smith, 1982). The payoff matrix for the sender will therefore have all negative entries (parameter c) for strategies A and B, and (by definition) 0 for strategy C.

Receivers, on the other hand, will profit from a call *if and only if they correctly interprets it*. That benefit is quantified with parameter b. If the actual predator is a leopard, acting as if an eagle is observed can be a costly mistake: monkeys flee into the bushes to escape from an eagle attack, but that is in fact exactly where leopards hide (R. M. Seyfarth & Cheney, 1997). The cost of mis-interpretation is quantified as parameter m. If receivers ignore all calls, their payoff is 0 (again, by definition). The payoff matrices in this simple example will thus look as in figure 5.

	receiver strategy		
sender strategy $\downarrow$	A'	B'	C'
A	$-c$	$-c$	$-c$
B	$-c$	$-c$	$-c$
C	0	0	0

(a) sender's payoff

	receiver strategy		
sender strategy $\downarrow$	A'	B'	C'
A	$+b$	$-m$	0
B	$-m$	$+b$	0
C	0	0	0

(b) receiver's payoff

Fig. 5. Payoff matrices in a simple alarm call system

It is clear that neither A nor B can be the stable strategy for the speaker; if the cost of calling, c, is non-negligible, the strategy of not communicating at all, C, is always

optimal. In explaining the evolution of communication, we thus face a **problem of cooperation**: if the benefits of communication are for the hearer, the sender has no incentive to give away information, or to incur risk. Dawkins and Krebs (1978) pointed out this problem with what they call the "classical ethological" view on animal communication, which takes communication as existing for the benefit of the group. Dawkins and Krebs have therefore suggested that communication should be understood as a form of manipulation, with the benefits of successful manipulation with the sender.

Others (e.g. Maynard Smith, 1965; Sherman, 1977; L. Cavalli-Sforza & Feldman, 1983) have argued that "altruistic" communication can evolve through kin selection. However, the appropriateness of kin selection for human language – where communication is typically with non-kin – has been called into question (Dessalles, 1998). Dessalles has instead argued for a form of "reciprocal altruism", where there is a real benefit for the sender, because it is rewarded with status in the population. Fitch (2004) reviews his and other arguments, but concludes that they are not convincing. He posits the "mother tongue" hypothesis – that human language developed primarily in a context of kin communication – as one of a number of factors that shaped human language in its evolution, and calls for further exploration of the role of kin selection in language evolution. We will come back to kin selection, and some of the misunderstandings about it (e.g. that kin selection requires kin recognition) in section 7.

In other circumstances, for instance sexual signalling, the problem is not so much in the willingness to send signals, because the senders benefit, but in the **honesty** of the signals. A large amount of work on the evolution of animal and human communication has been concerned with this problem, leading to what is now called "honest signalling theory" (the handicap principle, Zahavi, 1975, 1977; Grafen, 1990; Lachmann, Szamado, & Bergstrom, 2001). Hence, the problem of cooperation is pervasive in work on the evolution of communication, although its instantiations differ with different assumptions on the costs and benefits of communication, for both sender and receiver. Although the problem of cooperation is a consequence of careful considerations of payoff, strategy sets and invasibility, we will, because of its importance, add it as a separate point to the list of criteria of evolutionary explanations:

Criterion 8 (Problem of cooperation) *Evolutionary explanations of the evolution of language need to address the problem of cooperation, and demonstrate that senders will be willing to send honest signals, and that hearers will be willing to receive and believe the signal.*

Even if we find a scenario where successful communication is in the interest of both the speaker and the hearer, there is another problem that arises from the frequency-dependence of language evolution. We could call this the **problem of coordination**. If we ignore the non-cooperative strategies C and C', how does a population

of players coordinate their behaviours such that they play either A and A', or B and B'? That is, how do they agree on a shared code? This problem seems particularly difficult when we consider a series of innovations, as in Jackendoff's (2002) scenario of the evolution of human language. Each of these innovations needs to confer a fitness advantage if it is to spread the population, but it is difficult to see how a genuine innovation can be advantageous to the individual if it is not shared by the rest of the population (Zuidema & Boer, 2003).

Lewis (1969) showed that only "perfect" communication systems are "separating equilibria", which, if the role of "rationality" of the players is replaced by natural selection, corresponds to evolutionary stable states (Skyrms, 1996; Trapa & Nowak, 2000; Rooij, 2004). Models in this tradition make the following assumptions:

- There is no cost to communication;
- The interests of sender and receiver are perfectly aligned;
- There is a discrete set of signals and a discrete set of meanings, and the number of signals equals the number of meanings;
- All meanings are equally frequent and valuable;
- Every "perfect" mapping from meanings to signals is equally good (which implies that meanings have no relation to each other, signals have no relation to each other, and meanings have no natural relation to signals);
- The meaning–signal associations are innate and inherited from parent to child.

