Prototype Categorisation and the Emergence of a Lexicon in an Infinite World

Cyprian Laskowski

Linguistics and English Language

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

August 2006

Declaration

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

Cyprian Laskowski

Abstract

One of the least understood issues in language evolution is how hominins were able to ground and establish a shared lexicon. Recently, researchers have explored this issue using a variety of computational models, whose results have suggested that a shared lexicon could have emerged spontaneously through a process of self-organisation. However, these models have used psychologically unrecognised concept representations and an oversimplified environment. In this dissertation, I present a new computational model in an attempt to address these problems. Agents' category representations are inspired by prototype theory, having central members and graded membership. The environment consists of an infinite number of objects, and has a probabilistic structure which can be easily manipulated through model parameters. Despite the relatively complex model, simulation results are generally in line with previous ones and add further support to the self-organisation hypothesis. In addition, the speed and level of lexical convergence depend on the world structure, confirming that this is an aspect of past models which has seen too little attention. Future work should investigate the vast parameter space in further detail, and extend the simulations in various new directions.

Acknowledgements

This dissertation would have never materialised without the suggestions, advice, and encouragement of my supervisors, Jim Hurford and Andrew Smith: regardless of my outlook prior to our meetings, I invariably came out enthusiastic, focused and enlightened.

My parents and sibling have always stood behind me in good times and bad, and, well, they are simply awesome.

My fellow lab-dwellers Dave, Hannah, Justin, Karola, Mike, Sebastian, Stephen and many others offered useful feedback, humming tolerance, pleasing apples, and, refreshing midnight frisbee interludes on Buccleuch Place.

My former professor, Rose-Marie Déchaine, first got me interested in linguistic theory, encouraged me to apply for postgraduate studies, and recommended the University of Edinburgh to me, and vice versa.

But most importantly, my wonderful wife Megumi has somehow put up with my presence at home for an entire hour or so daily, only to see me feast on the delicious meals that she consistently concocted, take in her warm smile, and run frantically back to the lab.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Stages in Language Evolution

The evolution of language presents science with a very difficult problem, since it does not leave clear evidence in the fossil record (Davidson 2003), it does not have any close counterpart in animal communication (Hauser 1996), and it is a result of the complex interaction of at least three different dynamic systems operating on massively different timescales (Kirby 2002). Nevertheless, as a large body of recent research demonstrates, it is possible to study it indirectly from a wide variety of perspectives (Christiansen & Kirby 2003).

One of the key ideas that has emerged, although not unopposed (Lanyon 2006), is that language evolved in stages (Bickerton 1990). Jackendoff (1999) expanded on this proposal, postulating 11 different stages, which by hypothesis correspond with periods in hominin evolution. The earliest of these stages featured a very simple language, consisting of holistic utterances with no symbol concatenation or syntax. Although there are various ways of imagining the nature of such utterances (Wray 1998), in this dissertation I will assume, following Hurford (forthcoming), that the earliest words referred to objects and events in the world, and were understood situationally with the help of pragmatic inference.

In order for such a one-word stage to emerge, various preadaptations would had to have been in place (at least partially), including the physiological capacity to produce

and perceive signals, the ability to understand the intentions of others, and the motivation to communicate with others in the first place (Hurford 1999). Although all such preadaptations are very important and require further research, in this study I will focus on the last: the ability to associate non-situation-specific word forms with their referents in the world. In addition, I will assume, for the sake of concreteness that language had vocal origins, although arguments have also been made for a gestural route (Corballis 2003).

1.2   Words, Concepts, and Evolution

However, the relationships between word forms and referents are conventional and arbitrary, with no intrinsic association between them. Consequently, in order to use a word like "tree", it is necessary for the word form to be attached to something else in the mind. But obviously we cannot do so by planting trees in our brain; instead, we do so by forming a concept of a tree, which is associated with both the word "tree" and actual trees in the world. Thus concepts mediate the relationships between word forms and their referents (Pierce 1955).

Unfortunately, this need for conceptual intermediacy introduces two problems. First of all, as manifested by cross-linguistic differences, there are different possible ways to divide the world up into conceptual categories, even in such fundamental domains as space (Majid et al. 2004). Although such differences are surmountable in modern language-entrenched societies (except perhaps for second language learners), they are likely to have been significant among a group of prelinguistic hominins who were inventing and conventionalising their first words. Secondly, a conceptual capacity in animals and even prelinguistic human infants is often questioned, and some philosophers even argue that concepts cannot exist without words (e.g., Davidson 2001). If this is the case, then explaining how words first became established becomes much more difficult, unless one replaces concepts with something less sophisticated, such as perceptual categories (Barsalou 1999) or "proto-concepts" (Hurford forthcoming).

However, even if we grant that hominins were sufficiently equipped to use a holistic language on an individual level, there still remains a major mystery: how could a population of hominins agree on the meanings of their first words, without the help of language

(as occurs in dictionaries)? In particular, how did the meanings of the earliest words get grounded (Harnad 1990)?

Over the last twenty years, many researchers have been investigating this issue with the use of computer models (Steels 1997, Hurford 1989, Vogt in press). Having equipped abstract agents with certain cognitive abilities and exposed them to a simulated environment, many simulations have suggested that the emergence of a coherent lexicon can occur through a process of self-organisation (Ashby 1947). However, these studies are typically done using representations of the environment and of agents' categories that are unrealistic and oversimplified, making the agents' task too easy.

1.3 Dissertation Outline

In this dissertation, I try to address this concern by arguing for and designing a model with a more complex environmental structure and a more psychologically plausible category representation. I will first review the psychology literature on concepts, followed by an examination of past computational models. Then I will motivate and present my new model, and present some preliminary results. Finally, I will discuss the results and their relevance to language evolution.

CHAPTER 2

Concepts

2.1   Function

Categorisation is a fundamental cognitive skill of both humans and animals that helps make sense of the world. As Mervis & Rosch (1981) note, "without any categorisation an organism could not interact profitably with the infinitely distinguishable objects and events it experiences" (p. 94). In fact, not only do we constantly encounter new things in our environment, but even familiar things appear and behave differently in different circumstances. No two stimuli are ever exactly the same.

To illustrate, when we see a lion in the savanna, we do not treat it as if it were a completely new and unfamiliar object; rather, we recognise it as a lion, infer information about it, and respond accordingly. The same is true of animals: indeed, if antelopes, for example, couldn't distinguish lions from other things in their environment, then they would not have survived for very long in the competitive world of natural selection.

However, it is important to be cautious about drawing premature conclusions from categorisation abilities. In particular, a creature's ability to group two stimuli into the same category does not necessarily imply that it has a mental concept of that category. For instance, frogs instinctively lash out their tongues to catch flies (as opposed to leaves or other small things in the air). As Hurford (forthcoming) argues, this shows that their brains are in different states for the two kinds of stimuli. Yet it does not necessarily prove that there are sophisticated human-like mental representations or images in the frog's mind (Herrnstein 1991).

2.2 Features

Although it is clear that both animals and humans categorise stimuli, it is much less obvious exactly on what basis they do so. Intuitively, categories should be based on some kind of similarity, so that objects which are more similar to each other tend to be grouped together. But similarity is a vague and abstract term: precisely what are the criteria that are used to judge the degree of similarity between two things?

Perhaps the most natural candidates for categorisation lie in perception (Barsalou 1999): stimuli are similar insofar as they share perceptual features. For instance, apples can be easily distinguished from bananas by their size, shape, colour, smell and taste: all of these are properties that can be perceived through our senses. And indeed, these may be the only criteria available to infants: "From the infant's perspective, all that can be used to distinguish the categories is evidence that is readily perceptible in the immediate environment" (Murphy 2002, p. 293). Perceptual features are also heavily involved in animal categorisation, such as that of pigeons (Watanabe et al. 1995).

However, perceptual features are often insufficient, as many categories also require conceptual knowledge (Mandler et al. 1991). Furniture, for example, comes in all kinds of shapes, sizes and colours, and there are few perceptual features, if any, that distinguish furniture as a group. What they do share in common is their general function, an understanding of which requires experience and knowledge about the world. It is thus not surprising that there appears to be a shift in infant categorisation from being primarily perceptual to a mix of perceptual and conceptual (Oakes et al. 1997).

The distinction between conceptual information, perceptual features and concepts themselves is not as clear-cut as it may first appear, however. If we look more closely at commonly posited features, they almost invariably turn out to be based on preexisting concepts. For instance, suppose we characterise dogs as having features like "has four legs", "barks", "salivates", "wags tail", and "has large ears". All of these features are cognitively complex, and, in fact, seemingly composed of concepts themselves (Taylor 1995). This recursive problem is even more clear for conceptual information used in categorisation, such as "used at home" (a possible feature of furniture). One potential solution is offered by Schyns et al. (1998), who suggest a dynamically expanding set of

features, in which infants start with very low-level perceptual features, from which more complex features are created when required.

It is worth mentioning a potential source of confusion with feature terminology. Mervis & Rosch (1981) pointed out that one can make a distinction between representing features as discrete (e.g., [size=large]) or continuous (e.g., [size=0.3]). Discrete characterisations (often called features) are more intuitive, but they presuppose a "large" category. On the other hand, continuous representations (sometimes called dimensions) are more difficult to understand, but are in a relatively raw, unprocessed form. Nevertheless, it is generally possible to translate between the two representations (Smith & Medin 1981). In this dissertation, I will focus on a continuous "dimension" representation, but will use the term "feature" interchangeably when it seems more natural.

2.3 Structure

Even assuming that we can accurately identify an object's perceptual features, however, the question remains of how we use these features to categorise it. For instance, when looking at an orange, how does one determine that it's an orange, as opposed to a grapefruit, a plastic toy orange, or a small orange ball? Several theories have attempted to address this issue.

2.3.1   Classical Theory

In the classical theory, which can be traced back to Aristotle (Apostle 1980), concepts are represented as strict sets of necessary and sufficient conditions. This implies that all members of a category are equal, and that category boundaries are rigid. For instance, triangles are defined as three-straight-sided geometric figures, and it is normally trivial to decide whether or not a figure is a triangle. The elegance of the classical view is that determining category membership is a clean and unambiguous binary decision.

However, although efficient in principle, this theory has severe problems. Most seriously, for many concepts, it is virtually impossible to find solid definitional conditions that fit every member while excluding every non-member. In attempting to find a definition for the concept of games, Wittgenstein (1953) argued that it was impossible and that concepts were characterised instead by "family resemblances". Indeed, his argument

applies to just about any concept. Although classical categories are generally thought to work best in technical fields in which definitions are necessary, even the applicability of technical categories like "planet" and "strike" (in baseball) are sometimes unclear (Murphy 2002), and the same is even true of triangles which have been drawn sloppily. Consequently, the classical theory "has simply ceased to be a serious contender in the psychology of concepts" (Murphy 2002, p. 38).

2.3.2 *Prototype Theory*

In the 1970's, Eleanor Rosch and her colleagues followed up on Wittgenstein's (1953) argument and argued for a different psychological basis for categories, which has become known as prototype theory (e.g., Rosch 1978, Mervis & Rosch 1981). In stark contrast to the classical view, the emphasis in prototype theory is on the categories' centres rather than their boundaries. The best examples of categories are their prototypes, and the membership of an object depends on its degree of similarity to a category's prototype. Consequently, categories have unequal members, graded structure and fuzzy boundaries. For example, in the category of birds, sparrows and robins are normally typical members, while ostriches and penguins are peripheral.

Rosch's claims were supported by a large body of empirical studies. Many of these studies showed that "the prototypicality of items within a category can be shown to affect virtually all of the major dependent variables used as measures in psychological research" (Rosch 1978, p. 38), such as reaction time, speed of learning of new categories, order of category learning, order of spontaneous listing, and effect of priming (e.g., Rosch 1975, Rosch, Simpson & Miller 1976, Rosch, Mervis, Gray, Johnson & Boyes-Braem 1976).

Rosch also investigated goodness of category membership, by asking subjects to assign values on a scale to items (e.g., Rosch 1975). For instance, 200 college students were asked to rank items of furniture on a scale of 1 (high) to 7 (low), and there was generally good agreement between the subjects: out of sixty items investigated, chairs and sofas had the highest average scores (1.04), telephones the worst (6.68), with buffets and lamps in the middle (2.89 and 2.94, respectively).

Rosch also claimed that using prototypes is cognitively efficient (Rosch 1978), appealing to two principles. First, the category system should try to maximise information while

minimising cognitive effort. That is, it should form enough categories to allow for distinctions between stimuli, but not so many as to make unnecessary distinctions. Second, the world is not structured in an arbitrary way, but rather contains highly correlated structure. In particular, not all possible combinations of features occur in the world, some being much more probable than others. For instance, things with feathers tend to have beaks and to fly, while things with fur tend to have mouths and walk (Mervis & Rosch 1981). An efficient category system should exploit these correlations. In particular, knowing what category something belongs to should maximise what we know about its features, and, conversely, knowing a single feature of something should suggest what category it belongs to (what Rosch calls "cue validity"). Rosch argues for the usefulness of prototypes in such a structured world, concluding that they "appear to be just those members of a category that most reflect the redundancy structure of the category as a whole" Rosch (1978)(p. 37).

There are several issues that are problematic or at least vague in prototype theory, however. Perhaps most importantly, it is not clear what exactly is the nature of the prototype. As Taylor (1995) points out, a prototype can be thought of as either the central member (or members) of a category, or as "a schematic representation of the conceptual core of a category" (p. 59). Taylor (1995) and Murphy (2002) both argue for the latter, and in fact, Rosch & Mervis (1975) appears to favour it as well. In addition, what does it mean for something to have partial category membership? Lakoff (1987) draws a distinction between degree of membership and goodness of example, claiming that while sparrows may be better examples of birds (and thus more prototypical) than penguins, both are full-fledged members of the bird category. Croft & Cruse (2004) raise a similar issue with respect to fuzzy boundaries, arguing that an individual's category boundaries are sharp at any given time, although they may shift from time to time and vary from person to person. Finally, as Rosch (1978) herself points out, although experiments have revealed the pervasiveness of prototype effects in human categorisation, prototypes "only constrain but do not specify representation and process models" (p. 41).

Unlike its classical predecessor, prototype theory has, so far, generally stood the test of time. Although it has not undergone significant theoretical developments recently, it is still considered one of the major theories of categorisation (Murphy 2002). Moreover, prototype effects continue to be found in recent studies of categorisation in human adults

(Smith & Minda 1998), children (Hayes & Taplin 1993) and prelinguistic infants (Rubenstein et al. 1999), as well as, more debatably, in animals such as baboons (Depy et al. 1997) and even pigeons (Fagot 2000) (although these effects can sometimes be explained by alternative theories as well). Also, prototype theory has been applied not only to basic object categories (as in most of Rosch's studies), but also events (Croft 1990), adjectives (Smith et al. 1988), and even "ad hoc" categories like "things to take on a picnic" (Barsalou 1983).

2.3.3   Other Theories

Although still a major force in psychology, prototype theory has been seriously challenged by several theoretical alternatives, and is by no means the uncontested champion of concept theories. In this section I introduce two of the most dominant alternative proposals: exemplar theory and theory theory. However, since the computational model I develop in Chapter 4 is based on prototype theory, I will only give a very brief overview of these approaches here.

Exemplar theory is based on the originally revolutionary idea that concepts are represented in terms of remembered exemplars rather than summary representations (as assumed in both classical and prototype theory) (Medin & Schaffer 1978). As in the prototype view, categorisation depends on similarity, but in this case, a stimulus is compared to the individual exemplars in each category, rather than only the prototypes. In this theory, typicality effects thus derive from an item being similar to many exemplars of a certain category as opposed to other categories (Nosofsky & Palmeri 1997).

Theory theory emphasises that our individual concepts cannot be separated from our general knowledge of the world (Keil 1992, Murphy & Medin 1985). People have naive "theories" about the world which they use to categorise stimuli. These theories help decide which of an infinite number of possible features of objects is relevant to membership in a given category (Murphy 2002). They also facilitate the use of categorisation criteria that do not have clear featural bases. For instance, deciding whether or not something is a weapon depends primarily not on simple features, but on judging how effective it would be in hurting people (Barsalou 1985).

Although the classical theory is clearly inadequate, "there is no clear, dominant winner" from among the prototype, exemplar and theory theories (Murphy 2002, p. 488). For many phenomena, more than one theory can provide explanations in different but equally satisfying (or unsatisfying) ways. Indeed, Barsalou (1990) has argued that prototype and exemplar theory are indistinguishable. A full account of categorisation would seem to require an integration or compromise of the different theories. Indeed, recent attempts at doing this can be found in both theoretical and empirical research (e.g., Smith & Minda 1998, Murphy 2002).

2.4   Origin

2.4.1   *Humans*

Many researchers argue that human concepts and knowledge are at least partly innate. Fodor (1998) claims that humans have a system of basic innate concepts, which serve as foundations for the building of more complex concepts. Shepard (1994) argues that such innate concepts are evolutionarily driven, as individuals with certain innately internalised concepts would be favourably adapted to certain environments. Spelke (1994) overviews initial knowledge in infant cognition and argues that much of it is innate and triggered early in development, such as the ability to perceptually single out objects in the environment.

Of course, these innate concepts need to be supplemented by conceptual development ontogenetically in order to explain the full range of human concepts. The role of the environment is clearly crucial to much of conceptual development (Elman et al. 1996), as there are clearly cases which have no precedent in our evolutionary past and require experience to be learned (e.g., telephones). Human children not only develop more complex concepts through their interactions with the physical world, but also through their social interactions. (e.g., Keil 1992). Indeed, language itself appears to play an important role in conceptual development and thought (Bowerman & Levinson 2001, Gumperz & Levinson 1996).

2.4.2   Animals

Animal concepts are even trickier to study than human ones, largely because they are difficult to probe linguistically. In fact, it is disputed whether animals even have concepts at all (Davidson 2001). However, there is plenty of evidence that animals do have knowledge about things in the world (e.g., Tomasello & Call 1997), and whether or not we attribute the term "concepts" to this knowledge is largely a matter of definition (Hurford forthcoming). Indeed, Herrnstein (1991) presents a scale of different degrees of categorisation ability, ranging from the instinctive behaviour of venus flytraps to fully sophisticated abstract human concepts.

Compared to humans, however, a large proportion of animal concepts appears to be innate. For instance, new-born chicks, upon their first exposure to light, already have an instinct to approach objects which exhibit biological motion (i.e., living things) (Vallortigara et al. 2005). Similarly, infant vervet monkeys can produce and appropriately react to signals corresponding to different kinds of predators (Seyfarth & Cheney 1990).

On the other hand, animals are also capable of learning some new concepts or refining innate ones (e.g., Seyfarth & Cheney 1986). Various attempts have been made, with mixed success, to see if animals could learn abstract concepts, such as sameness, transitivity and number (Hurford forthcoming). The most sophisticated examples of learned concepts in animals are probably those of enculturated chimpanzees, who acquire a wide variety of knowledge and abilities that they do not have in the wild (Tomasello & Call 2004). However, even species which are phylogenetically distant from humans, such as parrots, have been shown capable of learning basic abstract relations (Pepperberg 1999).

2.5   Lexicalisation

It is important to distinguish conceptual abilities from signal-meaning associations (Hurford forthcoming). How do human children associate words to concepts, and to what extent can animals achieve this?

2.5.1 Humans

As dictionaries show, it is relatively easy to learn new words once a basic vocabulary has been developed. On the other hand, grounding one's first words and thus successfully overcoming Quine's (1960) dilemma is quite a feat. And yet virtually all human children do it all the time, producing their first words at roughly 1 year of age (Clark 2003) and beginning to understand their parents even earlier (Tincoff & Jusczyk 1999). By adulthood, humans effortlessly use tens of thousand of words, often in multiple languages.

In order to explain this, various word learning constraints and mechanisms have been postulated, supported by psychological studies with young children. The whole object bias states that children assume that words refer to entire objects, rather than their properties or parts (Macnamara 1972). According to the taxonomic bias, children believe that words refer to object kinds (taxonomic relations) rather than themes (thematic relations) (Markman & Hutchinson 1984). The shape bias claims that children are more prone to associate a word with things which have the same shape rather than some other common property (Landau et al. 1988). The mutual exclusivity bias (Markman 1989) and the principle of contrast (Clark 1987) are related proposals, based on the idea that children will initially assume that different words refer to different things. And Tomasello (2000) has emphasised the role of children's socio-cognitive abilities, and especially their understanding of others intentions, as important keys to word learning. With respect to all of these constraints, however, it is important to note that it is not clear whether they are specific to word learning or derive from more general cognitive mechanisms (Bloom 2000).

Finally, we should note that in the evolution of language, word invention must have initially been equally important to word learning. Since modern human children can rely on the preexistence of language vocabularies, very little is known on this subject, but children's invention of "proto-words" (Halliday 1975) and their use in parent-child interaction (Camaioni et al. 2003) have been documented.

2.5.2 Animals

In considering word learning in animals, a distinction should be made between signals in animal communication systems and the words of human language.

Although signals in animal communication systems, such as those of vervet monkeys, are mostly innate, they can be ontogenetically refined. For instance, young vervet monkeys sometimes produce a snake alarm call upon seeing a long thin stick, but such mistakes disappear over time (Seyfarth & Cheney 1980). Also, there are some signals, such as whale songs, which are culturally learned by individuals and transmitted to others (Tyack & Sayigh 1997). In addition, in some primate species, individuals listen to the vocalisations of others in different situations and pragmatically infer information from them (Seyfarth & Cheney 2006). According to Pika & Liebal (2006), apes also develop gestures and communicate intentionally with them, although these differ significantly from those of human children in that they are always used dyadically and imperatively.

As for animals trained in human language, the most successful have generally been apes, although surprisingly good results have also been found with parrots and dogs. Sue Savage-Rumbaugh and her colleagues have engaged in long term enculturation projects with bonobos and chimpanzees (Savage-Rumbaugh et al. 1993, 1990). They have argued that the symbol learning and production of one of their bonobos in particular occurred spontaneously and parallelled that of human children quite closely. Pepperberg (1999) has trained parrots using a competitive paradigm, and one of them can not only linguistically label objects, but even use and understand relational terms, such as "shape" and "colour". Finally, Kaminski et al. (2004) have reported a case of a dog who can understand about 200 human words and can apparently learn and retain new words from a single exposure using mutual exclusivity.

CHAPTER 3

Computational Modelling

3.1 Motivation

As mentioned in Chapter 1, the difficulties of studying language evolution have resulted in the innovation of a variety of alternative indirect techniques. One of the more fruitful paradigms over the last twenty years has been computational modelling. But what is the relevance of modern computers to complex real world phenomena such as the evolution of language?

Figure 3.1 shows how computer modelling can be used, based on a general argument made by Noble (1998). Suppose we wish to explain a real-world phenomenon of interest E_R, and we hypothesise that it emerges from a certain set of low-level assumptions, A_R. If we cannot test this hypothesis directly (as with language evolution), then a computational framework provides an alternative. We model the assumptions A_R with computational counterparts, A_M. We then explore the effects of A_M by running simulations with them, obtaining a simulated phenomenon E_M. Returning to the real world, our theory (i.e., $A_R \rightarrow E_R$) can be potentially supported insofar as E_M resembles E_R in its salient aspects. Notice that the process is cyclic: the results from simulations can refine our understanding of the phenomenon, leading us to modify our hypothesis, etc.

3.2 Dangers

Yet just as this argument highlights the potential of computational modelling, it also reveals some of its dangers. In particular, there is usually an implicit assumption in

15
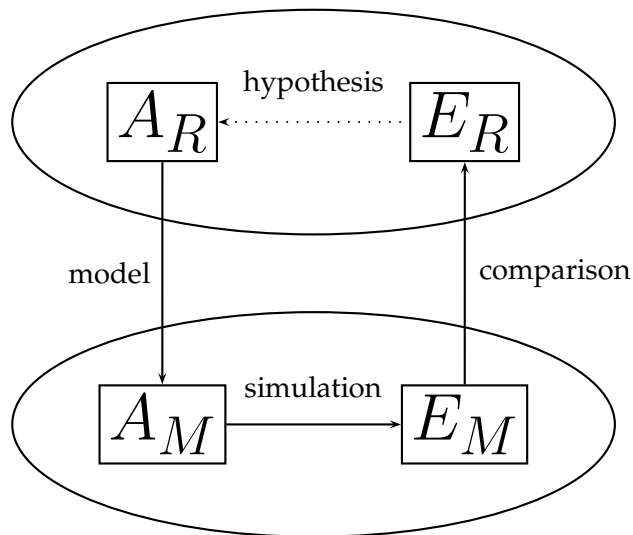
Figure 3.1: Computational modelling diagram: A visualisation of Noble's (1998) justification of computational modelling. E_R, the real-world phenomenon under study, is first hypothesised to be the result of A_R, a set of low-level real-world assumptions. These assumptions are modelled as A_M, and then simulations are run to obtain results, E_M. If E_M and E_R are sufficiently similar, then the hypothesis is supported.

computational models that the mappings between the model and the world, $A_R \rightarrow A_M$ and $E_M \rightarrow E_R$, are simpler than the strictly in-world mapping, $A_R \rightarrow E_R$. Such an assumption is perhaps necessary to justify the use of computational models, but it cannot be taken for granted. After all, computer models need to abstract and simplify the phenomena they investigate, and it's far from obvious that these transformations are inconsequential. Therefore, while computational models are good tools for exploring language evolution, their interpretation should be treated with great caution.

3.3 Previous models

Since the first attempts nearly twenty years ago (Hurford 1989), a large number of computational models has addressed a variety of phenomena in language evolution, such as phonological systems (de Boer 2000, Oudeyer 2006) and compositionality (Kirby 2000, De Beule & Bergen 2006). In this section, I will overview previous modelling work related specifically to "lexical emergence": that is, the evolutionary linguistic transition to a one-word stage, as described in Chapter 1. In accordance with my emphasis on

the conceptual bases of words, I organise this overview mainly in terms of conceptual representations.

3.3.1 Discrete concepts

Many models of language evolution, especially early ones, represent agents' concepts as atomic, innate and universal. That is, all agents are equipped with the same set of discrete meanings.

The earliest example of a lexical emergence model, Hurford (1989) used fixed, shared sets of both concepts and words. The agents' task was to converge on the same mappings between the two sets. There was no representation of an external environment, as the agents only interacted with each other. Hurford's main focus, however, was on a comparison of learning strategies. In fact, the focus of most models that employ discrete, atomic concepts is primarily no on lexical emergence itself, but some related phenomenon. For instance, Smith (2002) studied different word learning biases in a connectionist model, and their effects on the learning, maintenance, and construction of a lexicon. Models which explore the emergence of compositionality or the evolution of syntax also tend to use atomic meanings as components in the building of predicate logic utterances (e.g., Kirby 2000).

Steels (1996*a*) presented one of the first models in which concepts were not atomic, and there was no predefined set of word forms. Meanings were represented as vectors, with each element consisting of a feature type and an associated discrete value. There were nine possible meanings, one corresponding to each combination of feature values. Agents developed a shared lexicon by playing "language games", a paradigm that has been used in many subsequent models.

It is important to note that in Steels's (1996*a*) simulations, referents and concepts are indistinguishable. Agents perceive objects directly and do not need to convert them into internal category representations. Since the agents are exposed to the same set of objects, this means that in effect they trivially have identical conceptual systems consisting of the same set of discrete meanings. This makes the agents' task unrealistically simple. In order to overcome this problem, it is necessary to separate reference from meaning.

3.3.2   *Discrimination trees*

Steels (1996*b*) presents a model in which agents do have private internal category systems, and categories are grounded in interactions with an external environment of objects. Objects are represented as vectors of dimension values, with each dimension representing an abstract feature, and each value being a number in a continuous range. Consequently, rather than having discretisations provided to them by the model, agents must themselves divide a continuous world into categories.

Agent category systems are represented as binary "discrimination trees" in each dimension, with the top node representing a dimension's entire range of $[0, 1]$, and each other node partitioning its parent node exactly in half. Agents develop their categories through "discrimination games", in which they attempt to find or create categories that can distinguish one object in the environment from the others. Thus, an agent's category system gets more refined as it experiences more and more discrimination games.

Steels (1997) integrated the discrimination and language games, so that communicating about an object requires categorising it in a discrimination game, and this framework has been subsequently developed by various researchers. Van Looveren (2005) uses it to explore the transitions between key stages in language evolution. Kaplan (2000) compares and analyses lexical emergence in various models of different degrees of complexity. Smith (2003*a*) investigates how lexical emergence is affected when hearers get no explicit feedback from speakers, as well as the relationship between category system complexity and communicative success rates.

Although Steels' framework does satisfy the need to make agents' conceptual systems private and environmentally driven, there is still a major problem. The categories of the agents in Steels' discrimination trees are entirely of the classical genre: category membership of an object is a black-and-white issue, determined by whether or not an object is within the category's sharp boundaries. Models that are more in line with the psychology literature would have more credibility.

3.3.3 *Prototype representations*

Compared to models based on the classical view, prototype-based models of lexical emergence are relatively scarce. Vogt (2000), in a robotics paradigm, used a model in which categories had associated prototypes, which were represented as points in the feature space. Category membership of a stimulus is obtained by finding the prototype in one's category system which is closest to it. New categories can be created by adding new prototypes to the space, which effectively splits a previous category in two, much as in Steels's (1996*b*) discrimination trees.

Although there is a sense here in which there are privileged members, and stimuli are categorised by considering their degree of similarity to all the prototypes, this model is actually more classical than prototypical. Category membership is still an all-or-nothing judgement, and there is no sense in which categories overlap each other: rather, they partition the space along clear-cut boundaries determined by the set of prototypes and the similarity measure.

Belpaeme (2002) uses a relatively complicated prototype model that does incorporate the notions of graded membership and overlapping category boundaries. He represents each category as a radial basis function network (Hassoun 1995), which allows for each category to have 1 or more prototypes with varying weights (although in a subsequent model he uses a simplified representation with a single prototype: Belpaeme & Bleys 2005). A prototype is represented with a locally tuned unit, and a stimulus' similarity to the prototype is obtained by means of a multivariate gaussian function centred on the prototype. Although this system is flexible and powerful, the widths of the locally tuned units are fixed, so that the relevance of all features is constant across all categories. This idealisation is reasonable in Belpaeme's (2002) model, which is primarily concerned with basic colour categories, but is not ideal in general, as it cannot easily handle hierarchical relationships between categories.

These prototype models use representations that are certainly more psychologically plausible than their classical counterparts. Nevertheless, they still lack the idea of some features being more sensitive to others, and, in the case of Belpaeme's models, are specialised for a specific purpose.

3.3.4   Other representations

Although they are not the focus of this study, it is important to point out that atomic, classical and prototype models are not the only kinds of category representation that have been used in language evolution models. For example, de Jong (1997) used an adaptive subspace discrimination method, which fared better in robot categorisation than a prototype representation (de Jong & Vogt 1998). Webb (2005) adapts Steels's (1997) framework for discrimination and language games, but uses a more psychologically plausible representation based on self-organising maps. And Cangelosi & Harnad (2000) use neural network representations for agents attempting to distinguish categories based on their perceptual features, with the help of language.

3.3.5   World Structure

Although I have discussed it in the subsections above, it is worth emphasising that most lexical emergence models have very simple representations of an external environment. In Steels (1996*b*), there is a small and finite number of objects, which are in fact already represented with discrete features. Smith's (2003*a*) world is also finite, but object features are represented with continuous values. Crucially, these finite worlds are typically pregenerated, random, and with uncorrelated dimensions, making it possible for the agents to fine-tune their category systems to fit the world's idiosyncrasies; indeed, over the course of many simulations, categories tend to converge on individual objects, rather than groups of similar ones (Smith 2003*a*). However, Webb (2005) has shown that whether the world structure is random or ordered can have significant consequences.

A few models have attempted to use more realistic environments. In his colour-centred study, Belpaeme (2002) used a set of colour stimuli, which, although finite, was so large as to be effectively infinite. And models that have been performed with real robots, such as the Talking Heads experiments (Steels et al. 2002), are implicitly based on a continuous infinite world, since the robots use real sensory channels to detect objects, which will appear differently from different angles and different lighting conditions (Vogt 2000). Unfortunately, in these models, the world is necessarily complex, and it is difficult to manipulate and study the effect of its structure.

CHAPTER 4

A New Model

4.1 Motivation

As we have seen, many different lexical emergence models have already been studied. In considering the large body of results, Steels et al. (2002) compiled a list of conditions that appear to be required for models to be successful, pertaining to agent interaction, conceptualisation, signal usage, population structure, and environment representation. Given some basic prerequisites, the models suggest that lexical emergence could have occurred through a dynamic process of self-organisation.

However, we must always remember that we are exploring how language actually did evolve, and not under what conditions it could hypothetically have evolved most efficiently. Although natural selection no doubt played a role in language evolution (Pinker & Bloom 1990), it does have limits (Lenormand 2002), so we should favour lexical emergence models that are also psychologically plausible over those that simply optimise communication. It is far from obvious that more psychologically plausible representations will lead to better results. In fact, Steels et al. (2002) have noted that lexical emergence occurs relatively slowly in prototype-based models, and that agents do not converge on a coherent lexicon if the meaning search space is too large. It is possible, then, that as we make our models more realistic, the results will deteriorate, in which case, previous conclusions about language evolution based on positive results would be placed in serious jeopardy. Consequently, in order to gain confidence in the validity

of the self-organisation hypothesis and ascertain that we are not dealing with false positives (Allchin 1999), increasingly realistic models must be developed to see if past results were not due to overly simplistic or idealised assumptions.

In this chapter, I develop such a model, based on a psychologically supported category representation and a less simplistic but easily modifiable world. If the results are comparable to those of past models, then the self-organisation hypothesis will gain further support and validity. Otherwise, the theory will need to be reevaluated.

The model I will present has two distinguishing features:

1. A prototype category representation, which can handle hierarchical relationships.
2. An infinite and probabilistically structured world, with correlated and continuous dimension values.

4.2   Overview

In this model, agent categorisation is based on prototype theory and resembles that of Belpaeme & Bleys (2005). The most important difference is that the dimension sensitivities of categories (i.e., the degree to which deviation of a feature from the prototype affects its category membership) are not fixed, but rather reflect the environment in which categories are acquired and the history of their development. This allows for some categories and category dimensions to be broader and others narrower, providing a prototype counterpart for hierarchical categories.

Furthermore, the world in this model is infinite, clumpy and correlated. Rather than a static world with a small number of predefined objects or meanings, contexts of objects are generated anew each time, to reflect the fact that no two real-world stimuli are identical. However, they are not generated completely randomly. Dimension values have continuous pdfs (probability distribution functions) in which some dimension values are more likely than others, and there are correlations between the dimensions, in accordance with the observations made by Mervis & Rosch (1981) (see Section 2.3.2).

In other respects, the model is largely based on previous work. There is a population of agents who interact with an environment and with each other. The agents have two

kinds of experiences which bring about changes to their conceptual systems and lexi-cons: discrimination games and guessing games. In a discrimination game, the concep-tual system of a single agent develops as it tries to discriminate an object in its environ-ment. In a guessing game, one agent conveys a word to the hearer for an object in the common environment, and the hearer tries to figure out which object the speaker was referring to. The primary objective is to explore under what conditions a population of agents can converge on a high communicative success rate and a coherent lexicon. The model has various parameters which affect its structure and behaviour, and differ-ent conditions can be explored by running simulations with different configurations of these parameter values.

4.3 The world

4.3.1 *Contexts and Objects*

Every experience that an agent has is situated in a different environmental context. Each context is a set of objects, and the number of objects per context is determined by the parameter $conxt_{size}$. A context can be interpreted as the environment that the agent is currently interacting with.

The objects themselves are defined as vectors of real numbers between 0 and 1. The number of dimensions and hence the length of these vectors is specified by the parameter dim_{num}. Each dimension value is meant to represent an abstract property which can be used to characterise objects according to a single quantifiable criterion. For example, if we have a world of 3 dimensions and a context size of 2 objects, then one context might be {[0.31, 0.72, 0.89], [0.19, 0.76, 0.02]}.

4.3.2 *Object generation*

Objects in this model are constructed dynamically: objects are generated on the fly each time a context is needed. However, rather than being random, the dimensions are "clumpy": certain dimension values are statistically more likely than others (note that this sense of clumpy is different than in Smith (2003*a*), from where I borrow this term). In addition, values along the different dimensions are partially correlated with each other. The shape of this structure and the degree of correlation depend on three

model parameters pertaining to world structure specifically. Two of these parameters determine the pdf of dimension values within a single dimension, and the third defines the degree of correlation between the dimensions.

The parameter dim_{parts} determines how many partitions (or subintervals) the pdf has within a dimension (note that each dimension has the same distribution of features). The pdf has a peak in the middle of each partition, and reaches down to zero at its endpoints. For example, if the value of the parameter is 1, then there will be a strong bias towards values close to 0.5, with values becoming increasingly unlikely as they approach 0 or 1. If the value is 10, then the interval [0,1] will be divided into ten partitions (i.e., [0.0,0.1], [0.1,0.2], ..., [0.9,1.0]), with the probability distribution peaking at the ten midpoints.

dim_{clump} determines how biased the pdf is to the partition midpoints. If the value is close to 0, then the bias will be extremely small, and the distribution will be nearly uniform, with the midpoints of the partitions being only mildly favoured. If the value is large, then the bias will be large, and the dimension values will mostly be clustered within very short neighbourhoods of the partition midpoints.

dim_{correl} determines the degree of correlation between the dimensions, i.e., the extent to which different dimension values of the same object will tend to be similar to each other. If this parameter has a very small value (close to 0), then the dimension values are nearly independent of each other. On the other hand, if the value is large, then the dimension values are highly coupled.

Individual objects are generated independently, using a set of random numbers and a series of mathematical operations. In brief, first an initial random number is created on the interval [0,1], then a number is chosen for each dimension which is close to the initial number (due to dimension correlation), and finally these numbers are transformed into the actual object generation values (depending on world clumpiness). The basic object generation algorithm is shown in Table 4.1.

The definition of $\zeta(x)$ makes sure that the individual o_i values are all probabilistically more likely to be around the centres of the dimension partitions, although how likely this is depends on the value of dim_{clump}. $\phi_r(x)$ assures that the o_i values of an object will tend to be similar, to an extent determined by the value of dim_{correl}. Notice that $\zeta(x)$

Precondition A probability distribution function $\zeta(x)$ has been defined for dimension clumpiness (determined by the parameters dim_{clump} and dim_{parts}).

Procedure
1. A random number, r, is generated on the interval $[0, 1]$.
2. A probability distribution function, $\phi_r(x)$, is defined for dimension correlation, determined by the values of r and the parameter dim_{correl}.
3. A random number, r_i, is generated on the interval $[0, 1]$ for each dimension.
4. Each r_i is transformed into a value closer to r, c_i, in accordance with $\phi_r(x)$.
5. A dimension value is finally obtained for each object dimension by transforming each c_i, in accordance with $\zeta(x)$.

Table 4.1: Object generation.

is thus fixed for the entire simulation, while $phi_r(x)$ is defined anew each time a new object is being created. The precise manner in which these functions are defined and used is a bit complicated and left deliberately vague here: for full mathematical details, see Appendix B.

Figure 4.1 shows the distribution functions $\zeta(x)$ for one particular world, along with the generation of four objects. For each object, a random r is found, $\phi_r(x)$ is plotted, and a chart is shown which shows the steps in object generation, culminating in the object dimension values. Notice that all of the dimension values for all four objects are fairly close to the partition centre values. In addition, in most (but not all) cases, the dimension values for different objects are very close together, reflecting the relatively high level of dimension correlation.

4.4 The agents

4.4.1 Perceptual Spaces

Agents are equipped with a (possibly empty) private perceptual space and the ability to create and adapt categories within it. The dimensional structure of an agent's perceptual space is identical to that of the world. In other words, its number of dimensions is determined by the parameter dim_{num}, and each dimension spans the interval $[0, 1]$. Moreover, an agent maps the dimensional values of an object directly onto the same values in its perceptual space. This implies that there is no subjectivity due to perspective

$r=0.757$	dim1	dim2	dim3
r_i	0.149	0.966	0.569
c_i	0.653	0.797	0.738
o_i	**0.694**	**0.727**	**0.704**

$r=0.293$	dim1	dim2	dim3
r_i	0.995	0.810	0.698
c_i	0.549	0.373	0.346
o_i	**0.506**	**0.311**	**0.305**

$r=0.587$	dim1	dim2	dim3
r_i	0.934	0.897	0.202
c_i	0.650	0.636	0.533
o_i	**0.694**	**0.691**	**0.504**

$r=0.301$	dim1	dim2	dim3
r_i	0.154	0.617	0.962
c_i	0.283	0.338	0.464
o_i	**0.298**	**0.304**	**0.496**
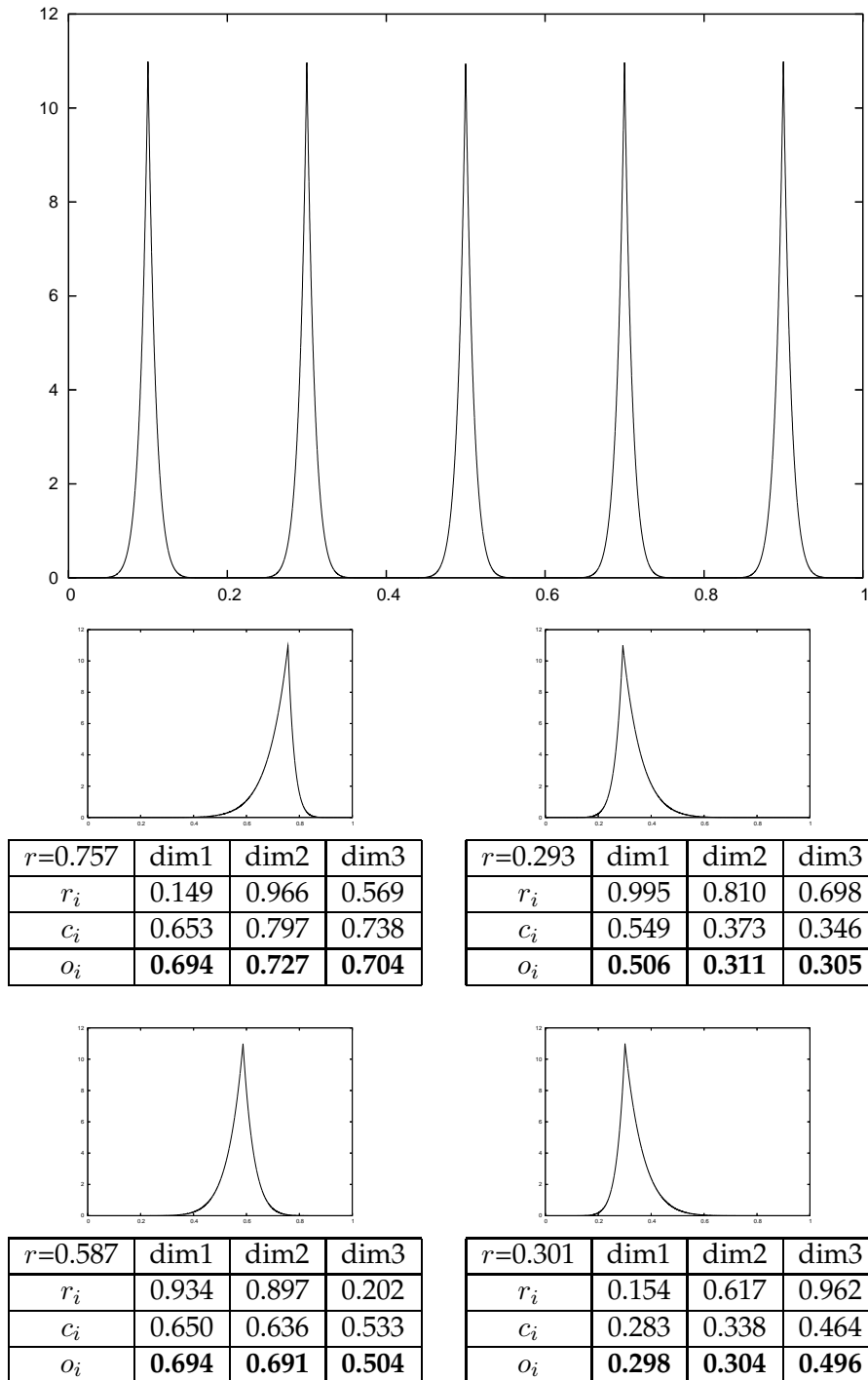
Figure 4.1: Generation of 4 objects: The top graph shows the distribution of a world dimension. Underneath, the generation of 4 objects is shown, following the algorithm described in Table 4.1. For each object is shown a plot of the function $\phi_r(x)$ and a table which shows the steps in the derivation of the object dimension values, o_i. The world structure parameters for these examples were set to $dim_{clump} = 10$, $dim_{correl} = 10$, $dim_{num} = 3$, and $dim_{parts} = 5$.

or context in perceiving an object's features. This idealisation makes the task simpler for

the agents, and has a non-trivial impact, as can be seen from experiments with robots,

in which variability in lighting and perspective do add a good deal of complexity to the problem of categorisation (Vogt 2000). Nevertheless, it is necessary to make some idealisations in order to restrict the range of phenomena and variables that the model can hope to explore.