It is easy to see why perfect communication systems are the only ESS's under these assumptions: if a communication system is sub-optimal, there must be synonymy: multiple signals are used for the same meaning. For the sender, however, it is always best to express a meaning m with the single signal s that has the highest chance of being understood, i.e. to avoid synonymy. The alternative signal(s) will thus not be used to express m anymore, and becomes available (through drift) for meanings that cannot be expressed yet. Hence, only "perfect" systems are stable against selection and drift.

It is clear, however, that all of these assumptions are violated in reality. Signals do have a cost, interests are not perfectly aligned, meanings and signals are not discrete, symbolic entities, but have similarity relations with themselves and each other, and, at least in human language, meaning–signal mappings are learnt and not innate. The problem of coordination thus remains a major open issue in the evolution of language, which we can add to the list of criteria:

Criterion 9 (Problem of coordination) *Explanations for the evolution of language need to deal with the problem of coordination, that is, show how, after each innovation, a shared code can be established and maintained.*

Much of the work on the evolution of language can be seen as dealing with this problem. A number of models, for instance, relax the innateness assumption above, and study, in computer simulations, the evolutionary success of a number of dif-

ferent strategies in word learning (Hurford, 1989; Oliphant, 1999; Smith, 2004). The payoff function in Hurford's model is the expected success in communication between a sender and a receiver (i.e. the game is cooperative; both sender and receiver benefit from success). Sender behaviour is characterised by a probabilistic mapping from a set of M meanings to a set of F signals; receiver behaviour by a probabilistic mapping from the signals to the meanings.

Hurford was interested in how these functions were learnt, and in the evolution of different learning strategies. The strategy set Hurford considered consisted of three strategies, termed imitator (that imitates the observed average sending and receiving behaviour in the population), calculator (that estimates the best send and receive functions based on observations of the population's receive and send behaviour respectively) and Saussurean learner (that chooses the same receive function as the calculator, but derives the send function from that receive function rather than from the receiving behaviour in the population). Hurford showed that Saussurean learners outcompete the other two learning strategies. These results were extended by Oliphant and Batali (1996), Oliphant (1999) and Smith (2004), among others. From these studies it emerged that learning strategies can evolve that give rise to "perfect" communication systems in a population.

Other models (e.g. Nowak & Krakauer, 1999), do not model such explicit learning rules, but do relax some of the other assumptions mentioned. More work is needed to study whether the results from these studies hold when learning is modelled explicitly. An encouraging result in this respect is due to Calvin Harley (1981). He studies the evolution of learning rules and showed that evolution will favour rules that *learn* the evolutionary stable strategy. Hence, results on Evolutionary Stable Strategies in innate communication systems, in principle carry over to situations where the same strategies are acquired in a learning process (Maynard Smith, 1982, chapter 4).

6 Levels of Selection

We have discussed some basic concepts from population genetics, which describe the change in frequencies of *genes*, and from evolutionary game theory, which describes the invasion and replacement of phenotypic *strategies* of individuals. The two approaches are obviously related, because the fitness of genes depends on the phenotypes they give rise to, and a strategy will only replace another strategy if all the genes necessary for that strategy are selected for and get established in a population. But the description of the evolutionary process in population genetics and evolutionary game theory are set at entirely different levels.

In Dawkins' (Dawkins, 1976) terminology, genes are *replicators*: they are the bits of information that get copied and transmitted – more or less intact – to the next

generation. Individuals are *vehicles* (Dawkins, 1976) or *reproducers* (Szathmáry, 1999). In sexual species, such as humans, a child is radically different from any one parent, because it inherits only 50% of the genes. Individuals, therefore, are not replicators, even though they are the obvious level of description when we talk about fitnesses and strategies.

If *replicators* and *reproducers* were the same objects, evolutionary dynamics would be relatively easy to describe. But in general, especially in sexual species, they are not. Genes are "packaged" – contained within the structured genome of an individual that lives within a structured population. That packaging makes the fate of a specific gene depend on the other genes it is associated with (genes that occur together more often or less often than would be expected on the basis of their frequencies alone, are said to be in *linkage disequilibrium*). If a gene a happens to be associated with a gene b that is under strong positive selection, gene a will increase in frequency even though it does not itself contribute to the fitness of its carrier ("genetic hitch-hiking", Hill & Robertson, 1966; Maynard Smith & Haigh, 1974). To predict the fate of a specific gene, we therefore need to know its statistical associations with other genes.