As for the agents' categories that are defined in these perceptual spaces, it is tempting to interpret them as concepts, but one must be cautious. In particular, as mentioned in Section 2.2, features recursively rely on other concepts and features. Indeed, it is not even clear if the agents' concepts are more or less basic than the dimensions used to categorisee them. As a result, it seems more sensible to think of the dimensions as lower level perceptual or sensory information. For example, one dimension could refer to sight, another to sound, etc. In that case, the dimensions can really be argued to be sensory channels (in the sense of Smith (2003*a*)). However, these kinds of coarse and unprocessed dimensions seem insufficient for developing actual concepts. Therefore, I will treat the dimensions as relatively low level perceptual features (which is why I refer to a perceptual rather than a conceptual space), but will generally avoid the use of the term "concept" in favour of the more vague term "category", and think of the dimensions and categories as the perceptual foundations of concepts. I will still sometimes refer to more complex features in examples, as they tend to be easier to understand.

4.4.2 Category Representation

In order to understand how categories are represented and defined over an agent's perceptual space, it is useful to first consider how categories work in classically oriented models. Recall that classical categories do not make distinctions between central and peripheral members, exhibit no graded membership, and have sharp boundaries. In effect, when an agent perceives an object from the environment and maps it onto a point in its perceptual space, it can make black-or-white category membership judgements for the objects with respect to each of its categories. We can think of each such category, then, as a binary function over the entire perceptual space, which maps each point to either 0 (not a member) or 1 (a member). This does not mean, however, that every object is only a member of one category: in hierarchal discrimination trees such as Smith's (2003*a*), for example, an object would be classified as a member of a particular leaf node and all of that node's ancestors.

In the current model, the binary nature of category judgements is replaced with a continuous scale from 0 to 1, in which the higher the value on the scale, the better the category membership. To consider a concrete example, with respect to a bird category, a sparrow might get a very high value (e.g., 0.99), a heron somewhat lower (e.g., 0.8), a penguin lower still (e.g., 0.6), a squirrel quite low (e.g., 0.3), and a rock nearly as low as possible (e.g., 0.01). Note that in such a system, every object is a member of every category to some extent, and it is the degree of membership that is crucial. At first, this may seem counter-intuitive and implausible: after all, how can one claim that a categorisation system in which squirrels and even rocks are judged to belong, however marginally, to a bird category, is more realistic than one that makes clear and "correct" judgements? However, recall that we are dealing here primarily with perceptual features and trying to match objects to categories. In terms of perceptual characteristics, squirrels certainly do share certain things in common with birds (such as having two eyes or engaging in biological motion), and indeed, no one would deny that squirrels are perceptually more bird-like than rocks.

Moreover, there is little threat of agents classifying rocks as birds for two reasons. First, as we will see in Section 4.4.3, agents will only use (or create) one category when categorising an object, and will try to choose the "best" fit, so a category will never be used for an object when there is a better alternative. Second, and more decisively, in order to make judgements of when to make new categories and when to use existing ones, it has been necessary to establish a threshold parameter, mem_{min}, which specifies the minimum degree of category membership required for an object to actually be matched to a category. Thus, the concepts in this model do not have entirely fuzzy boundaries. However, individual dimensions do have fuzzy edges, as the degree of acceptability in a given dimension depends on the object's other dimension values, and category boundaries may shift over time, due to category adjustments (see Section 4.5.4). This kind of fuzziness is thus in line with Croft & Cruse's (2004) view on category boundaries as being sharp but dynamic.

Let us proceed with technical specifics. A category's representation consists of two related parts, a prototype and a set of dimension sensitivities. The prototype refers to a point in the perceptual space, and specifies the dimension values that identify optimal members of the category. An object with the same values as the prototype would be

an example of the category's prototype, and, as we will see, will exhibit the maximum possible degree of category membership (i.e., 1).

Every category also has a set of dimension sensitivities, one for each dimension. A dimension sensitivity determines the extent to which deviations between an object's dimension value and the corresponding prototype dimension value affect the object's category membership. To give a real-world example, if we were judging whether something is a ball, we would consider shape and perhaps texture as very important, but not size. On the other hand, colour would be more important when classifying snow, while shape less so. The snow category would thus have a significantly higher dimension sensitivity in the colour dimension than the ball category.

4.4.3 Calculating Category Membership

A category can thus be thought of as a list of pairs of numbers which determine category membership over every point in the agent's perceptual space. For the purposes of our model, we require a precise membership measure with three properties:

- One point (the prototype) should yield the value 1.
- All other points should decrease around the prototype, symmetrically in each dimension.
- The rate at which the function drops in each dimension depends on the corresponding dimension sensitivity.

A good function with these properties in a single dimension, with support in the psychology literature (Shepard 1987), is the gaussian function:

$$G(x) = e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4.1}$$

A plot of the gaussian function is given in Figure 4.2, with $\mu = 0.4$ and $\sigma = 0.2$. The peak is at μ, and the speed at which it decreases is faster if σ is smaller. For our purposes, we can interpret μ as the prototype value and σ as the dimension sensitivity or width (keeping in mind that a high sensitivity corresponds to a small value of σ).
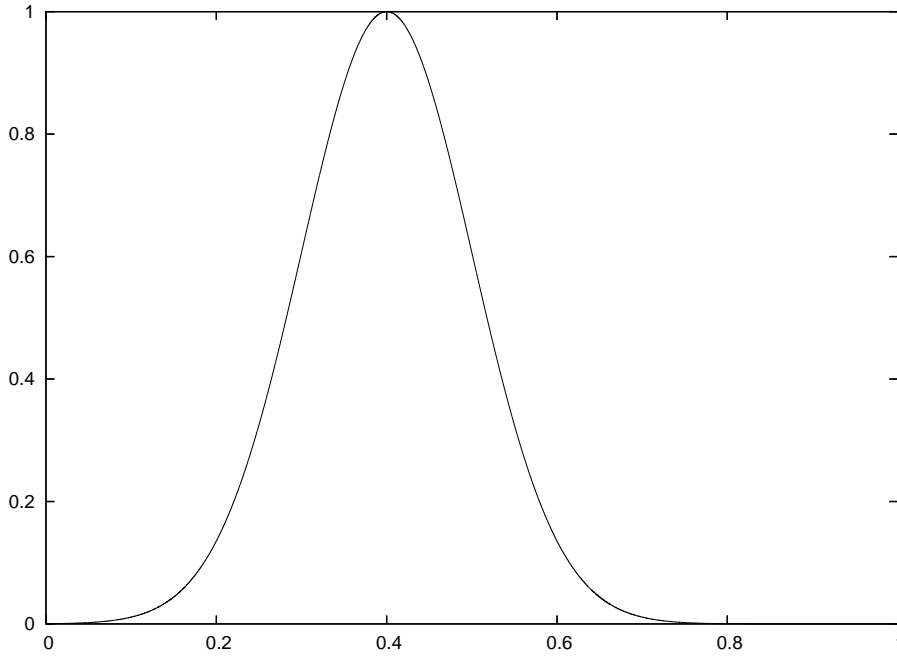
Figure 4.2: $G(x)$: A gaussian function, with $\mu = 0.4$ and $\sigma = 0.1$.

For the general case of multiple dimensions, one can use a multivariate gaussian function with diagonal covariances. However, in order to normalise for the number of dimensions (so that higher dimensional cases do not intrinsically yield lower category memberships), the Nth root of the gaussian function's value is taken, where N is dim_{num}. Therefore, the category membership function used to determine the degree of membership of an object o in a category c is

$$membership_c(o) = \left(\prod_{i=1}^{N} e^{-\frac{1}{2}\left(\frac{o_i - p_i}{s_i}\right)^2} \right)^{1/N} , \qquad (4.2)$$

where i identifies a dimension, with o_i being the object value, p_i the prototype value, and s_i the sensitivity. Notice that this function also has a maximum of 1 (when all object dimension values match the prototype) and never descends all the way to 0. An example of $membership_c(o)$ for an agent's category in a perceptual space of two dimensions is shown in Figure 4.3. Note the greater sensitivity of the function (i.e., the faster decline) along the x-axis than along the y-axis. Also, the object's value lies above the plane defined by mem_{min}, so this category could potentially be used to categorize the object.

Note also the greater sensitivity of the function (i.e., the faster decline) along the x-axis than along the y-axis.
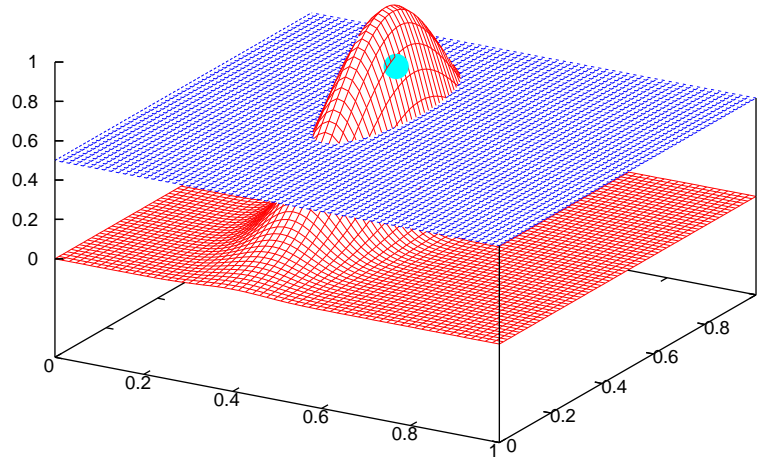
Figure 4.3: Category membership in 2 dimensions: $membership_c(o)$, the category membership function for an agent's category c in a conceptual space of two dimensions, with $p_0 = 0.4$, $s_0 = 0.05$, $p_1 = 0.6$, and $s_1 = 0.1$. The blue plane shows the value of the mem_{min}, and the light blue dot indicates the object being categorised: since the dot is above the plane, the object can be considered to be a member of the category.

It is more difficult to visualise category membership with more than two dimensions, but Figure 4.4 shows an example of a category in five dimensions, with the gaussian function for each dimension plotted independently on the same graph.

An example of a category member calculation with the category in Figure 4.4 applied to the object [0.12, 0.2, 0.51, 0.65, 0.9] is shown in Figure 4.5. Each object dimension value is shown with a line from the x-axis to where it meets the corresponding gaussian curve for that dimension (in the same colour). Notice that along one dimension (the grey one), the object matches the prototype perfectly. Also, the graph shows how the dimension sensitivities play an important role: the example object's fit for the category is the worst in the category's most sensitive dimension (the blue one), despite its value being closest to the prototype value in that dimension. In contrast, the object's dimension value is relatively far from the prototype in the category's least sensitive dimension (the green one), but this is of little consequence, since the curve decreases so slowly. The gaussian
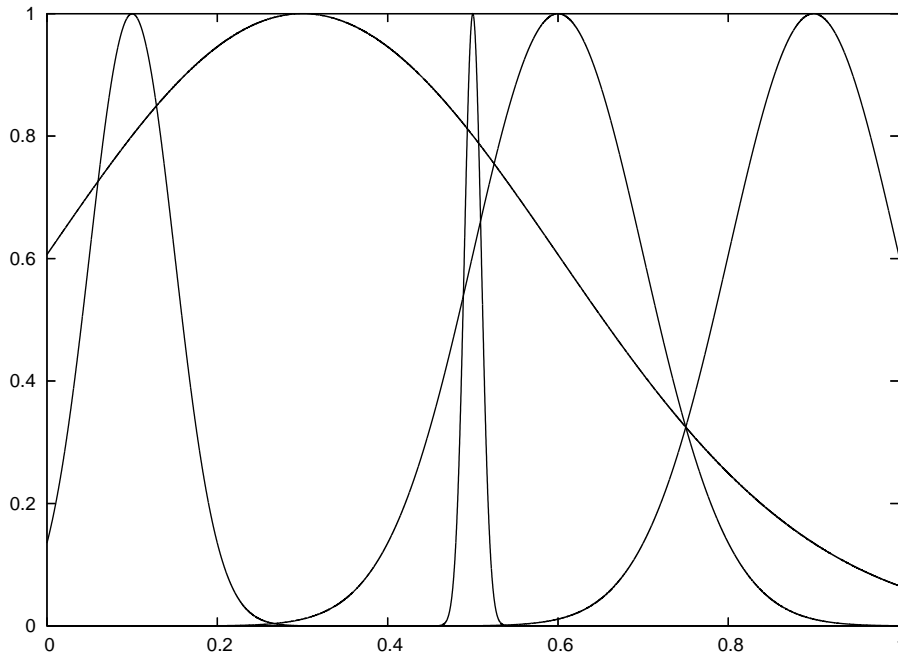
Figure 4.4: A 5-dimensional category: The category structure of a single category in a 5-dimensional perceptual space, with (left peak to right peak) $p_0 = 0.1$, $s_0 = 0.05$, $p_1 = 0.3$, $s_1 = 0.3$, $p_2 = 0.5$, $s_2 = 0.01$, $p_3 = 0.6$, $s_3 = 0.1$, $p_4 = 0.9$, $s_4 = 0.1$. The one-dimensional gaussian functions corresponding to each dimension are plotted independently.

values in the five dimensions (from left to right on the graph) are 0.9231, 0.9460, 0.6065, 0.8825, and 1.0. Multiplying these together and taking the fifth root yields a category membership of 0.8621. If this is above the value of mem_{min} and if it wins out over other categories, then this category is used to categorize the object.

4.4.4 Words

Each category is associated with a list of words, and each such association has a score attached to it. (This list can also be empty, in which case the category has not yet been lexicalised.) Thus, in this model, words are considered to be dependent on prerequisite categories: an agent's "lexicon" is embedded in its category system. If a category would be forgotten, for example (although that does not happen in this model), then the word would be lost as well.

Unlike categories, the representations used for words is extremely simple. They are strings of three characters with a consonant-vowel-consonant (CVC) structure (for a total of $21 \times 5 \times 21 = 2205$ possible distinct words), which are treated as discrete and symbolic.
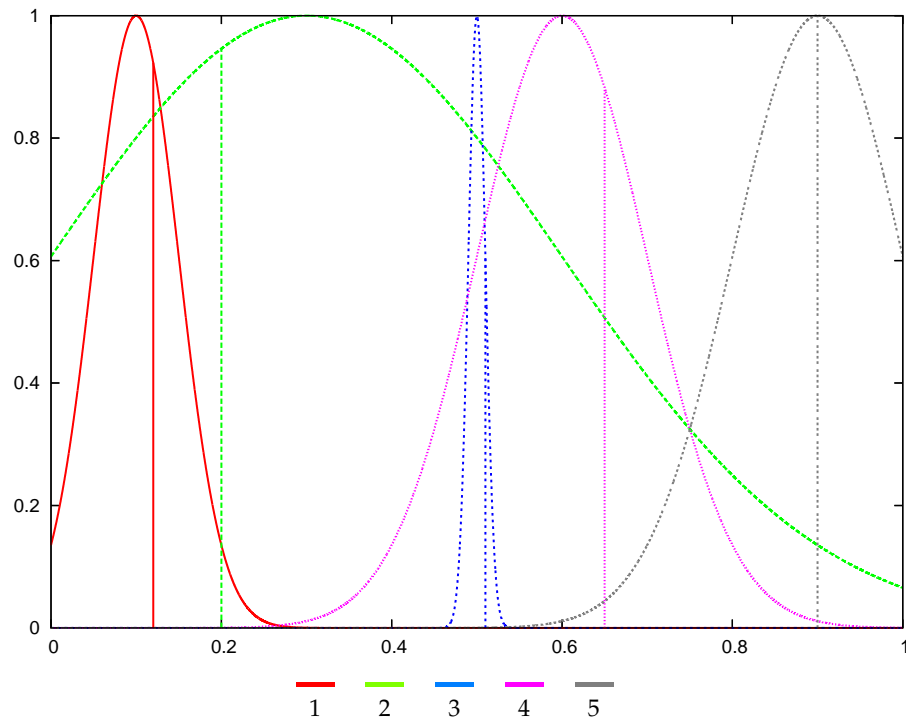
Figure 4.5: Category membership in 5 dimensions: The same category as in Figure 4.4 is plotted here, along with colour-matching dimension values for an object being categorised. The example object here has dimension values [0.12, 0.2, 0.51, 0.65, 0.9]. Applying the gaussians, and then taking the 5th root of their product gives a category membership of 0.8621.

In addition, agents always both produce and perceive words perfectly. Consequently, words like "kes" and "kez" are no closer together in any sense than "kes" and "wum".

Notice that this system, although simplistic, has potential counterparts for human language relationships, including synonymy, homonymy and polysemy. If two or more words are associated with one category, they are synonyms. If the same word is associated with two or more related or overlapping categories, they are arguably polysemous. And if two very different categories have the same word attached to them, then they are homonyms. We will see that all three of these cases emerge spontaneously in the simulations.

4.5   Discrimination Games

4.5.1   Overview

The discrimination games used in this model are based on the design developed by Steels (1996b). The fundamental objective and setup has been preserved, but some changes

have been made to the implementation, most of which are due to the prototype representation used here.

A discrimination game is a strictly private experience, involving a single agent and a world context. The agent's task is to select an object from the context at random (which I will call the topic), and uniquely categorize it, i.e., to find (or create, if necessary) a category which fits the topic but fails to fit the other context objects. Depending on the agent's current set of categories, the topic and the context, the agent will perform one of the three following operations: create an entirely new category, split off a new category from an existing one, or adjust an existing category.

Before explaining the procedure that establishes which operation is performed, I introduce two terms:

candidate category A category c is a candidate category for an object o if o's degree of membership in c is sufficient for it to be classified in that category. Recall that this condition amounts to $membership_c(o) >= mem_{min}$.

distinguishing category A category c is a distinguishing category for an object o if c is a candidate category of o and c is not a candidate category for any other object in the context (Smith 2003*a*).

The following procedure is used to determine which of the three operations above is used by the agent during a discrimination game. First, the agent searches its conceptual system and builds a list of all of the topic's candidate categories. If this list is empty, it means that no category is good enough to classify the object, and the agent must create an entirely new category. Otherwise, the agent searches the topic's candidate categories for distinguishing categories. If successful, it chooses the one with the highest membership, and this category is then adjusted to better fit the topic relative to the other context objects. If there are no distinguishing categories, then the agent finds the candidate category which has the most sensitive dimensions, by calculating the product of the dimension sensitivities for each category and choosing the one for which this product is the smallest. An even finer category is then created by creating a subcategory of this one. A formalised description of this algorithm is shown in Table 4.2.

Precondition An agent with a set of categories C is presented with a context O.

Procedure

1. Choose a topic object o_t randomly from O.
2. Find the set of candidate categories C_c for o_t (which is a subset of C).
3. If C_c is empty, add a new category, and terminate. Otherwise, move on to step 4.
4. Find the set of discriminating categories C_d for o_t (which is a subset of C_c).
5. If C_d is empty, find the category c_s in C_c which has the most sensitive dimensions, split off a new subcategory from c_s, and terminate. Otherwise, move on to step 6.
6. Find the category c_a in C_d which maximises $membership_{c_a}(o_t)$, adjust c_a, and terminate.

Table 4.2: Discrimination algorithm.

4.5.2 Creating a new category

When creating an entirely new category, both the topic and context play an important role, while the set of already established categories is irrelevant. The prototype's dimension values are taken directly from those of the topic, so that the topic would be given a category membership of 1 in the category, regardless of the dimension sensitivities or the other preexisting categories.

The calculation of the dimension sensitivities, on the other hand, is a little more complicated, and involves both the topic and the rest of the context. The idea is to use sensitivities which will guarantee that the other context objects will not qualify for membership in the new category, and yet not make the sensitivities too small, as this might make the category too restrictive to be of much use in the future. The strategy employed here to achieve this is to choose each dimension sensitivity s_i independently in such a way that if we were dealing with a one-dimensional world consisting of only the dimension i, then even the closest other object in the context would not pass the category membership threshold (falling short by a ratio determined by the parameter $adjust_{rat}$). This is done by solving the 1-dimensional category membership function (which is just the 1-dimensional Gaussian function given in Equation 4.1) for s_i, with p_i set to the topic value o_{t_i} (as this will become the prototype value), and the membership set to mem_{conxt}.

This parameter thus determines what the sensitivity of a dimension will be set to given the dimension values of the other context objects.

$$s_i = \frac{\min_{o \in O, o \neq o_t} |o_i - p_i|}{\sqrt{-2log(mem_{conxt})}} \tag{4.3}$$

We then independently obtain the s_i for each dimension in this way, and this will guarantee that every non-topic object in the context will have a membership of at most mem_{conxt}(a mathematical proof of this claim is shown in Appendix C).

4.5.3   Splitting off a new subcategory

Recall that subcategories are split off when a discrimination game finds that the agent has candidate categories for the topic, but no distinguishing categories. In this case, the agent looks for its most sensitive category, and splits off a new subcategory from it.

The most sensitive category c_s is defined to be the one for which the product of its dimension sensitivities is minimised, i.e.,

$$c_s = \underset{c \in C_c}{\arg\max} \prod_{i}^{N} s_i. \tag{4.4}$$

On the basis of this category, a new category is created. First, the dimension is sought in which the topic is most unlike each of the other context objects. In particular, in each dimension, we find the minimum difference between the topic value and the other dimension values, and then choose the dimension along which this difference is the largest:

$$i = \underset{i \in \{1,...,N\}}{\arg\max} \left(\min |o_{t_i} - o_i| \right) \tag{4.5}$$

i is the dimension along which the split is performed. The new category is defined as a subcategory of c_s by inheriting the prototype values and dimension sensitivities in all the dimensions except i. As for dimension i, its prototype and dimension values are calculated just as those of an entirely new category (see the preceding section).

Notice that this calculation is guaranteed to make the new category a better fit for the topic than the superordinate category it was derived from. On the other hand, although very unlikely, it is not strictly guaranteed to make the new category a discriminating category for the topic, or even that the new category will have more sensitive dimensions than the category it was derived from. For computational reasons, further calculations which would turn this unlikelihood into impossibility have been avoided.

4.5.4   Adjusting an existing category

When at least one discriminating category is found for the topic object, then no new category needs to be added, and the agent only makes minor adjustments to the category, in order to make it fit the topic better relative to the context. I have already explained how discriminating categories are found: here I explain how these adjustments are made.

Having chosen the discriminating category for which the topic object's membership is maximised, the agent adjusts its prototype value and dimension sensitivity in each dimension independently.

A prototype's dimension value is shifted towards the topic object value. But by how much? Certainly, the amount of shift should depend on the distance between the topic and prototype dimension values. If the topic and prototype are equal, then the shift should be 0, and the larger the distance between them, the greater the shift should be, without of course ever actually being larger than the distance. However, since in this model the dimension sensitivities are not fixed (as they are in Belpaeme (2002), for example), and the importance of the distance between prototype and object values with regards to determining category membership depends on them, they should have an effect on the amount of shift as well. For example, if prototype and topic are 0.05 apart in a particular dimension, then this will have little impact (and thus little need for adjustment) if the dimension sensitivity has a high value like 0.3, but will be very significant if the sensitivity is only, say, 0.01. As a result, the particular function that has been chosen to calculate the amount of shift is a function of the distance between the prototype and topic values, and the ratio of this distance over the dimension sensitivity:

$$shift_i(d_i, r_i) = d_i - \frac{1}{r_i + \frac{1}{d_i}}, \text{where } d_i = |p_i - t_i| \text{ and } r_i = \frac{d_i}{s_i} \tag{4.6}$$

Note that when $d_i = 0$, $shift_i = 0$, and as r_i gets larger and larger, $shift_i$ asymptotically approaches d_i.

Once the amount of shift required is calculated, the prototype's dimension value is updated accordingly. Table 4.3 shows the resultant values for a fixed prototype starting value of 0.6, and topic values and dimension sensitivities covering a range of values. Note how the changes are larger for smaller values of s_i (i.e., larger dimension sensitivities), and that even in such cases, the shift is usually quite small. Also, not all combinations of the variables are relevant, because if r_i is large, then the gaussian function would yield a tiny value, in which case the topic would never have been identified as a member of the category. Accordingly, values corresponding to large ratios (i.e., $r_i > 4$) are not shown in the table.

$s_i \backslash t_i$	0.400	0.425	0.450	0.475	0.500	0.525	0.550	0.575	0.600
0.025	0.471	0.486	0.495	0.499	0.500	0.501	0.505	0.514	0.529
0.050	0.483	0.492	0.498	0.500	0.500	0.500	0.502	0.508	0.517
0.075	0.488	0.495	0.498	0.500	0.500	0.500	0.502	0.505	0.512
0.100	0.491	0.496	0.499	0.500	0.500	0.500	0.501	0.504	0.509
0.125	0.493	0.497	0.499	0.500	0.500	0.500	0.501	0.503	0.507
0.150	0.494	0.497	0.499	0.500	0.500	0.500	0.501	0.503	0.506
0.175	0.495	0.498	0.499	0.500	0.500	0.500	0.501	0.502	0.505
0.200	0.495	0.498	0.499	0.500	0.500	0.500	0.501	0.502	0.505

Table 4.3: Prototype shifting: Example of what values a category prototype in dimension i would be shifted to from a value of 0.5 for various possible combinations of topic dimensions t_i and dimension sensitivities s_i (rounded to 3 decimal points).

Having adjusted the prototype centre for a dimension, the agent then adjusts the corresponding dimension sensitivity. The guiding idea here is that, just as in category creation, any adjustment of the category sensitivities should aim towards making the topic more distinguishable from the other context objects. In contrast with the category creation and splitting cases, however, the topic is not used as a model to create a new prototype, and so the topic and prototype values will differ as well. In addition, the adjustments should not be too large, since otherwise the category's structure changes drastically with each adjustment, leading to an unstable conceptual system. Consequently,

adjustment occurs using the new dimension sensitivity calculation (Equation (4.3)) based on the non-topic context objects, and a parameter which affects how much this calculation should affect the adjustment.

In particular, a target dimension sensitivity, $s_{i_{target}}$, is calculated based on the non-topic context objects in exactly the same manner as in the cases of creating a new category or subcategory (i.e., using Equation (4.3)). This time, however, the result does not determine the dimension sensitivity independently, but only plays a role in a larger calculation. The extent by which the actual category's sensitivity is changed towards this target is determined by the parameter $adjust_{rat}$. If $adjust_{rat} = 0.0$, then the sensitivity is not changed at all, making the above calculations in fact unnecessary. If $adjust_{rat} = 1.0$, then the sensitivity is set to $s_{i_{target}}$. For intermediate values of $adjust_{rat}$, the new sensitivity is an interpolation between the old one and the target. In short, the new sensitivity is calculated and updated by:

$$s_{i_{new}} = s_{i_{old}} + adjust_{ratio} \times \left(s_{i_{target}} - s_{i_{old}}\right) \qquad (4.7)$$

4.5.5 Examples

In this section, I will demonstrate the mechanisms of the discrimination game by looking at individual games in a single run of discrimination games. In order to facilitate visualisation and avoid clutter, I have used a run with only a one-dimension and a context size of two (i.e., $dim_{num} = 1$ and $conxt_{size} = 2$). The other relevant parameters used in this example were: $mem_{min} = 0.8$, $mem_{conxt} = 0.1$, $adjust_{rat} = 0.5$, $dim_{clump} = 10.0$ and $dim_{parts} = 5$.

Each figure below consists of two graphs, showing one discrimination game. The x-axis shows dimension values, and the y-axis indicates category membership. The agent's category representations are shown with blue gaussian curves, while the context object dimension values are indicated with vertical green lines, with the topic specified further with a vertical text label. The left graph shows the agent's set of categories before the discrimination game, while the right graph (b) shows the same set after modification via the discrimination game. As a result, each (b) graph contains either one more curve than the corresponding (a) graph (if a new category was created or split off), or the

same number of curves but with one curve altered (if an existing category was adjusted). Finally, the graphs also show the parameters mem_{min} with a horizontal red line, and mem_{conxt} with a horizontal light blue line.

Figure 4.6 shows what happens in an agent's very first discrimination game, provided it has not been equipped with any innate categories (as was the case in all the simulations investigated here). In (a), the agent does not yet have any categories, so it must create one. The new category shown in (b) has a peak at the topic's dimension value. Since the dimension sensitivity has been calculated according to Equation (4.4)), the nearest object to the topic (in a context size of 2 like here, this is just trivially the non-topic object) has a membership of mem_{conxt}.
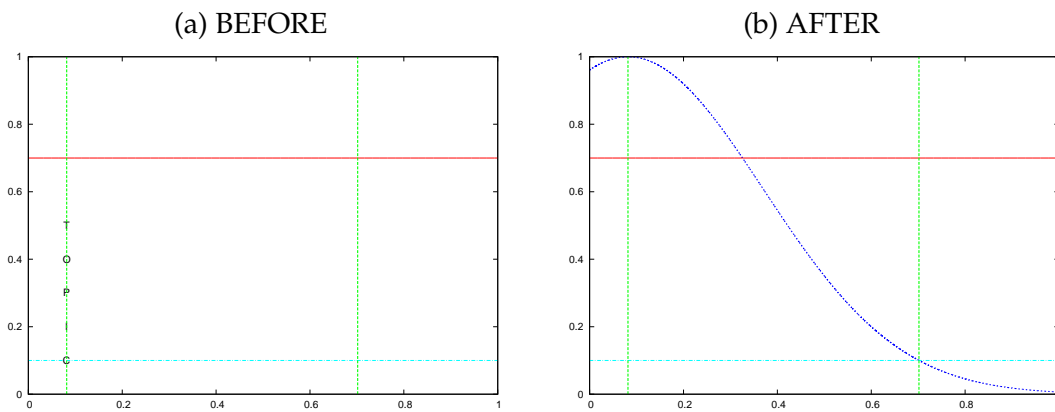
Figure 4.6: Discrimination game example: Adding the first category.

The immediately following discrimination game is shown in Figure 4.7. The agent starts with the single category system that resulted from the first discrimination game. The topic's category membership is clearly inadequate in the category: in the graph, this can be clearly seen by noting that the topic's green line intersects the blue category curve much lower than the red mem_{min} line. Consequently, just as in the first discrimination game, it creates a new category. Note that in this case, the new category's dimension sensitivity is much stronger, because the topic and other context object's value varied by a lesser extent. The agent's category sensitivity thus reflects the degree of discrimination that was required when it was created.

In the next game (Figure 4.8), the topic is coincidentally quite similar to what it was in the previous game, around 0.5 (although not entirely coincidentally, since the world struc-ture parameters make 0.5 a relatively likely object dimension value: see Section 4.3.2).
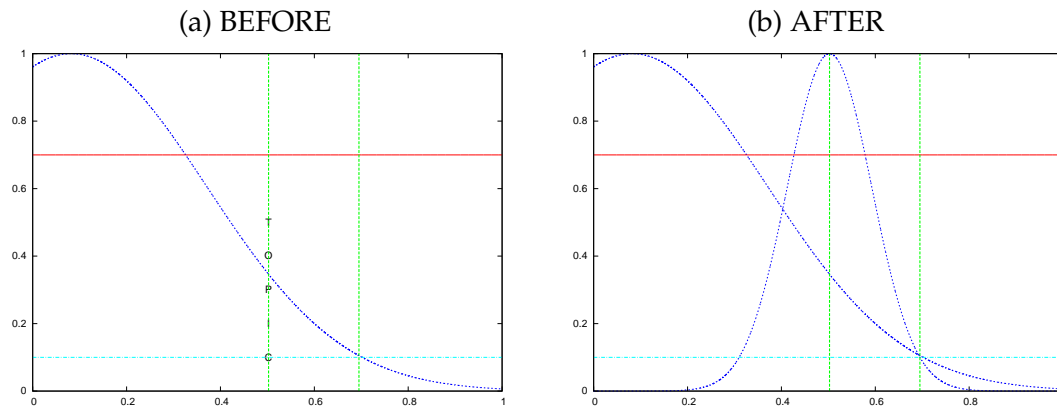
Figure 4.7: Discrimination game example: Adding another category.

The category then created is thus an excellent fit for the current topic, while matching the other object very poorly. As a result, in this case, the category adjustment mechanism is followed. First, the category centre is shifted slightly in the direction of the topic, as defined in Equation (4.6). The shift is virtually indiscernible, however, since the topic and prototype are extremely close, and the dimension sensitivity is not too strong. Next, the dimension sensitivity is modified. A potential dimension sensitivity (according to Equation 4.4) is calculated which would give the non-topic object a membership of mem_{conxt}, and the category's sensitivity is updated according to Equation (4.7). Note that the sensitivity does not get changed so much that the non-topic object's membership actually reaches mem_{conxt}. This would have occurred had $adjust_{rat}$ been set to 1, but as it is only 0.5, the shift only occurs part of the way there. The net result of this discrimination game is thus that we still have two categories, but that one has become wider, reflecting the fact that it has now been used in a situation which required less discrimination than before.

Figure 4.8: Discrimination game example: Widening a category.

Figure 4.9 shows an example of what happens when there is a context object whose dimension value is very close to that of the topic. A very strong degree of dimension sensitivity is required in this case, so the resultan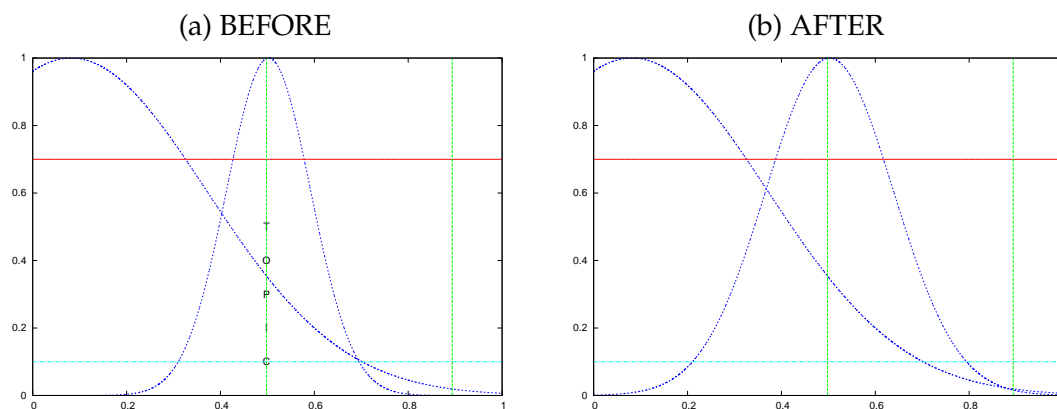t category is extremely thin. Note, however, that if the agent ever uses this category again, it is likely to encounter it in a more merciful context, in which case the category will be broadened significantly in a category adjustment like the one above.

(a) BEFORE   (b) AFTER
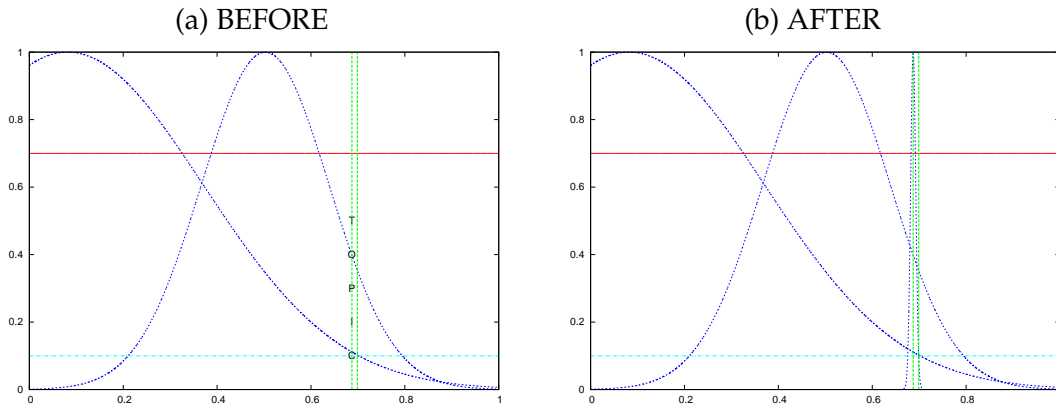
Figure 4.9: Discrimination game example: Adding a narrow category.

The next figure (4.10) represents a discrimination several games later, and is an example of a category splitting off. As in the previous case, the other context object and topic dimension values are very close. In this case, they both have sufficient membership in the rightmost category, making that a candidate but not discriminating category for the topic. A subcategory is therefore created. However, for a 1-dimensional case like the one exemplified here, it is worth noting that splitting off a category and adding a new category actually do exactly the same thing. Recall from Section 4.5.3 that when a category is split off, all prototype values and dimension sensitivities except one are copied to it from its parent, while the last is treated as for new categories. But of course, when there is only one dimension, it means that nothing is copied from the parent. Nevertheless, we do end up with a category which can be called a subcategory, in that it is much like a more refined version of its parent.

Finally, Figure 4.11 shows the 500th discrimination game in this run. A large number of categories have been built, with varying degrees of sensitivity. Note, however, that the prototypes cluster around the values 0.1, 0.3, 0.5, 0.7 and 0.9, which are precisely the midpoints of the world partitions.

(a) BEFORE (b) AFTER
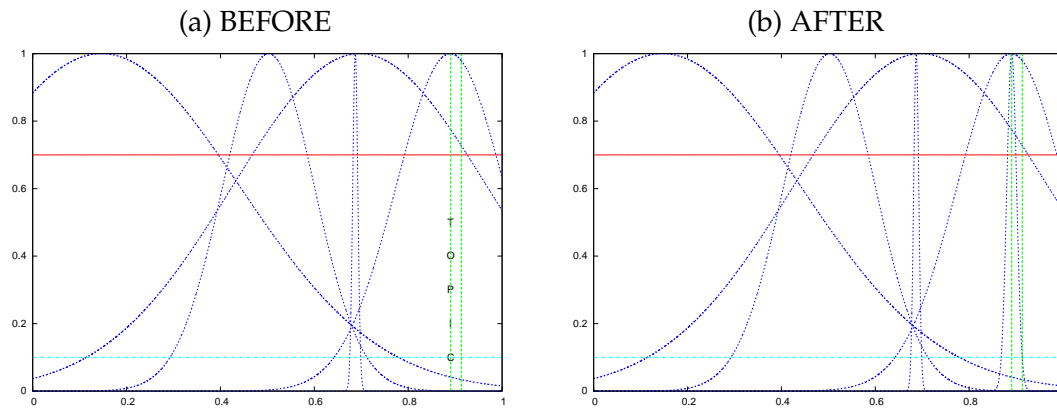
Figure 4.10: Discrimination game example: Splitting a category.

(a) BEFORE (b) AFTER
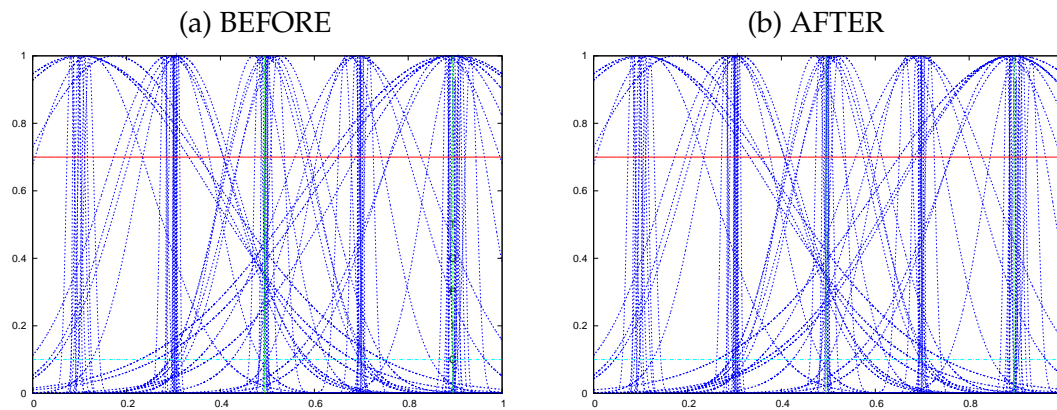
Figure 4.11: Discrimination game example: After 500 discrimination games.