To make things even more complicated, not just the gene frequencies change; also the associations themselves change in evolution. The *physical linkage* between genes on a chromosome tends to keep these genes together, but *recombination* breaks up these associations. *Sexual selection* on the other hand generates associations between for instance, the preferences of the females and the selected traits of the males. Finally, *epistasis* also generates linkage disequilibrium, because if genes are much better in combination than they are apart, natural selection itself will make the combination more frequent than expected by chance. N. Barton and Turelli (1991) and Kirkpatrick, Johnson, and Barton (2002) have developed a mathematical framework to describe the dynamics of such *multi-locus evolution*; however, they take fitnesses as given and do not yet provide a bridge to the fitness concept in phenotypic models.

Hence, the relation between gene frequency change and adaptation at the level of the individual (such as language) is not at all trivial. The problem with the gene as the level of description is that we don't know the relevant fitness coefficients, because our knowledge of life, death and reproduction is almost entirely specified at the level of the individual. But the problem with the individual as level of description, is that we are not necessarily justified in assuming that natural selection corresponds to optimisation. Do the results from game-theoretic analyses translate to fitness coefficients of the genes that underlie the strategies? How do we relate the fitness coefficients, and the fundamental results about evolution as optimisation by Fisher and Wright, to adaptation on the level of individuals?

For the purposes of this paper, it would take too far to investigate the contributions of Grafen Grafen (2003) and others to relate population genetics and evolution-

ary game theory. However, a few important implications for language evolution research are worth making explicit. First, a "strategy" in a game-theoretic analysis will typically be coded for by many genes (*pleiotropy*). So if alleles $a_1, a_2 \ldots a_n$ at loci 1 to n are needed for an evolutionarily stable strategy A, we need each of these alleles to represent a step in the right direction. In technical terms, we need *additive genetic variance*; Maynard Smith (1982) argues that additive genetic variance is common in nature, and that this is therefore a reasonable assumption to make in game-theoretic analyses. We need to be aware, however, that we ignore all the phenomena of multi-locus evolution in game-theoretic analyses of language, issues that we ultimately need to deal with:

Criterion 10 (Levels of selection) *Explanations for the evolution of language need to relate selection at the level of individuals or groups to changes in gene frequencies. That is, they need to specify and relate the assumed levels of description for selection and heritability.*

Second, an important (methodological) observation is that there is no single best level of description; researchers make a heuristic choice about the level at which they will describe the evolutionary dynamics. Every model will only be an approximation, and it depends on the phenomenon of interest at which level the evolutionary process is most adequately described. Below, we will briefly discuss kin selection, and show, using the Price equation, why for the phenomena of social evolution the population structure is a crucial level of description that is left out in standard game-theoretic models.

7 Social Evolution & Kin Selection

The techniques from social evolution theory (Hamilton, 1964a, 1964b; Frank, 1998) could fill a whole separate paper; we will therefore keep the discussion brief. One fundamental equation, the **Price equation** (Price, 1970), is useful, however, to highlight a silent assumption in game-theoretic models, and to illustrate the issue of multiple levels of selection. We will follow Frank (1998) and Andy Gardner (p.c.) in the derivation of the Price equation. Like Wright's equation (9), it can be interpreted as describing the change in the frequency of a gene, but more generally it describes the change in the value of any trait z.

Price introduces his equation as follows:

> *"Gene frequency change is the basic event in biological evolution. The following equation [...], which gives frequency change under selection from one generation to the next for a single gene or for any linear function of any number of genes at any number of loci, holds for any sort of dominance or epistasis, for sexual or asexual reproduction, for random or nonrandom mating, for diploid, haploid*

or polyploid species, and even for imaginary species with more than two sexes" (Price, 1970, p.520)

We are interested in the change in frequency of a specific trait z in the population between the present ($\overline{z}$) and the next generation ($\overline{z}'$). If we divide up the population in M units of frequency $q_1 \dots q_M$ (these units are, for instance, individuals or groups, depending on the level of selection the equation is meant to describe), and we know their fitness $w_1 \dots w_M$ and their frequencies of trait z: $z_1 \dots z_M$, then the change of the trait's frequency in the whole population is given by:

$$\begin{aligned}\Delta\overline{z} &= \overline{z}' - \overline{z} \\ &= \sum_i q_i' z_i' - \overline{z} \\ &= \sum_i q_i \frac{w_i}{\overline{w}}(z_i + \Delta z_i) - \overline{z}\end{aligned} \tag{12}$$

The main diffference between Wright's and Price's equations is in the term Δz_i. This indicates a change in prevalence of trait z in group i, independent of the fitness of group i.