When there are more than three dimensions, things become more complicated and more difficult to visualise. Due to lack of space, I will not discuss examples in multiple dimensions here, but the reader is invited to study Figures 4.12 and 4.13, which show the 1st through 9th and the 500th discrimination games for one run, with the same parameter values as in the previous example, except that $dim_{num} = 3$ and $dim_{clump} = 10$. For compactness, I show only the result of each discrimination game.
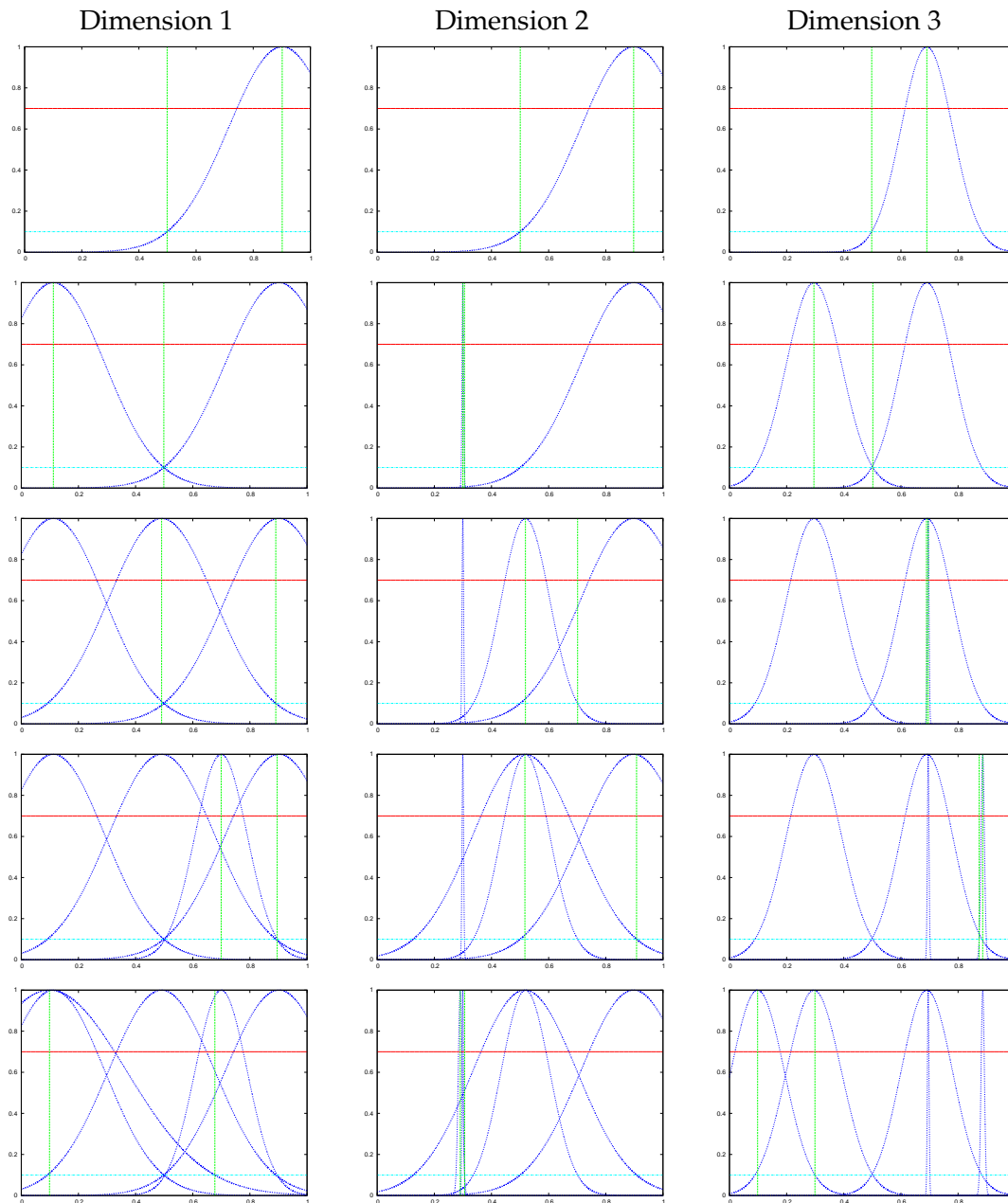
Figure 4.12: Discrimination game examples with 3 dimensions: 1st, 2nd, 3rd, 4th, and 5th discrimination games.
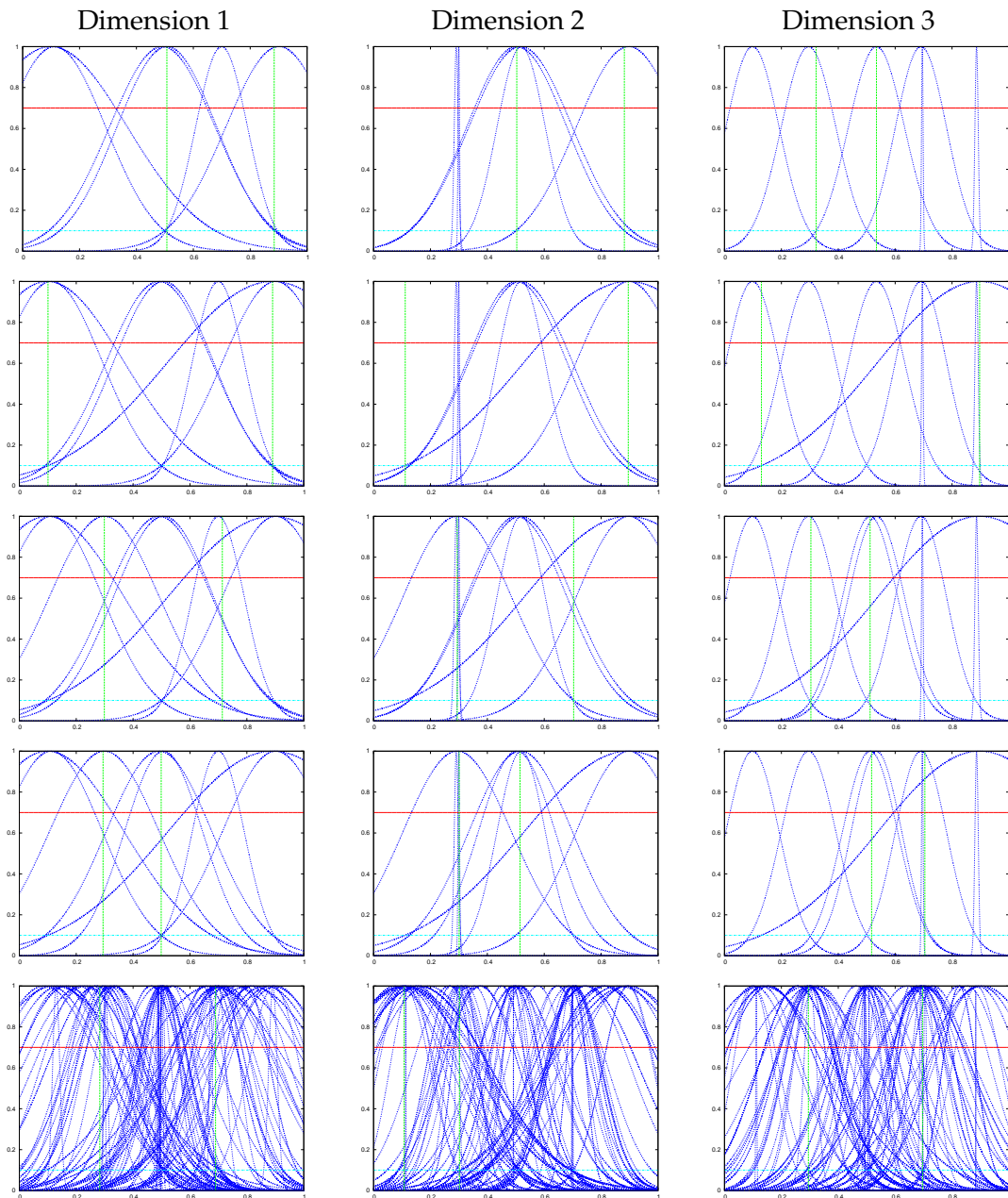
Figure 4.13: Discrimination game examples with 3 dimensions: 6th, 7th, 8th, 9th, and 500th discrimination games.

4.6   Guessing Games

In addition to discrimination games, agents also engage in social linguistic interactions with each other. Since these games are not dependent on the representation used for objects or categories, the algorithm used here closely follows Steels's (2001) guessing game, which is one of several related language games (Vogt & Coumans 2003). Consequently, I will describe this aspect of the model only very briefly. A formalised outline of the algorithm is given in Table 4.4.

Precondition A speaker agent A_s and a hearer agent A_h are presented with a context O.
Procedure
1. A_s chooses a topic object o_t randomly from O.
2. A_s carries out a discrimination game on the o_t, obtaining a category c_s.
3. A_s utters a word w to express c_s.
4. A_h finds the best fit, if any, between w, a context object, o_g, and a category, c_g.
5. If $o_g = o_t$, then the game is a success.
6. Otherwise, if $o_g \neq o_t$ (or if no o_g was found), then the game is a failure. The speaker points out o_t to the hearer, and the hearer carries out a discrimination game on o_t to obtain a category c_h.
7. c_s, c_g and c_h are updated based in accordance with the success or failure of the game.

Table 4.4: Guessing game algorithm.

Steps 1 and 2 are identical to what happens for the discrimination game with a topic o_t, yielding a category c_s. In Step 3, the agent checks whether there are words associated with c_s. If there are, it utters the one with the strongest association to c_s. If there aren't, it invents a random word and utters it.

The hearer's task in Step 4 is the most difficult and computationally expensive. It first finds the set of all of its categories which have w associated with them, C_w. If there are no such categories, the guessing game is a failure. Otherwise, it finds the category c_g in C_w and the object o_g in the context for which the membership is maximised, i.e.,

$$(c_g, o_g) = \underset{(c \in C_w, o \in O)}{\arg\max} \; (membership_c(o)) . \tag{4.8}$$

However, if $membershi_{c_g}(o_g) < mem_{min}$, then c_g does not qualify as a candidate category for o_g, and the guessing game fails. Otherwise, the hearer indicates its guess o_g to the speaker.

In Step 5, the speaker then checks trivially whether the guess object is the topic. If it is, the game is a success. Otherwise, the speaker corrects the hearer by pointing to the topic. In that case, the hearer undergoes a discrimination game with the topic and context objects and obtains a category c_h (i.e., the same game that the speaker underwent in Step 2).

Once the guessing game ends, both speaker and hearer update their lexicon. This means updating the word associations of the agent's discrimination category c_s, the hearer's guessed category c_h, and, if the hearer failed to guess or guessed incorrectly, the hearer's discrimination category c_g. The specific adjustments made to each category depend on whether the guessing game was successful and whether the word w is already associated with the category. Figure 4.14 lists the exact changes that are made.

Success	**Failure**
If $w \in c_{s_W}$, $$\forall v \in c_{s_W}, \ v_a = \left\{ \begin{array}{ll} v_a + 1 & \text{if } v = w \\ v_a - 1 & \text{if } v \neq w \end{array} \right. .$$ Otherwise, $$\begin{aligned} c_{s_W} &= c_{s_W} \cup \{w\} \\ w_a &= 0. \end{aligned}$$	If $w \in c_{s_W}$, $$\forall v \in c_{s_W}, \ v_a = \left\{ \begin{array}{ll} v_a - 1 & \text{if } v = w \\ v_a + 1 & \text{if } v \neq w \end{array} \right. .$$ Otherwise, $$\begin{aligned} c_{s_W} &= c_{s_W} \cup \{w\} \\ w_a &= 0. \end{aligned}$$
$$\forall v \in c_{g_W}, \ v_a = \left\{ \begin{array}{ll} v_a + 1 & \text{if } v = w \\ v_a - 1 & \text{if } v \neq w \end{array} \right. .$$	$$\forall v \in c_{g_W}, \ v_a = \left\{ \begin{array}{ll} v_a - 1 & \text{if } v = w \\ v_a + 1 & \text{if } v \neq w \end{array} \right. .$$
	If $w \in c_{h_W}$, $$\forall v \in c_{s_W}, \ v_a = \left\{ \begin{array}{ll} v_a + 1 & \text{if } v = w \\ v_a - 1 & \text{if } v \neq w \end{array} \right. .$$ Otherwise, $$\begin{aligned} c_{s_W} &= c_{s_W} \cup \{w\} \\ w_a &= 0. \end{aligned}$$

Figure 4.14: Lexical adjustments: Adjustments made to category word associations after a guessing game, depending on whether the game was a failure or not. c_W denotes the set of word associations of category c, w is the word used in the guessing game, c_s is the resultant category of the speaker's discrimination game, c_g is the category that the hearer used in the guessing (if the hearer did make a guess), and c_h is the resultant category of the hearer's discrimination game (if communication failed).

CHAPTER 5

Results

5.1 Overview

In this chapter, I will present the results of simulations. All simulations have been run with the model described in Chapter 4, but under various sets of conditions, as defined by different parameter configurations (see Appendix A for a summary of the model parameters).

With the exception of Section 5.3.6, all of the simulations consisted solely of a large number of guessing games. A simulation consists of a sequence of epochs, each of which contains 100 guessing games. The number of epochs in a simulation is defined by the parameter $epoch_{num}$. For each epoch, the communicative success rate, κ, is defined as the ratio of the number of successful guessing games within the epoch. Plots of κ over the full set of epochs are used to demonstrate how the communicative success of the population changes. In order to obtain a single number that measures the final communicative success rate of a simulation, $\overline{\kappa}$ is defined as the average of κ over the final 10 epochs of a simulation. Another value that is sometimes given is λ, which is the average number of categories per agent at the end of the simulation, giving a simple idea of the complexity of the agents' category systems.

Due to space constraints, the results given here are not comprehensive, but give only a suggestive sample of what the model might reveal. In addition, recall that the model has quite a number of parameters, all of which can theoretically take an infinite number of different values. As a result, the parameter space is very large, and it is impossible to

search through all possible combinations of parameter values. Therefore, although many important patterns will most likely be missed, it is necessary here to limit the scope and methodology of the exploration.

The remainder of this chapter is thus organised in four sections. First, I will investigate how the model works in a few simple cases in some detail, in order to have a baseline that other results can be compared to. Second, I will briefly investigate the effects of each of the parameters, by varying one or two at a time, while keeping the others constant. These investigations will be extremely short, and are meant to be suggestive of the model's potential. Third, I will make a brief attempt at comparing agents' conceptual structures, and relating them to communicative success. Finally, I will look at a single word in a single simulation and survey the agents' categories which are associated with it.

5.2   Base cases

The model here uses a relatively complex representation of both world structure and agents' category systems. Consequently, the first issue to explore is whether a population of agents can even communicate successfully at all under the simplest possible conditions.

Accordingly, the first set of simulations was performed with a very simple parameter configuration. A completely random world (i.e., $dim_{clump} = 0$) was used with only 1 dimension, and minimal population and context sizes of 2 each. The other parameters were set to values which generally worked well in preliminary experiments (note that the parameters unmentioned here have no effect when the world is random and 1-dimensional): $mem_{min} = 0.9$, $conxt_{size} = 0.1$, and $adjust_{rat} = 0.0$. Fifty simulations with these parameter settings were run, each consisting of 200 epochs. Figure 5.1 shows the communicative success rate of the runs, κ, over the epochs, overlaid on one graph (only 20 runs are shown to reduce clutter).

It is evident that under these simplistic conditions, the agents quickly converge on a very high communicative success rate. In all the simulations, κ quickly reached 90% within 5 epochs (i.e., 500 guessing games). And in the majority of the runs, κ reached 95% and stayed close to it. For a few runs, communication would occasionally dip a little for some epochs, but even then κ did not drop below around 90%.

Figure 5.1: Random 1-dimensional world: κ for 20 superimposed runs under very simple conditions, with 2 agents and 2 context objects.

These, however, are the results from using a completely random world. What happens if, in contrast, we take a highly clumpy world? Figure 5.2 shows the results with nearly the same parameter values as for Figure 5.1, but with $dim_{clump} = 100$ and $dim_{parts} = 5$.

In this case also, convergence is even faster and communicative success even higher than in the random case. Table 5.1 shows the average and standard deviation of both $\overline{\kappa}$ and λ for each of the two plotted cases (but using 50 runs each). Four other versions of a structured world are also tabulated, based on different values for dim_{parts}.

dim_{clump}	dim_{parts}	$\overline{\kappa}$		λ	
0	N/A	98.0	0.9	323	11
100	1	99.2	0.3	325	13
100	2	99.2	0.3	326	13
100	5	99.2	0.3	340	10
100	20	99.0	0.4	394	12
100	100	98.9	0.3	303	12

Table 5.1: 1-dimensional world: Mean and standard deviation $\overline{\kappa}$ and λ values over 50 simulation runs in a 1-dimensional world under different world structure conditions.

It should be noted that dim_{clump} (and dim_{correl} as well) can theoretically take any value in $[0, \infty)$, but the mathematical calculations require higher and higher precision as the
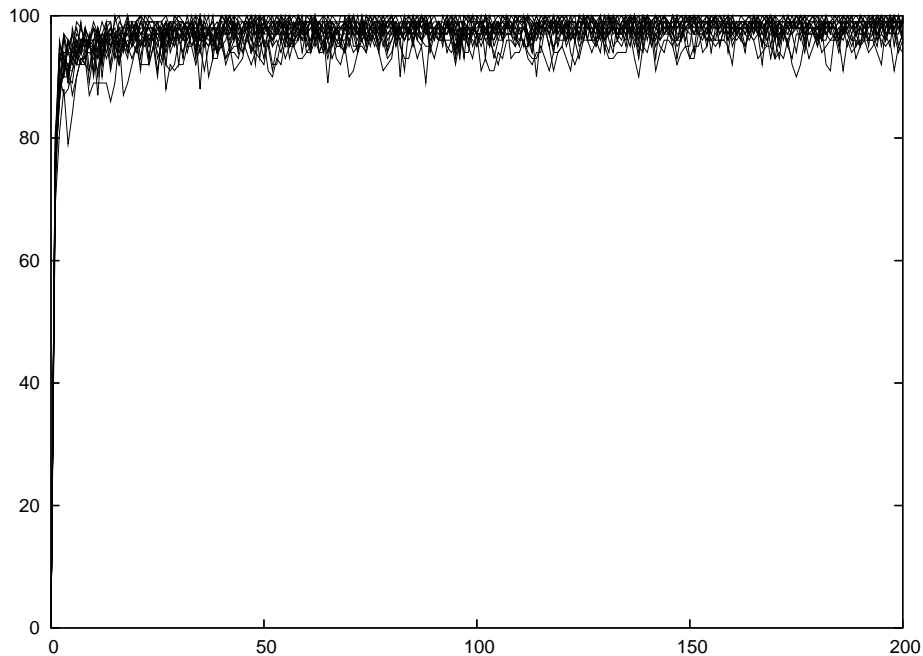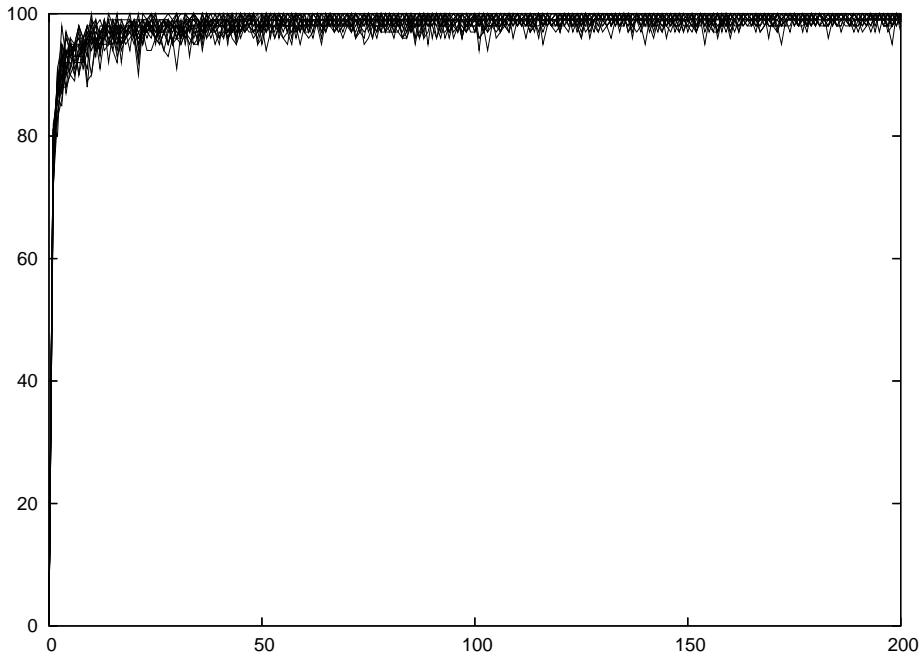
Figure 5.2: Clumpy 1-dimensional world: κ for 20 superimposed runs under very simple conditions, with 2 agents and 2 context objects, in a clumpy world with 5 partitions.

value goes up, and will eventually break down when using computers (for instance, incorrect calculations were very much in evidence when this parameter was set at around 500).

It is not difficult to imagine why a clumpy world is more conducive to better communication. In such a world, an agent's many category centres will tend to be found coinciding with the probabilistically dense areas of the world's structure. When $dim_{parts} = 5$, many of the categories prototypes were clustered around the midpoints of the world's dimension's partitions. Since different agents interact with the same world, they both end up with roughly the same kind of structure to their category system. On the other hand, when the world structure is totally random, the agent's category systems will depend largely on their personal discrimination game histories. Consequently, there need be a priori no similarity between the agents' category systems. In light of this, it is surprising that agents still manage to reach a high $\overline{\kappa}$. However, here they only need to distinguish between two objects, and they are still likely over time to associate words with categories that are similar to the lexicalised categories of the other agent.

Notice that the number of partitions in the world's one dimension didn't affect the results significantly. There is perhaps a slight drop in $\overline{\kappa}$ as the number of partitions goes up, but

this difference might not be statistically significant. Indeed, regardless of the number of partitions, $\overline{\kappa}$ is clearly larger than it is for the random world. This is interesting because we can take the number of partitions as a parallel to the number of objects in models with finite pregenerated worlds. For instance, in a highly clumpy one-dimensional world with five partitions, generated objects tend strongly to have dimension values very close to 0.1, 0.3, 0.5, 0.7, and 0.9. In other words, there are 5 "natural kinds" in our world, with some variation within each kind. From this perspective, we might imagine that the more partitions we have, the more natural kinds there are in the world. Having more partitions makes it more likely that context objects will belong to different natural kinds, and thus should facilitate discrimination. On the other hand, having more basic natural kinds makes it less likely that individual natural kinds will be deeply divided up into subcategories. As a result, when two objects do fall in the same partition, the system with fewer natural kinds will tend to handle them more successfully. The results here suggest that these two arguments more or less cancel each other out.

Perhaps surprisingly, communicative rates reach high levels even when there is only one partition. In such a world, each context is virtually guaranteed to present agents with very similar objects. Indeed, it is not at first obvious how a highly clumpy world with only one partition is different from a compressed version of a random world. However, even within the single partition, the world does have a bias towards its centre (0.5), and thus a structure, from which the agents appear to be benefiting.

Also, the number of categories that agents create is consistently between around 300 and 400, for all partition values, as well as the random world. This shows that a random world does not need more categories to discriminate objects from each other. The categories may be distributed differently when the world is clumpy, but then the agents need to make finer distinctions among similar objects. This result also shows that the stability of the category system is not related directly to communicative effectiveness.

However, a one-dimensional world is ultimately not very interesting, especially since this model is concerned with investigating the effects of both dimension correlations and clumpiness in the world on agents' communication. In most of the simulation runs presented in this chapter, three dimensions have been used. This number has been cho-

sen because it is enough to investigate the effects of dimension correlations, but not so much as to lead to overly complex and time-consuming computations.

Figure 5.3 shows the results of 20 simulation runs for each of four 3-dimensional world types, varying between extremes in both clumpiness and dimension correlation. Since I will use the same world structures in subsequent sections, I give these worlds names: W_r for a completely random world with no clumpiness and no dimension correlation ($dim_{correl} = 0, dim_{clump} = 0$), W_o for a highly correlated world with no clumpiness ($dim_{correl} = 100, dim_{clump} = 0$), W_l for a highly clumpy world with no dimension correlations ($dim_{correl} = 0, dim_{clump} = 100$), and W_s for a highly structured that is both clumpy and correlated ($dim_{correl} = 100, dim_{clump} = 100$).
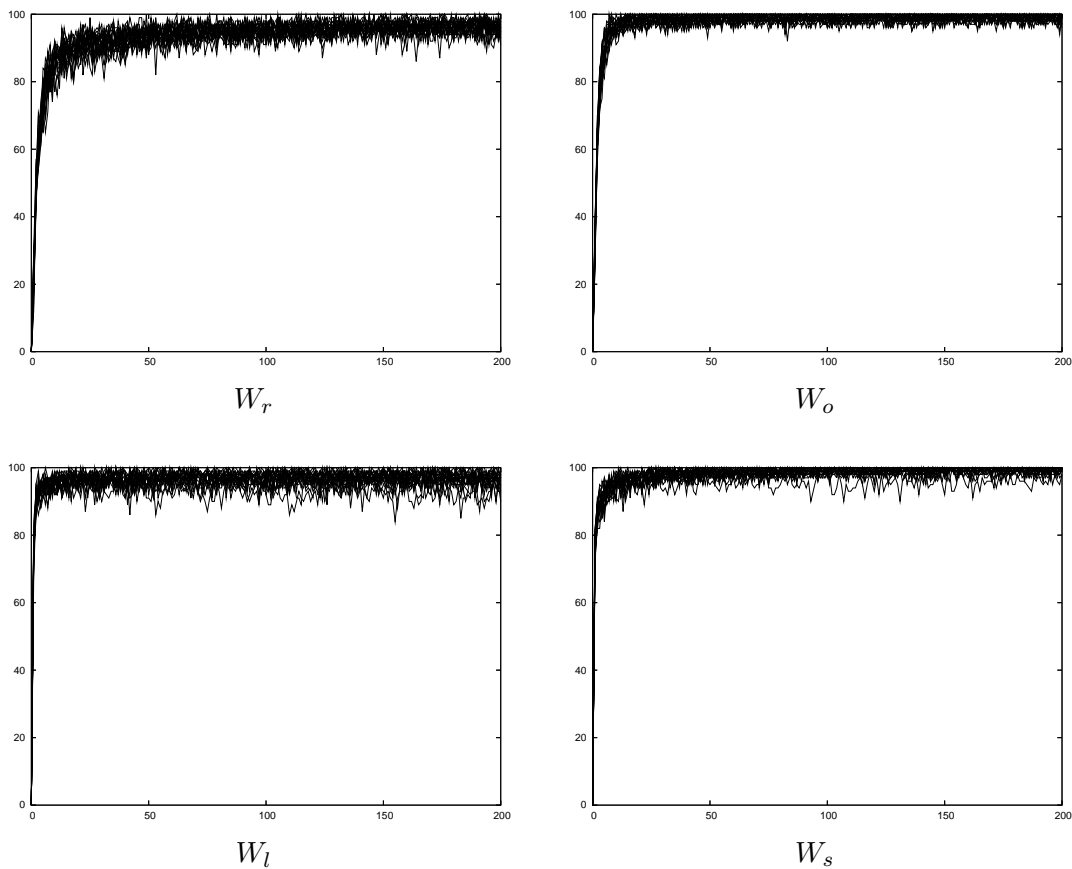

Figure 5.3: 3-dimensional worlds: κ for 20 superimposed runs in different world structures.

In all four cases, a high degree of communicative success is reached quickly and maintained, comparably to the 1-dimensional case. It is perhaps surprising that communication has not deteriorated much with more dimensions, as the world structure is more

complex. However, note that this higher complexity also allows for more variation between objects, which facilitates discrimination. Objects can have very similar values in two dimensions, but a greater difference in the third dimension may be enough to make them easily distinguished.

Table 5.2 shows averages over 50 runs for the four worlds depicted in Figure 5.3. As the tables suggest and can be vaguely seen on the graphs, a higher degree of world clumpiness helps communicative success get a little better, as it did in the 1-dimensional case.

World	dim_{correl}	dim_{clump}	$\overline{\kappa}$		λ	
W_r	0	0	96.3	0.7	505	29
W_o	0	100	98.9	0.4	343	15
W_l	100	0	96.6	1.4	281	10
W_s	100	100	99.1	0.5	331	12

Table 5.2: 3-dimensional world: Mean and standard deviation $\overline{\kappa}$ and λ values over 50 simulation runs in different world structures.

On the other hand, the effect of dimension correlation appears to be marginal in these cases. This may be because dimension correlation introduces a trade-off: you get a more structured and predictable world, but you also get significantly less variation. For instance, if the dimensions have 5 partitions, then an uncorrelated but clumpy 3-dimensional world would have, from an objective point of view, 125 (5^3) different general kinds of objects, one corresponding to each combination of partition centres across the three dimensions. But if the world was instead highly correlated, then only 5 of those 125 kinds would occur often (i.e., those for which the dimension values fell in the same partitions). We will see later, however, that dimension correlation does make a difference under some other conditions.

5.3 Parameter Effects

5.3.1 Population size

The simulations in the preceding section involved populations of only two agents. However, humans and other primates are highly social animals who live in groups. This section explores how communicative success is affected by larger populations.

Figure 5.4 plots the effects of population size under four different world structures. Each graph shows one run of 500 epochs each for population sizes of 2, 4, 8, 16, 32, 64, and 128. Clearly, population size in general decreases eventual communicative success, but only minorly.
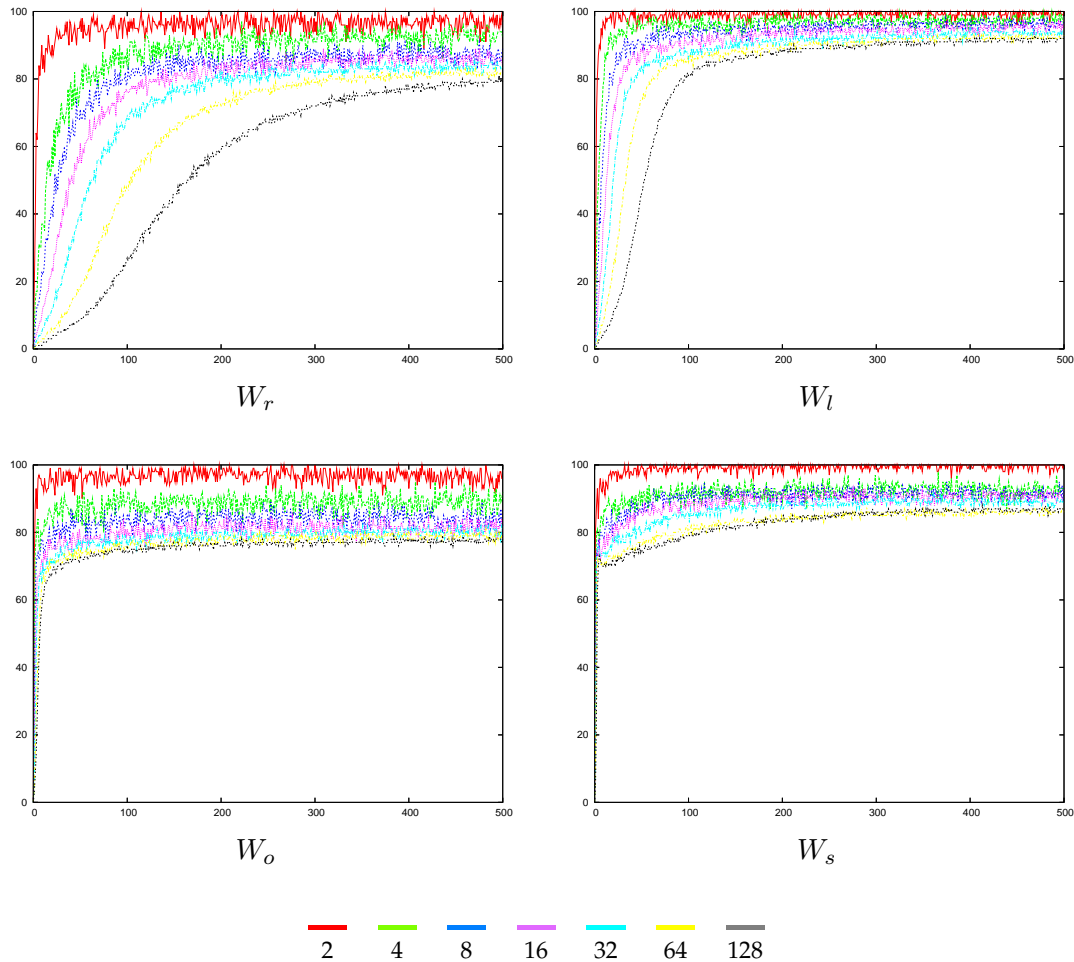
Figure 5.4: Varying pop_{size} in four different world structures.

Less obviously, high dimension correlation appears to have a significant effect on the rate of convergence, especially with higher populations. When the correlation is high (i.e., in W_o and W_s), κ reaches 70% within a few epochs, even in a very clumpy world with 128 agents. In contrast, when there is no dimension correlation, κ does not reach 70% until about 250 epochs in the random world and over 50 in the clumpy one. However, note that κ eventually does win out in the uncorrelated cases, while the correlated cases seem to reach a plateau. These effects can be seen in Figure 5.5, where the same runs for the four world types are plotted together on one graph each for the smallest (2) and largest (128) population sizes.
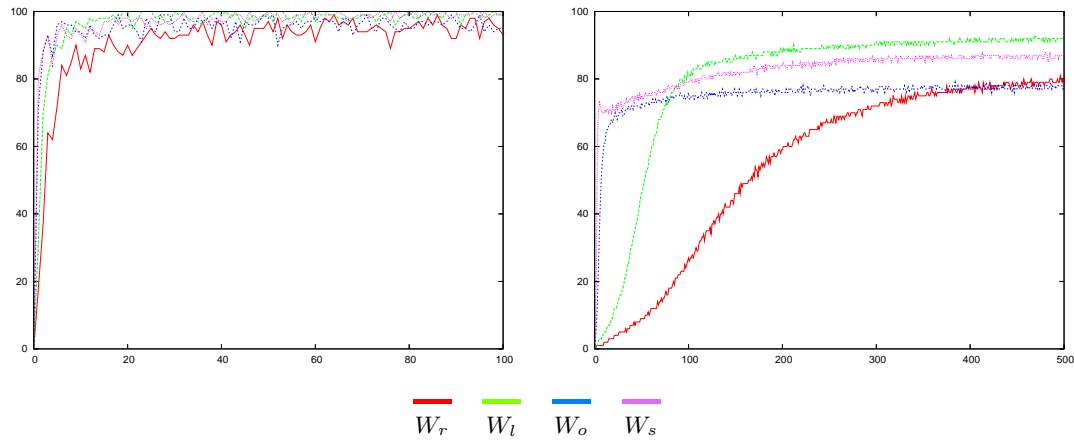
Figure 5.5: Population sizes and world structures: The left graph shows κ for 2 agents in four different world structures, and the right shows the same thing for 128 agents.

5.3.2 Context size

If agents' joint attention skills are not very advanced, they may need to choose from amongst a larger set of context objects to ascertain what speaking conspecifics are referring to. Figure 5.6 shows the effect of increasing context size in the four different world structures we have considered above.

As the graphs reveal, increasing context size deteriorates communication, and does so more significantly than increasing population size. However, this effect is more pronounced for uncorrelated worlds: correlated worlds appear to be relatively robust to larger contexts.

The reason for this appears to be related to the number of categories that agents create. In the runs shown here with context sizes of 16, λ was 42,427 and 38,176 for the two uncorrelated worlds, as opposed to only 1798 and 2682 for the two correlated ones. Since the simulations consisted of 50,000 guessing games (500 epochs of 100 games each), this means that in the uncorrelated cases, the discrimination games were resulting in the creation or splitting off of new categories around 80% of the time. But in any game when a category is just being created, it has not yet been lexicalised, so the game is guaranteed to fail. So the question becomes why context size should have such a deleterious effect on the discrimination game in an uncorrelated world. This result requires further investigation.
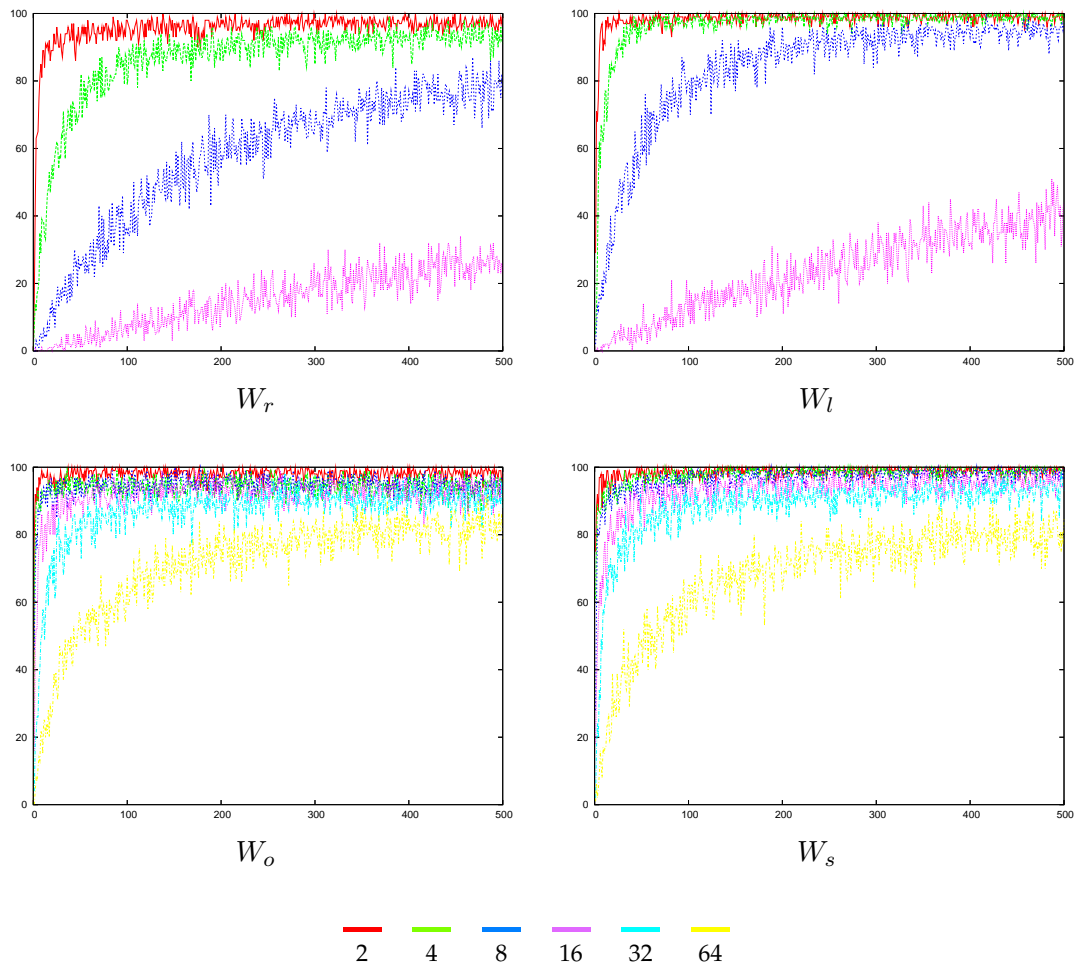
Figure 5.6: Varying $conxt_{size}$ in four different world structures. Note that simulations with context sizes beyond 16 were not run for W_r and W_l, because results were already very poor with 16.

5.3.3   World Structure

The simulations that I have presented so far have been run in somewhat extreme world structures, in order to investigate what effect the object generation parameters might have when other parameters are varied. But we should expect that the structure of our perceived world probably has a structure that lies somewhere in between. Dimensions are correlated, but not to the point where they match perfectly (e.g., most birds fly, but some do not). And some dimension values are more likely than others (e.g., animals typically have an even number of legs, unless they are injured). Of course, it is absurd to claim that the natural world corresponds to any specific values of these two parameters, since they are already idealisations in other ways (e.g., each dimension's distribution is identical). Nevertheless, if we grant that dimension correlation and clumpiness in

the world is substantial but not extreme, then it is important to look at how the model behaves with intermediate values for dim_{clump} and dim_{correl}.

Table 5.3 shows $\overline{\kappa}$ and λ from one simulation run of 500 epochs each for 25 combinations of dim_{correl} and dim_{clump}. In these runs, the population and context sizes were both set to 4, because this makes it easier to see patterns. The results suggest that communicative success is best when the world is highly clumpy but uncorrelated. On the other hand, the numbers of categories that agents form is highest when the world is completely random, i.e, neither clumpy nor correlated. These results agree with the findings shown for the extreme cases discussed in Section 5.2.

co\cl	0	1	5	20	100	co\cl	0	1	5	20	100
0	79.7	84.2	92.1	95.2	96.6	0	4039	3517	1718	1424	1476
1	73.0	79.2	93.2	95.2	95.6	1	4814	3982	1871	1492	1494
5	72.4	79.0	89.4	90.8	91.4	5	2470	2230	1450	1320	1302
20	73.5	73.9	82.7	83.9	84.3	20	691	698	736	740	782
100	78.8	80.0	82.1	81.2	84.7	100	684	678	748	743	781

Table 5.3: World structure: $\overline{\kappa}$ (left) and λ (right) for different world structures, defined by different dim_{clump} (row) and dim_{correl} (column) values.

However, the tables do not show the rate of convergence. In the Figure 5.7(left), dim_{clump} is fixed (i.e., $dim_{correl} = 5$), and κ is plotted for each of the five simulation runs with different values of dim_{clump}. Clearly, when the value of dim_{clump} is low, communicative success converges at a slower rate and stabilises at a lower value, although the value does not seem to make much difference between the final three cases where $dim_{clump} >= 5$.

In Figure 5.7(right), the opposite case is shown ($dim_{clump} = 5$ and dim_{correl} fluctuates). These results show that although higher dimension correlations eventually stabilise at lower communicative success values, they do lead to faster convergence. This is probably because high dimension correlation corresponds to less natural kinds, so it is easier to establish a simple lexicon based on them, but generally harder to distinguish objects from each other, since they are more likely to fall under the same general natural kind.

5.3.4 Dimensions

Most of the simulations looked at so far are based in a 3-dimensional world. As a result, agents perceive three features of objects, and build corresponding categories in three
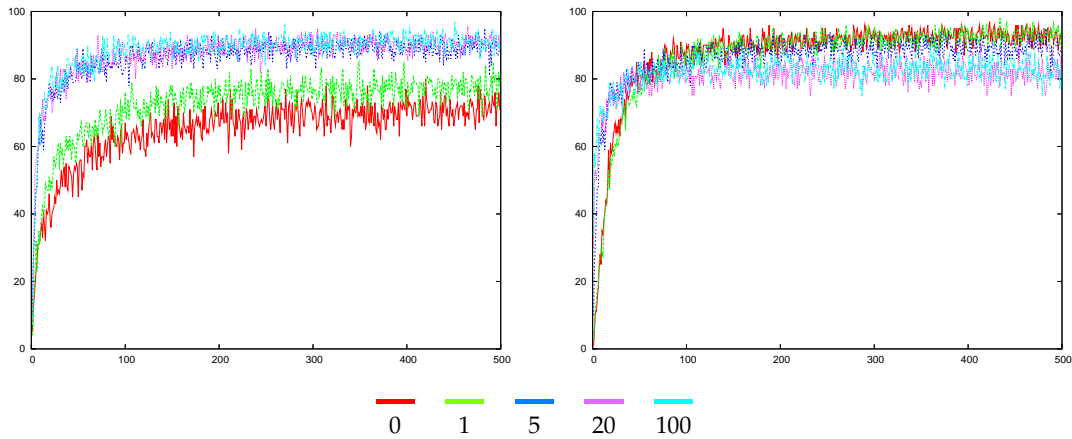
Figure 5.7: World structure: Varying dim_{clump} (left) and dim_{correl} (right).

dimensions. However, this number is arbitrary, and, in fact, humans most likely use much more than just three features when categorising objects in the world. Therefore, it is important to consider the scalability of the model to a large number of dimensions.

Figure 5.8 shows the effects of various numbers of dimensions in our four world structures (with $pop_{size} = 2$ and $conxt_{size} = 2$). The results are very clear. In an uncorrelated world, communication breaks down significantly when the number of dimensions goes past about 5: in an 8-dimensional world, the agents barely even reach 10% success rates, even after 50,000 guessing games. In contrast, a highly correlated world is quite robust to and arguably even unaffected by higher numbers of dimensions.