Multiplying both sides of this equation with $\overline{w}$, and rearranging gives:

$$\begin{aligned}\overline{w}\Delta\overline{z} &= \sum_i q_i w_i z_i + \sum_i q_i w_i \Delta z_i - \overline{w}\,\overline{z} \\ &= \underbrace{\sum_i q_i w_i z_i - \overline{w}\,\overline{z}}_{\text{Cov}_{[w,z]}} + \underbrace{\sum_i q_i w_i \Delta z_i}_{E_{[w\Delta z]}}\end{aligned} \tag{13}$$

As indicated, the terms in equation (13) correspond to the definitions of the *covariance* between fitness and trait value, and *expected value* of the product of fitness and change in trait frequency. Hence, the process of evolution can be elegantly summarised in the Price equation, as follows:

$$\overline{w}\Delta\overline{z} = \underbrace{\text{Cov}[w,z]}_{\text{selection}} + \underbrace{E[w\Delta z]}_{\text{transmission}} \tag{14}$$

The Price equation partitions the process of evolution into a term that describes the effects of selection (traits that are associated strongly with fitness will be selected for most effectively), and a term that describes the effects of (biased) transmission (the index i is the index of the parent; hence Δz_i describes the change in the trait value – from a particular parent to all its offspring – regardless of selection).

We will use the Price equation to investigate the evolution of an altruistic trait,

such as the alarm calls discussed in the previous section. Consider a population, subdivided in N groups $G_1 \dots G_N$, each of size M individuals. An individual has a level of altruistic behavior $z_{ij} \geq 0$. We can ask: under which circumstances will this trait evolve? In each group G_i, individuals benefit from the amount of altruism in that group, labelled as $z_i = \sum_j z_{ij}$; the total benefit is bz_i. The jth individual in that group, however, also suffers a cost from being altruistic, indicated with c; the cost is thus cz_{ij}. The fitness of the jth individual in the ith group is now given by:

$$w_{ij} = \alpha + bz_i - cz_{ij}, \tag{15}$$

where α is a baseline fitness (not dependent on the presence or absence of the altruistic trait). The fitness of the ith group is given by:

$$w_i = \frac{1}{M} \sum_j w_{ij} = \alpha + (b - c)z_i. \tag{16}$$

Hence, an individual's fitness (her relative contribution to the total offspring of the group) depends on the amount of altruism received and the amount of altruism given. If the cost c of being altruistic is larger than 0, it is always best for an individual to be selfish, as this cost is subtracted from its fitness. The group's fitness (the relative contribution of this group's offspring in the total offspring of the whole population) depends on the total amount of altruism given. If the cost c of altruism is lower than the benefit b, it is always best *for the group* if all individuals are altruistic, as this maximizes z_i.

The evolutionary process within each group i can be described with a Price equation, as in equation (14). If we assume there is no transmission bias, the equation simplifies to:

$$\overline{w_{ij}}\Delta\overline{z_{ij}} = w_i \Delta z_i = \text{Cov}_j[w_{ij}, z_{ij}]. \tag{17}$$

The evolutionary process at the level of the whole population is also described with a Price equation, where the transmission term concerns the within-group dynamics of equation (17):

$$\begin{aligned} \overline{w_i}\Delta\overline{z_i} &= \text{Cov}_i[w_i, z_i] + E_i[w_i \Delta z_i] \\ &= \text{Cov}_i[w_i, z_i] + E_i[\text{Cov}_j[w_{ij}, z_{ij}]]. \end{aligned} \tag{18}$$

Whereas equation 17 is not different from Price's equation (as changes in frequency only depend on fitness), equation 18 can only be formulated using Wright's equation. It takes into account both the change in group frequency due to group fitness, and models the change in frequency of trait z within each group as biased transmission.

The covariance in equation 18 can be replaced by a regression and variance term, because (by definition) $\text{Cov}(x, y) = \beta(x, y)\text{Var}(y)$, where $\beta(x, y)$ is the regression coefficient between x and y. This gives the following equation:

$$\overline{w_i}\Delta\overline{z_i} = \beta(w_i, z_i)\text{Var}_i[z_i] + E_i[\beta(w_{ij}, z_{ij})\text{Var}_j[z_{ij}]]. \tag{19}$$

These regression terms β can be read off directly from equations (15) and (16), because they correspond to the slope of the fitness functions, i.e. $\beta(w_i, z_i) = b - c$ and $\beta(w_{ij}, z_{ij}) = -c$. Substituting these values into equation (19) and rearranging gives:

$$\begin{aligned}
\overline{w_i}\Delta\overline{z_i} &= (b-c)\text{Var}_i[z_i] + E_i[-c\text{Var}_j[z_{ij}]] \\
&= (b-c)\text{Var}_i[z_i] - cE_i[\text{Var}_j[z_{ij}]] \\
&= b\text{Var}_i[z_i] - c(\text{Var}_i[z_i] + E_i[\text{Var}_j[z_{ij}]]) \\
&= b\text{Var}_i[z_i] - c\text{Var}_{\text{total}} \\
&= \left(b\frac{\text{Var}_i[z_i]}{\text{Var}_{\text{total}}} - c\right)\text{Var}_{\text{total}},
\end{aligned} \tag{20}$$

where $\text{Var}_{\text{total}}$ is the total variance.