Upon reflection, this result is not surprising. In a completely uncorrelated world, values in one dimension can occur with all values in the other dimensions (mitigated by the dimension clumpiness of course). In other words, we get a total set of dimension combinations (Garner 1974), of the type that Rosch was arguing does not occur in the natural world (e.g., things with wings don't have leaves). At the other extreme, if the world was completely correlated so that the feature values of objects were identical in every dimension (which is not possible in this model since this would require $dim_{correl} = \infty$), then it would be virtually just like having a 1-dimensional world. Indeed, one could then think of the N dimensions of the world as actually 1 dimension which was simply a conjunction of the original N ones. In practice, when $dim_{correl} = 100$, there are occasionally cases where objects have very differing values in different dimensions, but generally they are nearly identical.
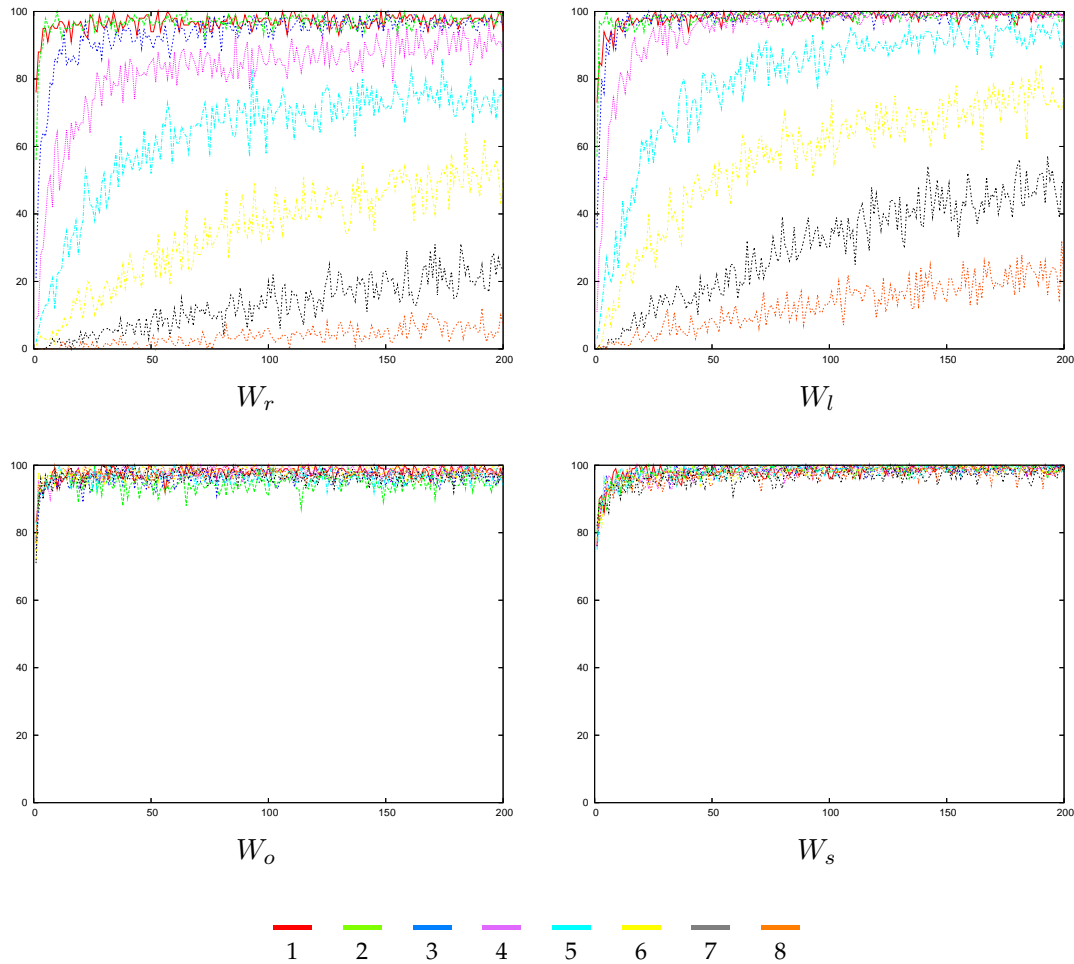
Figure 5.8: Varying dim_{num} in different world structures.

Under this logic, the more correlated the world, the more stable the category system ought to be with an increasing number of dimensions. In fact, this is perhaps best exemplified by the number of categories that agents create, which is shown in Table 5.4 for the simulations shown in Figure 5.8. Note that the average number of categories explodes quickly in the uncorrelated worlds, but is roughly constant (or even decreases) when the dimensions are highly correlated.

5.3.5 Category membership and adjustment

The category membership parameters (mem_{min} and mem_{conxt}) and category adjustment parameter ($adjust_{rat}$) are annoying but necessary parameters. Ideally, their actual values would be inconsequential, so that we could fix them and focus on more interesting parameters. However, intuitively this seems unlikely, since they affect what choices are

d\w	W_r	W_l	W_o	W_s
1	308	355	335	344
2	262	362	300	322
3	569	340	282	349
4	1766	806	274	328
5	5470	3102	264	356
6	12016	7097	266	370
7	17222	13330	264	379
8	19254	17214	267	365

Table 5.4: World dimensions: λ for different combinations of world dimensions (rows) with world structures (columns).

made during discrimination games and the structure of the resulting categories. In particular, the relative values of the two membership parameters should make an impact.

Table 5.5 presents charts for the four world types we have been looking at, showing $\overline{\kappa}$ and λ for different combinations of mem_{min} and mem_{conxt} ($adjust_{rat}$ has been set to 0 for simplicity). Although this is a very sparse sampling of the possible parameter configurations and the $\overline{\kappa}$ values come from only one run each, there does appear to be a general pattern that the best results occur with higher values of mem_{min} and lower values of mem_{conxt} (although this is certainly violated in W_o).

c\m	0.5	0.7	0.9
0.1	80.1	87.6	96.4
0.3	75.3	80.9	93.1
0.5	78.9	80.8	93.3

W_r

c\m	0.5	0.7	0.9
0.1	83.2	94.5	99.6
0.3	77.3	89.9	97.9
0.5	76.4	84.1	97.4

W_l

c\m	0.5	0.7	0.9
0.1	83.5	85.7	91.6
0.3	86.5	85.0	93.7
0.5	86.8	87.9	87.3

W_o

c\m	0.5	0.7	0.9
0.1	87.8	95.8	98.8
0.3	92.6	90.9	97.4
0.5	89.9	92.9	94.6

W_s

Table 5.5: Category membership: $\overline{\kappa}$ for different combinations of mem_{min} (m) and mem_{conxt} (c), in four different world structures.

As for $adjust_{rat}$, Figure 5.8 shows the results for the simple cases of 2 agents and 2 context objects over 200 epochs. $adjust_{rat}$ is sampled at values of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. Almost across the board, the lower the adjustment values the better. In particular,

full adjustment (i.e., $adjust_{rat} = 1.0$) leads to quite poor results. This is a somewhat surprising result, because it suggests that the agents' category dimension sensitivities are best left constant once created. In that case, further experience can only shift the prototype (as it is not affected by this parameter), but not modify the category's shape. It would seem, then, that the benefit of having a category maintain the same extension in the world (and perhaps more importantly, the same extension as its fellow agents' corresponding categories) outweighs the benefit of adapting one's category structure due to experience. This begs the question of whether the model would do even better if the prototype didn't shift either, but currently no parameter controls this behaviour, so it cannot be tested.
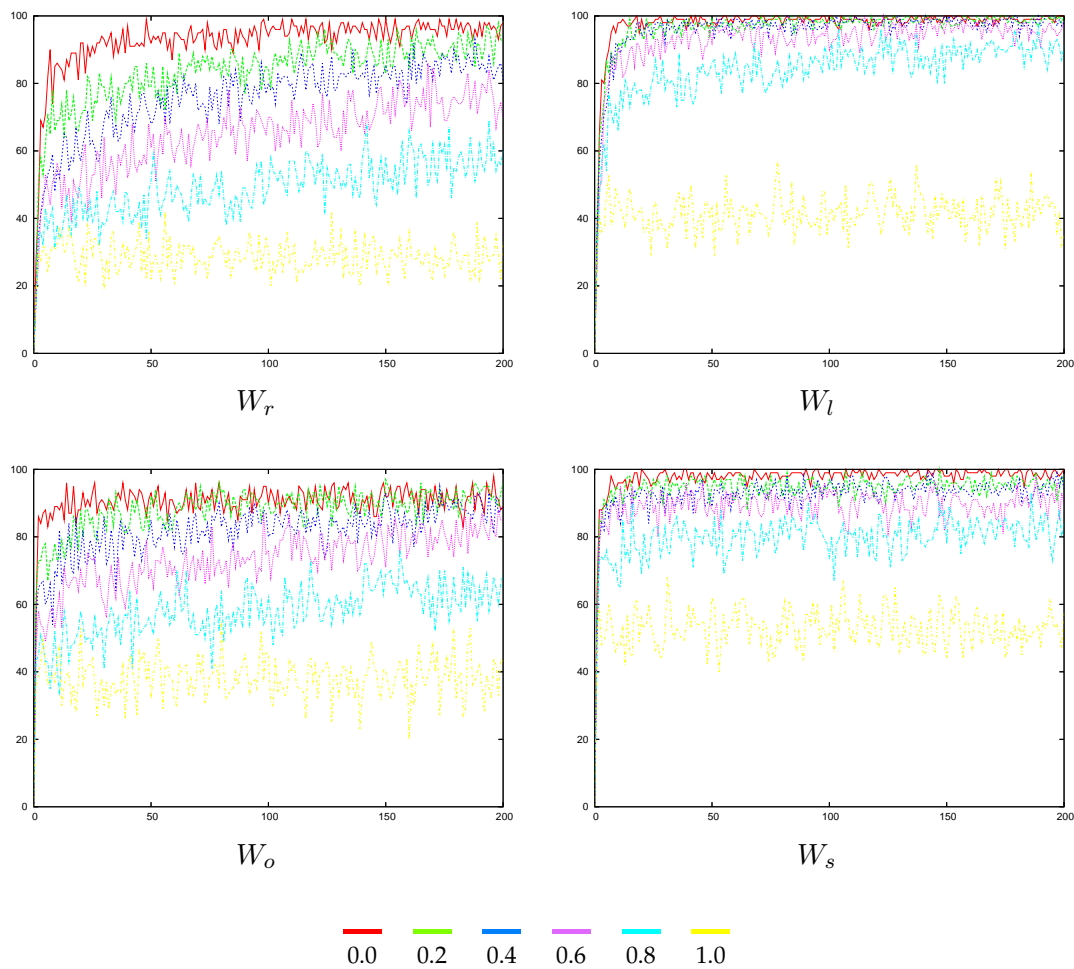

Figure 5.9: Varying $adjust_{rat}$.

These "optimal" values for the three parameters (i.e., $mem_{min} = 0.9$, $mem_{conxt} = 0.1$, and $adjust_{rat} = 0$) have been used for the other simulations, without further justification.

5.3.6 Non-linguistic games

There is a potentially significant objection to the set up of the simulations presented thus far: all the simulations have consisted solely of guessing games. This means that all agent categorisation occurs in a social linguistic setting involving two agents and a common environmental context. But clearly not all of human categorisation is due to linguistic discourse, especially in infancy when children do not yet produce or understand language but are able to interact with and categorize objects (Mandler 2004). Upon closer inspection of the guessing game algorithm (see Table 4.4), there is good reason to believe that this setup makes the agents' jobs too easy. Notice that in a guessing game, there are always 1 or 2 discrimination games. The speaker always undergoes a discrimination game. While the hearer undergoes no discrimination games if it guesses correctly, if it fails then it undergoes *exactly the same* discrimination game as the agent did (the same topic and the same context). Consequently, in the first few guessing games of a simulation, which inevitably fail, the agents (in a 2-agent population) build up exactly the same category systems, which only begin to diverge with the first guessing game success. Even after that, however, a failed guessing game results in the same discrimination game for speaker and hearer, and if $adjust_{rat}$ is 0, as in most of the simulations above, the category systems will only diverge significantly when the agents create or split off new categories.

For this reason, a few simulations were run to see if the addition of private discrimination games would have a negative impact on communication. The parameter $preling_{dgs}$ defines the number of games prior to the beginning of the epochs (as parallels to infant prelinguistic experiences), while $epoch_{dgs}$ specifies the number of discrimination games within epochs (as parallels to adult non-linguistic experiences). Table 5.6 shows the values of $\overline{\kappa}$ after 200 epochs for four different world structures, with different values of $preling_{dgs}$ and $epoch_{dgs}$. Although there does appear to be some deterioration when private discrimination games are included, the effect is not large, with all simulations reaching at least 90%. This is even the case when agents undergo 5000 initial discrimination games each and experience five times as many discrimination games as guessing games in each epoch ($epoch_{ggs} = 100$ in all the simulations). Although it is possible that

the effects may be more significant in more complicated world structures and larger population and context sizes, these preliminary results suggest that agents can communicate successfully even when they have different individual categorisation histories.

i\e	0	100	500
0	96.5	93.6	92.3
5000	93.8	92.4	93.1

W_r

i\e	0	100	500
0	98.6	98.4	98.1
5000	97.7	98.1	98.2

W_l

i\e	0	100	500
0	97.5	91.5	91.0
5000	92.1	90.0	92.0

W_o

i\e	0	100	500
0	99.0	94.7	94.8
5000	95.4	96.2	94.6

W_s

Table 5.6: Non-linguistic games: $\overline{\kappa}$ for different combinations of $preling_{dgs}$ and $epoch_{dgs}$, in four different world structures.

5.4 Dimension Specialisation

Unlike discrimination trees, in which the creation of a new category at any point can only occur in a finite number of ways (Smith 2003a), category creation in this model is highly undetermined, and depends largely on the randomly generated context and topic object values. Indeed, the categorisation history of agents in each simulation is unique. As a result, it is difficult to analyse the structure of the agents' category systems. It is not impossible, however, and in this section I present one example method, which simplistically measures the relative refinement of an agent's dimensions.

We first take an agent's category system at the end of a simulation, and for each dimension, we calculate the average of the dimension sensitivities across all of the agent's categories. If N is the number of dimensions dim_{num}, and C is an agent's set of categories, then we have

$$\forall i \in \{1, .., N\}, \overline{s_i} = \sum_{c \in C} s_{c_i}. \tag{5.1}$$

Then we define the dimension specialisation ϵ of an agent to be the ratio between the largest difference among the $\overline{s_i}$'s and the largest s_i:

$$\epsilon = \frac{\max_{i,j\in\{1,..,N\},i\neq j} |\overline{s_i} - \overline{s_j}|}{\max_{i\in\{1,..,N\}} \overline{s_i}} \tag{5.2}$$

This measure gives us a rough idea of how specialised the agent's most sensitive dimension is compared to its least sensitive dimension. A value of 0 would mean that all dimensions are on average refined to exactly the same degree. As ϵ approaches 1, one dimension becomes much more refined relative to another.

Table 5.7 shows the average values of ϵ across both agents and all 50 simulation runs from the three-dimensional base case in Section 5.2, along with the standard deviations, expressed as percents. Notice that there is more specialisation in the more random worlds, but even there, the specialisation is quite small. This suggests that agents in this model do not normally develop one dimension in significantly more detail than others. Such a finding conflicts with that of (Smith 2003*b*), where agents sometimes became "experts" of different dimensions, and yet still managed a high communication success rate.

World	μ	σ
W_r	12.9%	8.2%
W_l	9.6%	5.3%
W_o	4.8%	3.1%
W_s	3.1%	1.8%

Table 5.7: Dimension specialisation: Mean and standard deviation of ϵ calculations in four different world structures.

5.5   A single word in a population

Until now, I have treated word meanings as associated with concepts. The relations between a single word to different concepts has not been examined. But ultimately, it would be useful if we could also treat words the way that dictionaries do, i.e., by reversing the relationship and considering what concepts are associated with a given word. Moreover, since words are public, it makes sense in this context to look not just at a single agent's associations, but on those of the entire population. Therefore, I will now look at the case of a single word from a single simulation run, and examine what categories it is associated with across the entire population.

The simulation was run with a population size of 4, a context size of 5, and a world with three uncorrelated but fairly clumpy dimensions with five partitions. The epoch-related parameters were set to $preling_{dgs} = 1000$, $epoch_{num} = 500$, $epoch_{dgs} = 100$, and $epoch_{ggs} = 200$. By the end of the 500 epochs, the agents communicated with a 90.5% success rate, and had category systems averaging 3021 categories per agent. Figure 5.10 shows κ for this simulation.
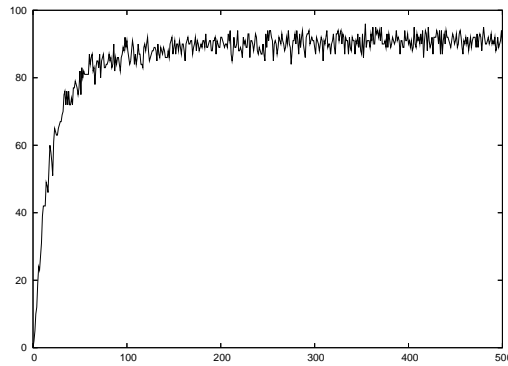
Figure 5.10: κ for the simulation used in the single word example.

Table 5.8 lists all the categories of all the agents (at the end of the simulation) that are associated with the word "bax". Each category is listed with an index for convenience, a letter identifying an agent, the score of the category's association with the word, whether or not "bax" is optimal for this category (i.e., whether it has the highest associated score), and the three prototype dimension values and dimension sensitivities that define the category.

The most striking thing about this list of categories is its length: among only 4 agents, there are 30 categories associated with the word "bax". However, an inspection of the category prototypes reveals that, asides from one exception (i.e., category #15), all the categories for which "bax" is the primary word have prototypes lying close to [0.7,0.9,0.3]. On the other hand, most of the categories for which "bax" is not optimal have prototype centres lying elsewhere in the perceptual space, with the exception of a few of the categories of agent D. Notice also that there are a few cases here of subcategories being associated with the same word (e.g., category #5 is a subcategory of category #1). Finally, the number of categories per agent which are associated with "bax" varies widely from 2 for agent B to 16 for agent D. Although this seems unreasonably large, agents still converge on successful communication, so apparently it does not pose a serious problem.

Ind	Agt	Scr	Opt	p_1	s_1	p_2	s_2	p_3	s_3
1	A	229	Y	0.69992	0.27690	0.89916	0.18032	0.29525	0.07449
2	A	7	Y	0.71183	0.00850	0.91190	0.18785	0.31815	0.08598
3	A	3	Y	0.70080	0.08967	0.89737	0.09262	0.29614	0.01061
4	A	1	Y	0.69925	0.18482	0.88870	0.07968	0.31266	0.00615
5	A	0	Y	0.71996	0.01169	0.89916	0.18032	0.29525	0.07449
6	A	0	N	0.67983	0.08118	0.88499	0.08568	0.33482	0.02429
7	A	-7	N	0.89806	0.17253	0.50078	0.08935	0.70099	0.08965
8	A	-210	N	0.50126	0.08953	0.29953	0.08326	0.30065	0.08382
9	B	238	Y	0.69861	0.09054	0.89789	0.18602	0.29864	0.09254
10	B	7	Y	0.70148	0.00578	0.89789	0.18602	0.29864	0.09254
11	C	188	Y	0.70581	0.07531	0.89252	0.08047	0.29873	0.19896
12	C	44	Y	0.69522	0.01080	0.91346	0.10011	0.30091	0.09416
13	C	16	Y	0.68897	0.08825	0.89792	0.00382	0.29749	0.09338
14	C	7	Y	0.70199	0.09020	0.91392	0.00637	0.29138	0.08549
15	C	0	Y	0.49952	0.08903	0.30989	0.00347	0.30118	0.09335
16	C	-2	N	0.89846	0.00645	0.46625	0.07776	0.69778	0.09132
17	D	63	Y	0.69526	0.08021	0.89779	0.17984	0.29600	0.00734
18	D	44	Y	0.70027	0.00556	0.90124	0.10119	0.30311	0.08894
19	D	29	Y	0.68776	0.17648	0.88932	0.08631	0.30960	0.00907
20	D	24	Y	0.70652	0.09448	0.88506	0.07985	0.30137	0.00376
21	D	11	Y	0.69495	0.01723	0.92959	0.18909	0.30535	0.01295
22	D	9	Y	0.71001	0.08806	0.89790	0.18986	0.28441	0.00726
23	D	6	Y	0.70931	0.08775	0.91236	0.18216	0.30345	0.00151
24	D	2	Y	0.70584	0.08694	0.89680	0.00563	0.30803	0.00707
25	D	0	Y	0.72212	0.08124	0.91490	0.00017	0.30447	0.09515
26	D	0	Y	0.71086	0.00527	0.90498	0.00302	0.27690	0.00144
27	D	0	Y	0.73241	0.01897	0.91581	0.27312	0.29137	0.00486
29	D	0	N	0.69101	0.08362	0.90931	0.09986	0.33165	0.02106
28	D	-1	Y	0.89826	0.00482	0.50178	0.09494	0.72184	0.01053
30	D	-1	N	0.68889	0.00353	0.87014	0.36764	0.30548	0.08689
31	D	-2	N	0.71608	0.10700	0.88624	0.17329	0.28556	0.07710
32	D	-133	N	0.50186	0.08953	0.29802	0.08326	0.30151	0.08382

Table 5.8: A word in a population: All of the categories associated with the word "bax" for all the agents in a population. Ind is a convenient index for a category, Agt is a letter identifying the agent, Scr is the word score, and Opt is Y if "bax" is the word with the highest score for the category. The p_i's and s_i's are the prototype dimension values and dimension sensitivities, respectively.

Indeed, the results suggest that perhaps it is too simplistic too think of words as being attached to individual simple categories. Words seem to associate themselves with many closely clustered categories, so that they could be said to be broadly associated with a subspace of the perceptual space which more or less contains those categories. This subspace could be thought of as a higher-level cognitive entity corresponding more closely

to a concept. Under this interpretation, while a perceptual category has a single proto-type, a concept is potentially associated with many. In fact, one could speculate that such a concept has a representation that brings together exemplar and prototype theory, as it is associated with multiple points in the perceptual space which originated from actual encountered instances.

CHAPTER 6

Discussion

6.1   Comparison to past results

Since previous models have used very different representations of the world and agents'
categories, many of the results from Chapter 5 do not have obvious analogs. However,
some of the experiments do have close parallels with previous work, and have obtained
similar results.