This is equivalent to Hamilton's (Hamilton, 1964a, 1964b) result on kin selection, which says that an altruistic trait can evolve if the benefit b times the relatedness r is larger than the cost c:

$$br > c. \tag{21}$$

For the equivalence to be true, we have to take $r = \frac{\text{Var}_{i[z_i]}}{\text{Var}_{\text{total}}}$. If the benefits of trait z, weighted with the relatedness within a group, are larger than the costs, i.e. $rb > c$, then $\Delta\overline{z}$ will be positive, i.e. evolution will favour the trait even if it harms the individual.

The derivation using the Price equation highlights the correct interpretation of *relatedness* which is often misinterpreted. As this derivation shows, the relatedness term r is *not* the fraction of genes two individuals share (*identity by descent*), as is commonly assumed (e.g. Okasha, 2003). Rather, it is a statistical association between the trait of interest in one individual and the trait in the individual she interacts with. Therefore, the relatedness between two individuals can even be negative. This simply means that the individuals are less related to each other than to a random third individual in the population (Hamilton, 1970). If the association is high enough, altruistic traits can be favoured by natural selection. That is, if (for whatever reason) altruists are surrounded by other altruists, they benefit more from the altruism received than from the altruism offered (and conversely, if it is low

enough, natural selection can favour *spite* – behaviours that harm both the actor and the recipient; Hamilton, 1970; A. Gardner & West, 2004).

Interactions within kin-groups (and kin recognition) are an important mechanism for this association to arise (hence Maynard Smith's term "kin selection"), but not the only one. Subdivision of a population in groups is another mechanism (such "group selection" is thus a form of kin selection). Hamilton himself suggested a third mechanism, that of "green beards". If the same gene complex that codes for an altruistic trait, also codes for an external marker (i.e. a green beard), altruists can choose to interact preferentially with each other. This is of interest for language evolution, because language itself could be such a green beard, if related individuals can recognise each other based on features in their language. Sociolinguistic research has established that people are indeed more disposed to be friendly to people who have the same accent than to others ((referentie toevoegen)).

This would help people to cooperate more with people with whom they are closely related, aiding kin selection for cooperative behavior. Thus kin selection seems the most promising solution for the problem of cooperation that we introduced in section 5. It would certainly be worthwhile to study formal models of kin selection, that take into account the details of human communication.

8 Cultural Evolution

Dawkins (1976) emphasised that the principle of natural selection is not restricted to genes or individuals. In every situation where one can identify replicators, heritable variation and selection, a process of adaptation can take place. For instance, cultural inventions (or "memes", Dawkins, 1976) – religion, technology, fashion or indeed words and grammatical rules – undergo evolution if there are mechanisms for cultural transmission and cultural selection.

Since Dawkin's book, many wildly speculative theories have been launched under the heading "memetics", which have given this new field a bad reputation. Nevertheless, the basic idea is sound and open to serious investigation (Mesoudi, Whiten, & Laland, 2004). For a start, all mathematical models and criteria discussed in this paper apply to cultural evolution as well, although the Hardy-Weinberg (section 2) is not valid when blending inheritance occurs (such as would be the case in speech sounds, which can change continuously). The idea of viewing historical language change as a form of evolution is particularly attractive because, on the one hand, it makes the extensive mathematical toolkit of evolutionary biology available to linguistics, and on the other hand, is presents evolutionists with an enormous body of new data.

We need formal models of the cultural evolution of language, in which we can

deal with all the constraints on evolutionary models that I listed in this paper. Although many authors have noted the parallels between biological evolution and language change, including Darwin (1871, p.91), only recently have people started to study the cultural evolution of language in such a formal framework. Some relevant mathematical models are those of L. L. Cavalli-Sforza and Feldman (1981), Niyogi (2002) and Yang (2000). These authors look at the competition between two or more languages, with no qualitative differences between languages. Simulation models such as those of Kirby (1998) and Batali (2002) look at more open-ended systems, with more explicit formalisms for grammar and learning.

One problem is that is not so easy to decide on the appropriate units of selection. For instance, Kirby (2000) described the dynamics in his simulation model with context-free grammar rules as replicators under selection for more reliable replication. In later papers, however, he argued that the analogy between biological and cultural evolution in this case breaks down (Kirby, 2002). This is because the grammatical rules are *induced* from observable language, whereas in biological evolution genes are *inherited*, with no feedback from phenotype to genotype (other than through the effects of selection). This is known as the "central dogma of molecular biology". This observation is correct, of course, but it does not mean we cannot describe the dynamics in models such as Kirby's using the tools from evolutionary biology. The effects of induction in language change are a form of "directed mutation", and can be included, for instance, in the Price Equation in the transmission term. More work is needed to work this out with concrete examples.

9 Conclusions

In this paper we have discussed a variety of models from population genetics, evolutionary game-theory and social evolution theory. We have used these models to make a list of criteria for evolutionary scenarios of the biological and cultural evolution of language. Of course, no single study of the evolution of language (or any other biological trait) will be able to satisfy all "criteria". Rather, these criteria serve as *targets* for language evolution research. Hence, when confronted with a scenario for the biological or cultural evolution of language, we should ask:

(1) What are the units of inheritance and selection the scenario assumes? Genes? Memes?
(2) What is the scope of variation in these genes or memes? That is, what is the assumed set of possible traits/strategies available for evolution?
(3) What are the selection pressures? That is, what is the assumed payoff for each of these possible traits in each possible context?
(4) For every innovation in the scenario, will it indeed be favoured by selection when extremely rare? If not, is there a non-negligible chance it could get established by stochastic effects, or get frequent enough to be favoured by se-

lection?

(5) Does the assumed series of changes in the scenario indeed constitute a path of ever-increasing fitness? That is, is there a path of fit intermediates from start to finish?
(6) How much time will each of the innovations take to get established?
(7) Is there for every transition sufficient variation, but not too much?
(8) How does the scenario explain that speakers maintain the willingness to speak honestly, and that hearers continue to listen and believe the information received? That is, how does it solve the problem of cooperation?
(9) How does the scenario explain that speakers and hearers, after every innovation, agree on which signals refer to which meanings? That is, how does it solve the problem of coordination?
(10) How does the scenario relate dynamics at different levels of description – genes, strategies, individuals, groups, languages?

When asking these questions about published theories of the evolution of language, it quickly becomes clear that we are very far – from the point of view of mathematical evolutionary biology – of a satisfactory evolutionary scenario. Triumphant claims such as Pinker's (2000) that "game theorists have demonstrated the evolvability of the most striking features of language" should therefore be taken with a grain of salt. There is, as of yet, no complete and formal scenario for the evolution of *any* of the unique design features of human language (Hockett, 1960), that satisfies all criteria above. In particular, the questions about sufficient time and variation and the problems of cooperation and coordination have been ignored too often (although discussed game theoretic models represent significant progress on these issues).

Some have argued that a scientific theory of language evolution is simply not feasible, because there is too little data to select among many plausible scenarios (e.g. Chomsky, 1972; Lewontin, 1998). Interestingly, asking the questions above also shows that such pessimism is premature: rather than too many plausible scenarios for the evolution of language, we still have far too few that come close to meeting the criteria from evolutionary biology (as is probably the case for all major transitions in evolution, Maynard Smith & Szathmáry, 1995). Of course, it is possible that there really is a paucity of data, and that ultimately, there will be multiple scenarios of the evolutionary history of language that are coherent and consistent with the empirical facts. But only when we have precise scenarios of the evolution of language and worked out ways to test empirically the plausibility of one scenario against another, can we conclude – if that turns out to be the case – that a single, plausible account of the origins of natural language is beyond reach.

Acknowledgments WZ is funded by the Netherlands Organisation for Scientific Research (Exacte Wetenschappen), project number 612.066.405. BdB is funded by the Netherlands Organization for Scientific Research (Geesteswetenschappen),

project number 276-75-007.

References

Barton, G. E., & Berwick, R. C. (1987). *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Barton, N., & Partridge, L. (2000). Limits to natural selection. *BioEssays*, *22*, 1075-1084.

Barton, N., & Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics*, *127*(1), 229-255.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models.* Cambridge, UK: Cambridge University Press.

Briscoe, T. (Ed.). (2002). *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press.

Cavalli-Sforza, L., & Feldman, M. (1983). Paradox of the evolution of communication and of social interactivity. *Proc. Nat. Acad. Sci. USA, 80*, 2017–2021.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.

Chomsky, N. (1972). *Language and mind*. Harcourt, Brace and World. (Extended edition)

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished doctoral dissertation, University of Edinburgh, Scotland.

Christiansen, M. H., & Kirby, S. (Eds.). (2003). *Language evolution*. Oxford, UK: Oxford University Press.

Coyne, J., Barton, N., & Turelli, M. (2000). Is Wright's shifting balance process important in evolution? *Evolution*, *54*(1), 306-317.

Crow, J. F. (1999). Hardy, Weinberg and language impediments. *Genetics*, *152*(3), 821-825.

Darwin, C. (1859). *The origin of species – by means of natural selection or the preservation of favoured races in the struggle for life*. London: Murray. ((this edition, New York: The New American Library, 1958))

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray. (Reprinted in 1981 by Princeton University Press)

Dawkins, R. (1976). *The selfish gene*. Oxford University Press. (This edition 1989)

Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University Press.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: an evolutionary approach.* Oxford, UK: Blackwell Scientific Publications.

Deacon, T. (1997). *Symbolic species, the co-evolution of language and the human brain*. The Penguin Press.

Deacon, T. (2000). Evolutionary perspectives on language and brain plasticity. *Journal of Communication Disorders, 33*(4), 273-290.

Dessalles, J.-L. (1998). Altruism, status, and the origin of relevance. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases.* Cambridge, UK: Cambridge University Press.

Dessalles, J.-L. (2007). *Why we talk: The evolutionary origins of language*. Oxford: Oxford University Press.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567-631.

Dunbar, R. (1998). Theory of mind and the evolution of language. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases.* Cambridge, UK: Cambridge University Press.

Eigen, M. (1971). Self-organization of matter and the evolution of biological macro-molecules. *Naturwissenschaften, 58*, 465-523.

Fisher, R. A. (1922). On the dominance ratio. *Proc Roy Soc Edin, 42*, 321-431.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Oxford, UK: Clarendon Press.

Fisher, S., & Marcus, G. (2006). The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics, 7*, 9-20.

Fitch, W. T. (2004). Kin selection and "mother tongues": A neglected component in language evolution. In K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (p. 275-296). Cambridge, MA: MIT Press.

Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.

Gardner, A., & West, S. (2004). Spite and the scale of competition. *Journal of Evolutionary Biology*. (in press)

Gardner, R., & Gardner, B. (1969). Teaching sign language to a chimpanzee. *Science, 165*, 664-672.

Gold, E. M. (1967). Language identification in the limit. *Information and Control (now Information and Computation), 10*, 447–474.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology, 144*, 517–546.

Grafen, A. (2003). Fisher the evolutionary biologist. *Journal of the Royal Statistical Society: Series D (The Statistican), 52*, 319-329.

Haldane, J. B. S. (1932). *The causes of evolution.* New York: Longmans.

Hamilton, W. (1964a). The genetical evolution of social behaviour. i. *Journal of Theoretical Biology, 7*(1), 1-16.

Hamilton, W. (1964b). The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology, 7*(1), 17-52.

Hamilton, W. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature, 228*(277), 1218-20.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28*, 49-50.

Harley, C. (1981). Learning the evolutionarily stable strategy. *Journal of Theoretical Biology, 89*(4), 611-33.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 317–337). Oxford, UK: Oxford University Press.

Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.

Hill, W., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research, 8*(269-294).

Hockett, C. (1960). The origin of speech. *Scientific American, 203*, 88-111.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77*(2), 187-222.

Hurford, J. R. (2008). *The origins of meaning: Language in the light of evolution*. Oxford: Oxford University Press.

Hurford, J. R., Studdert-Kennedy, M., & Knight, C. (Eds.). (1998). *Approaches to the evolution of language: social and cognitive bases*. Cambridge, UK: Cambridge University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Science, 3*(7).

Joshi, A., Vijay-Shanker, K., & Weir, D. (1991). The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S. Shieber, & T. Wasow (Eds.), *Foundational issues in natural language processing* (p. 21-82). Cambridge MA: MIT Press.

Kirby, S. (1994). Adaptive explanations for language universals: A model of Hawkins' performance theory. *Sprachtypologie und Universalienforshung, 47*, 186-210.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language.* Cambridge, UK: Cambridge University Press.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form.* Cambridge, UK: Cambridge University Press.

Kirby, S. (2002). Natural language from artificial life. *Artificial Life, 8*(2), 185–215.

Kirkpatrick, M., Johnson, T., & Barton, N. (2002). General models of multilocus evolution. *Genetics, 161*(4), 1727-50.

Komarova, N., Niyogi, P., & Nowak, M. (2001). The evolutionary dynamics of grammar acquisition. *J. Theor. Biology, 209*(1), 43-59.

Lachmann, M., Szamado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences USA, 98*, 13189-13194.

Lai, C., Fisher, S., Hurst, J., Vargha-Khadem, F., & Monaco, A. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*(6855), 519-23.

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

Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science* (Vols. 4: Methods, models, and conceptual issues). Cambridge, MA: MIT Press.

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language, 48*, 839-862.

MacNeilage, P. F. (2008). *The origin of speech.* Oxford: Oxford University Press.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature, 201*, 1145-1147.

Maynard Smith, J. (1965). The evolution of alarm calls. *The American Naturalist, 99*, 59-63.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press, Cambridge, England.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet.Res., 23*, 23-35.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*, 15-18.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: W.H. Freeman.

Mesoudi, A., Whiten, A., & Laland, K. (2004). Perspective: is human cultural evolution darwinian? evidence reviewed from the perspective of the origin of species. *Evolution, 58*(1), 1-11.

Mitchener, G., & Nowak, M. A. (2002). Competitive exclusion and coexistence of universal grammars. *Bull Math Biol, 65*, 67-93.

Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. London: Continuum International Publishing Group.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA, 36*, 48-49.

Niyogi, P. (2002). Theories of cultural evolution and their applications to language change. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models.* Cambridge University Press.

Nowak, M. A., Komarova, N., & Niyogi, P. (2001). Evolution of universal grammar. *Science, 291*, 114-118.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*, 611-617.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA, 96*, 8028-8033.

Okasha, S. (2003). Biological altruism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy.*

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior, 7*(3/4).

Oliphant, M., & Batali, J. (1996). Learning and the emergence of coordinated communication. *Center for research on language newsletter, 11*(1), 1-46.

Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature, 348*, 27-33.

Pinker, S. (2000). Survival of the clearest. *Nature, 404*, 441-442. (News and Views)

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it. *Cognition, 95*(2), 201-236.

Price, G. R. (1970). Selection and covariance. *Nature, 227*, 520-521.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Blackwell.

Provine, W. (1986). *Sewall Wright and evolutionary biology*. Chicago, IL: University of Chicago Press.

Rooij, R. van. (2004). Evolution of conventional meaning and conversational principles. *Synthese (Knowledge, Rationality and Action), 139*, 331-366.

Roughgarden, J. (1979). *Theory of population genetics and evolutionary ecology: An introduction*. New York: Macmillan. (Reprinted 1987.)

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General, 115*(3), 211–235.

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science, 210*, 801-803.

Seyfarth, R. M., & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (p. 249-273). Cambridge, U.K.: Cambridge University Press.

Sherman, P. W. (1977). Nepotism and the evolution of alarm calls. *Science, 197*, 1246-1253.

Siegelmann, H., & Sontag, E. (1991). *Neural networks are universal computing devices* (Tech. Rep. Nos. SYCON–91–08). Rutgers Center for Systems and Control.

Skyrms, B. (1996). *Evolution of the social contract*. Cambridge, UK: Cambridge University Press.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology, 228*(1), 127-142.

Számadó, S., & Szathmáry, E. (2006). Selective scenarios for the emergence of natural language. *Trends in cognitive sciences, 21*(10), 555-561.

Szathmáry, E. (1999). The first replicators. In L. Keller (Ed.), *Levels of selection in evolution* (p. 31-52). Princeton, NJ: Princeton University Press.

Terrace, H. S. (1979). *Nim*. New York: Knopf.

Tomasello, M., & Bates, E. (Eds.). (2001). *Language development: The essential readings*. Malden, MA: Blackwell.

Trapa, P. E., & Nowak, M. A. (2000). Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, *41*(2), 172-188.

Truswell, R. (to appear). Constituency and bonobo comprehension.

Wang, W. S.-Y., Ke, J. Y., & Minett, J. W. (2004). Computational studies of language evolution. In C. Huang & W. Lenders (Eds.), *Computational linguistics and beyond* (p. 65-106). Academia Sinica: Institute of Linguistics.

Weinberg, W. (1908). Über den nachweis der Vererbung beim Menschen. *Jahresh. Wuertt. Ver. vaterl. Natkd.*, *64*, 369-382.

Wilke, C. (2005). Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology*, *5*, 44.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the sixth international congress on genetics* (p. 355-366).

Yang, C. D. (2000, October). Internal and external forces in language change. *Language Variation and Change*, *12*(3), 231-250.

Zahavi, A. (1975). Mate selection - a selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214.

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology*, *67*, 603–605.

Zuidema, W., & Boer, B. de. (2003). How did we get from there to here in the evolution of language? *Behavioral and Brain Sciences*, *26*(6), 694-695.