Most fundamentally, many researchers have found that a population of agents who start
with empty category systems and lexicons can converge on a high degree of communica-
tive success. Moreover, the shape of the convergence curves tends to be similar: starting,
inevitably, at a communicative success of 0%, success increases at a fast rate initially
and then seems to level off at or asymptotically approach some value (e.g., Smith 2003*a*,
Steels & Belpaeme 2005, Webb 2005).

However, both the speed of convergence and the eventual communicative success rate
tend to be worse than in other models. For instance, in Steels & Kaplan (1998), a popula-
tion of 10 agents reached a communicative success rate of 90% within about 2000 games,
and eventually reached 100%. In contrast, simulations run using the current model with
that many agents never reach 100%, often do not attain a success rate of 90%, and usu-
ally take much longer to converge (although this depends on the world structure: see
Figure 5.4).

The main reason for this is that every object an agent encounters is a new, previously un-seen object. Categories cannot get specially tuned to match individual objects precisely, and new categories do not cease to be created in discrimination games (as it is always possible that a new context occurs with two extremely similar objects, for example). On the other hand, in other simulations, the world is typically finite and usually consists of a fairly small number of objects (e.g., in the example above from Steels & Kaplan 1998, it was 10). Over the course of many games, the same objects and even the same context can recur. As a result, it is easier for the agents to converge on a good set of categories and words that may be adapted to that particular set. This can be seen from the fact that when such models introduce new objects during a simulation, the communicative success goes down significantly for a time, even if it had been at 100% (Steels & Kaplan 2002).

Another reason why the categorisation model used here may be intrinsically slower is that there is higher potential for variation between agents compared to other models. For instance, Smith (2003*a*) has studied discrimination trees in detail, and demonstrated how variation in their growth is quite constrained. Indeed, after a finite number of discrim-ination games, there are only a finite number of possible category systems. In contrast, in this model, since objects and prototypes are both real-valued vectors in a continuous space, there is a mathematically uncountable number of possible category systems even after only a single discrimination game.

This model has yielded similar general results to other models in terms of the effects of population size, context size, and number of dimensions. Not surprisingly, for each parameter, higher values generally result in slightly lower communicative success and a slower rate of convergence (e.g., Smith et al. forthcoming, Vogt & Coumans 2003, Divina & Vogt forthcoming). However, I have shown that the significance of varying these and other parameters depends on the world structure. For instance, an increase in the number of dimensions had an overwhelmingly negative effect, but only in uncorrelated worlds. This result resembles that of Webb (2005), who found that a well-structured world resulted in higher communicative success.

Finally, the relationship between meaning similarity and communicative success has been studied in various models (Smith 2003*a*) but has not yet been looked at here in de-

tail. Preliminary investigations suggest that, in contrast to Smith's (2003*a*) simulations, agents do not become increasingly sensitive to particular dimensions. On the other hand, the more general finding that communicative success is correlated but consistently superior to meaning similarity between agents most probably holds here as well, at least in the cases where communicative success is very high, as it is difficult to see how the agents' category systems could develop so closely given the huge range of possibilities.

Indeed, one of the potential drawbacks of this model is that it is difficult to analyse and compare agents' category systems, even compared to other prototype models. For instance, Belpaeme (2002) developed mathematical measures for comparing agents' prototype-based category systems. However, the measures he used cannot be directly adapted because they rely on the fact that the gaussian functions he uses to represent prototypes have fixed and equal widths. As demonstrated in Section 5.4, the agents' conceptual structures are not an inaccessible black box, but more consideration would be required in developing more sophisticated measures.

6.2   Evaluation

It is important to step back and scrutinise how lexical emergence models in general, and this model in particular, actually relate to real language evolution. To do this, I return to Noble's (1998) justification of computational simulations (see Figure 3.1). Recall that the usefulness of computational modelling rests on a relatively faithful mapping between A_R and A_M, and a comparison between E_M and E_R. In our case, this amounts to asking two questions, which I will try to take a brief stab at here:

(1) How well does the model reflect our knowledge and theory about how lexical emergence really occurred?

(2) How closely do the simulation results correspond to the actual one-word stage that presumably occurred in real language evolution?

Question (1) has been addressed intermittently throughout this dissertation. The categorisation of agents in this model is based on an idealised model of prototype theory, that incorporates the ideas of prototypical members and graded membership (see Section 2.3.2). At any one time, however, categories do not have fuzzy boundaries, although,

as we have seen, it has been disputed whether human categories do either. Three points must be borne in mind, however. First, the idealisations that I have used may be significant: for instance, agents all use the same dimensions when categorising objects, and detect them perfectly. Second, prototype theory has been challenged with moderate success by other concept theories, such as exemplar and theory theories (Murphy 2002). Third, the hominins who developed a holistic language did not necessarily categorize in the same way as we do now: although some studies have suggested that animals do use prototypes as well (White et al. 1993), the evidence is still very scant.

The model here has also explored worlds that are infinite, but imbued with a fair amount of structure, that could be easily manipulated. The unbounded number of objects in the world reflects the fact that we are constantly encountering new objects (although perhaps typically belonging to the same category as previously seen objects), and that even the same object can appear differently in different circumstances. The clumpiness and correlation of the dimensions reflects the fact that not all dimension values are equally likely, and that some feature combinations are more common than others (for instance, it is rare to see dogs with three legs, and one never sees dogs with wings). These are idealisations, however, and it would be impossible and even absurd to attempt to pin down values for the world parameters which would reflect how the world is actually structured.

As for the simulation dynamics, I have borrowed the general mechanisms directly from the line of work begun by Steels (1996*a,b*). The discrimination game presents an idealised procedure for agents to develop categories based on interactions with the environment, and the guessing games add on top of this a two-agent interaction using rudimentary linguistic communication. However, these are not the only kinds of interactions possible: indeed, Vogt Vogt & Coumans (2003) have compared the guessing game with other candidate language games and discovered that their results can vary significantly. In this model, the details of the algorithms also needed to be modified to conform with the types of category representations used. Although I have attempted to make decisions that were environmentally driven and as psychologically plausible as possible, the low-level decisions were sometimes nearly arbitrary and alternative implementations were not attempted.

Question (2) is even trickier to handle. We are meant to evaluate the model by comparing the simulation results with the phenomenon under study. Yet that phenomenon itself is not well understood. If we were modelling the acquisition of a language by a human child, then at least we could compare the results to the behaviour of actual language learners. In the case of language evolution, however, and lexical emergence in particular, we have no direct access to the actual phenomenon under study. Indeed, although one-word stage has been advocated (Jackendoff 1999) and often presupposed (Gong et al. 2004), we cannot be sure what such a stage was like, how quickly it emerged, or even whether it existed at all. As a result, it is most difficult to relate simulation results to actual language eovlution.

In particular, there is a tendency to search for the conditions under which behaviour is optimised, which in this case normally means that a high communicative success rate is reached. But how do we judge how successful linguistic communication actually was in its earliest stages? On one hand, it seems like there would be no motivation to invent and learn words if they were not useful right away, inviting us to believe that a high rate should be sought. But we could equally well argue that the last prelinguistic hominins had survived evolution without linguistic means, and that language may have initially been merely an optimisation on communication, rather than a requirement. From this point of view, even low rates of occasional lexical success are useful, in that they are better than no attempts and thus no success at all. Hominins who did "try" to start language off would thus fare better than those who didn't. Of course, this argument too requires caution, as it presumes that even initially language users held a selective evolutionary advantage.

Consequently, the model I have used here cannot independently answer any general questions with authority. To really learn how lexical emergence occurred in the evolution of language, we must continue to accumulate converging evidence from a vast variety of different disciplines, and to continue to build more realistic computational models. I believe that the current model has, however, contributed to our understanding of lexical emergence by replicating past results with a more realistic representation, and added support for the self-organisation explanation.

6.3   Future work

This dissertation presents a new model of lexical emergence that has just begun to be explored, and further research could proceed in several directions. The parameter space should be explored in more detail, with different combinations being tested, especially under less ideal conditions. Simulation results should be analysed statistically to determine the significance of the results, which can at present only be claimed to be suggestive. The simplistic one-generation model could be modified to allow for agent removal and addition (to simulate evolution), and the learning of a lexicon by new agents from older agents (Smith et al. 2003). Noise could be added to the model, to see if the idealisations used here are significant (Steels & Kaplan 1998). Efforts should be made to motivate low-level decisions in the categorisation algorithm with psychological findings. Real human language experiments should be performed and the results integrated with those of the simulations. Other models could be designed using the same model of world generation but different categorisation theories, such as exemplar-based models (following Nosofsky et al. 1994), as well as the previously used discrimination trees (Smith 2003*a*). The prototype model could be used and tested with robots, as in the Talking Heads Experiments (Steels & Kaplan 2002). The effects of Lexicalisation of concepts could be explored Lupyan (2005), and integrated with the developmental studies on the subject (Bowerman & Levinson 2001, Gopnik & Choi 1990). Memory effects could be built in, such as enhancement or fading of categories and word forms based on usage history (Belpaeme 2002). And in general, such computational research should inform and be informed by diverse fields such as psychology, linguistics, philosophy, biology, archaeology, neuroscience and computer science.

6.4   Conclusion

Many computational models have shown that lexical emergence can occur via a process of self-organisation among a population of agents, but they have tended to use unrealistic representations. In this dissertation, I have presented a more realistic model of agent categorisation, and exposed agents to an infinite, clumpy and correlated world. Although the preliminary results are only suggestive, they do show that lexical emergence does emerge under various conditions. In time, the agents do converge on a coherent lexicon, so that they are able to identify the objects that their interlocutors are

referring to with a high level of success. However, the rate at which this convergence occurs, the level that it eventually reaches and other aspects of the simulation dynamics are sensitive to the different model parameters. In particular, I have argued that world structure, an issue largely unaddressed in other models, has a significant effect on how scalable the system is in terms of other parameters. However, much exploration remains to be done to explore the dynamics of this model in more detail, and it should be considered as just one of many steps to validating the growing body of simulation results.

APPENDIX A

Simulation Parameters

Below is a full list of the simulation parameters used in the model, along with their range of possible values and a brief description. All parameters are set once at the beginning of each simulation with a configuration file, and kept constant throughout the simulation.

Parameter	Values	Description
pop_{size}	$\{2, 3, ...\}$	Number of agents in the population.
$conxt_{size}$	$\{2, 3, ...\}$	Number of objects in each environmental context.
dim_{num}	$\{1, 2, ...\}$	Number of object dimensions which are detected by agents.
dim_{parts}	$[1, \infty)$	Number of partitions into which the dimensions are divided.
dim_{clump}	$[0, \infty)$	Degree of clumpiness within a dimension.
dim_{correl}	$[0, \infty)$	Degree of correlation between the dimensions of objects.
mem_{min}	$[0, 1]$	Minimum category membership for an object to be considered a member of a category.
mem_{conxt}	$[0, 1]$	"Target" category membership for the non-topic context objects during category creation.
$adjust_{rat}$	$[0, 1]$	Ratio by which a category's dimension sensitivities are changed during category adjustment.
$preling_{dgs}$	$\{1, 2, ...\}$	Number of discrimination games per agent prior to the onset of the epochs.
$epoch_{dgs}$	$\{1, 2, ...\}$	Number of discrimination games per agent per epoch.
$epoch_{ggs}$	$\{1, 2, ...\}$	Average number of guessing games per agent per epoch.

APPENDIX B

Mathematical details of object generation

In this appendix, I will demonstrate the mathematical details of how an object is generated by looking at a specific world configuration, showing how the probability distribution functions are defined, and using them to generate a single object.

Suppose that the world structure parameter values for a simulation are dim_{num}= 4, dim_{parts}= 5, dim_{clump}= 9.2, and dim_{correl}= 4.5. As we will see, this configuration will result in objects that have four fairly correlated dimensions, with dimension values strongly biased towards 0.1, 0.3, 0.5, 0.7, and 0.9.

The values generated for a dimension are selected according to a probability distribution function, ζ. The interval $[0, 1]$ is divided into five equal partitions, and each partition is further subdivided into two equal halves. ζ is then defined piecewise in terms of one function for each of the ten subintervals: ζ_1, ζ_2, ..., ζ_{10}. The probability distribution over the first partition is then defined as $\zeta_1(x) = cx^{9.2}$, $\zeta_2(x) = c(1/5 - x)^{9.2}$, with c (a normalising constant) uniquely determined by the condition $\int_0^{0.1} \zeta_1(x) = \int_{0.1}^{0.2} \zeta_2(x) = 1/10$. The other partitions have exactly the same distribution functions, except of course that their domains are shifted appropriately.

In order to generate numbers according to $\zeta(x)$, we need a function that takes random numbers from the interval [0,1] and maps them back onto the same interval but probabilistically according to $\zeta(x)$. It can be verified that such a function θ can be obtained by taking the inverse of the indefinite integral of the probability distribution function, i.e.,

$\theta(x) = \left(\int \zeta(x) \right)^{-1}$. Figure B.1 shows both $\theta(x)$ and $\zeta(x)$ for the world structure in this example.

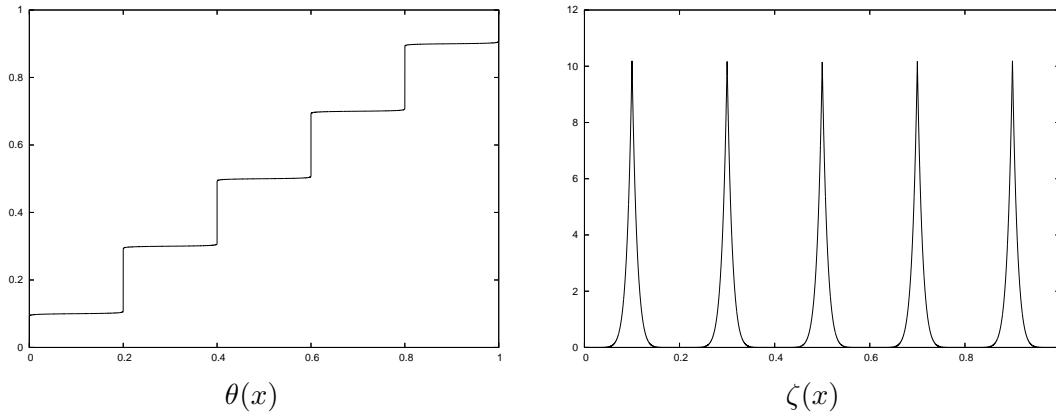
$$\theta(x) \qquad\qquad\qquad\qquad \zeta(x)$$

Figure B.1: Object generation: Feeding random numbers in $[0, 1]$ through $\theta(x)$ gives numbers according to the probability distribution $\zeta(x)$. This mechanism is used to make dimension values clumpy across all world objects.

At this point, we could take four random numbers between 0 and 1, plug them through θ, and create a new object by setting its dimension values to the resulting outputs. If we generated many objects in this way, the distribution of values in each dimension would reflect our probability distribution function ζ. We would then have a dynamic world of objects with certain dimension values being more common than others.

However, this is not quite sufficient. The above mechanism generates a value for each dimension independently, so there are no correlations between the dimensions. This would be like having a world in which having feathers and flying were completely independent properties. In such a world, agents could not infer from something having feathers that it could probably fly, or vice versa, so that no features would serve as a cue for any other features or object kinds. And, indeed, this is reflected in the fact that the parameter dim_{clump} has had no role in the above mechanism. In order to incorporate dimension dependency into the object generation algorithm, we cannot simply use four random numbers to feed through θ.

Consequently, when a new object is to be generated, we first obtain a single random number, r, on the interval $[0, 1]$. Suppose, for concreteness, that $r = 0.3$. This number is used to set up a two-part piecewise probability distribution function ϕ (called ϕ_r in Section 4.3.2 but abbreviated here), with the first piece (ϕ_1) defined on the domain $[0, 0.3]$, and the second (ϕ_2) on the domain $[0.3, 1]$. The distribution function has the

shape $\phi_1(x) = a_1 x^{4.5}$ and $\phi_2 = a_2(1-x)^{4.5}$, where a_1 and a_2 are uniquely determined by the conditions $\phi_1(0.3) = \phi_2(0.3)$ and $\int_0^1 \phi = 1$.

As we did for ζ, we obtain a function ψ that will give us random numbers in proportion to the values of our probability distribution function ϕ by taking the inverse of the integral, i.e., $\psi(x) = (\int \phi(x))^{-1}$. Figure B.2 shows both $\psi(x)$ and $\phi(x)$ for this object generation.
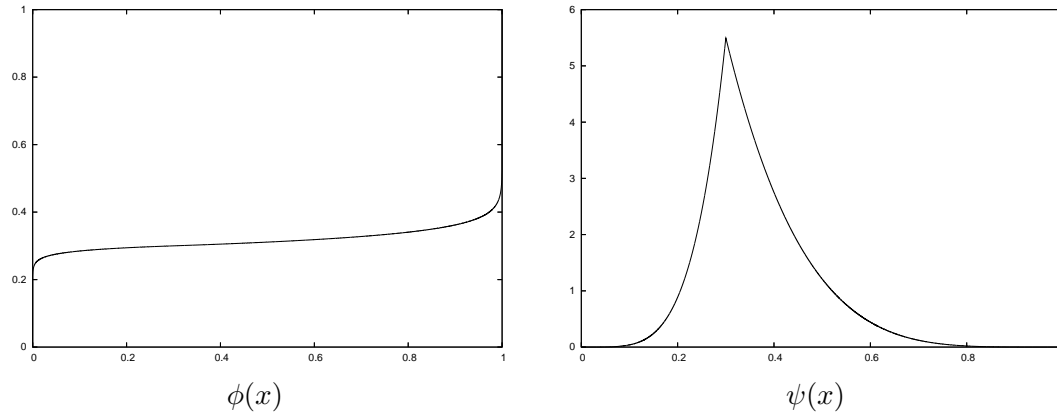

$$\phi(x) \qquad\qquad \psi(x)$$

Figure B.2: Object generation: Feeding random numbers in $[0,1]$ through $\psi(x)$ gives numbers according to the probability distribution $\phi(x)$. This mechanism is used to make the dimensions values of the object being generated correlated with each other.

Having defined ψ, the algorithm now uses it to generate a set of values, one for each of the four dimensions, by plugging four random numbers through ψ. This gives four values that are similar to each other. These values are then plugged through the dimension distribution function θ above, moving these values closer to the partition midpoints.

To exemplify, say we obtain four random numbers 0.6822, 0.0817, 0.3524, and 0.7604. Plugging each of them through $\psi(x)$ gives us 0.3936, 0.2368, 0.3098, and 0.4240. Notice that these values are all closer to 0.3 then the numbers from which they came, which reflects the correlation that exists between the dimensions. When these values are now fed through $\theta(x)$, we get the new object's dimension values: 0.3237, 0.2908, 0.3010, and 0.4869. These values are indeed closer to the partition midpoints: three of them are close to 0.3, while one is close to 0.5, showing that although there are strong correlations between the dimensions, these are still probabilistic, and there are exceptions.

APPENDIX C

Mathematical proofs

$$\forall o \in \{q \in O | q \neq o_t\},$$

$$membership_c(o) = \left(\prod_{i=1}^{N} e^{-\frac{1}{2}\left(\frac{o_i - p_i}{s_i}\right)^2} \right)^{1/N}$$

$$= \left(\prod_{i=1}^{N} e^{-\frac{1}{2}\left(\sqrt{-2log(mr)}\right)^2 \left(\frac{|o_i - p_i|}{\min_{q \in O}|q_i - p_i|}\right)^2} \right)^{1/N}$$

$$= \left(\prod_{i}^{N} e^{(log(m))\left(\frac{|o_i - p_i|}{\min_{q \in O}|q_i - p_i|}\right)^2} \right)^{1/N}$$

$$\leq \left(\prod_{i}^{N} e^{(log(m))} \right)^{1/N}$$

$$= \left(\prod_{i}^{N} (m) \right)^{1/N}$$

$$= m$$

Table C.1: Proof of bound on context object category membership: When a new category is created, the category membership of non-topic context objects is bounded above by the mem_{conxt}. o_t is the topic, O is the context, p_i is the prototype value in dimension i, s_i is the dimension sensitivity in dimension i, m is an abbreviation for mem_{conxt}, and N is an abbreviation for dim_{num}.

References

Allchin, D. (1999), 'Negative results as positive knowledge, and zeroing in on significant problems', *Marine Ecology Progress Series* **191**, 303–305.

Apostle, H. G. (1980), *Aristotle's categories and propositions (de interpretatione)*, Peripatetic Press, Grinnell, IA.

Ashby, W. R. (1947), 'Principles of self-organizing dynamic systems', *Journal of General Psychology* **37**, 125–128.

Barsalou, L. (1990), On the indistinguishability of exemplar memory and abstraction in category representation, *in* T. K. Srull & R. S. Wyer, eds, 'Advances in social cognition', Vol. 3, Lawrence Erlbaum, Hillsdale, NJ, pp. 61–88.

Barsalou, L. W. (1983), 'Ad hoc categories', *Memory and Cognition* **11**, 211–227.

Barsalou, L. W. (1985), 'Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11**, 629–654.

Barsalou, L. W. (1999), 'Perceptual symbol systems', *Behavioral and Brain Sciences* **22**, 577–660.

Belpaeme, T. (2002), Factors influencing the origins of colour categories, PhD thesis, Vrje Universiteit Brussel.

Belpaeme, T. & Bleys, J. (2005), 'Explaining universal color categories through a constrained acquisition process', *Adaptive Behavior* **13**, 293–310.

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

Bloom, P. (2000), *How children learn the meanings of words*, MIT Press, Cambridge, MA.

Bowerman, M. & Levinson, S. C., eds (2001), *Language acquisition and conceptual development*, Cambridge University Press, Cambridge.

Camaioni, L., Aureli, T., Bellagamba, F. & Fogel, A. (2003), 'A longitudinal examination of the transition to symbolic communication in the second year of life', *Infant and Child Development* **12**, 1–26.

Cangelosi, A. & Harnad, S. (2000), 'The adaptive advantage of symbolic theft over sensorimotor toil: grounding language in perceptual categories', *Evolution of Communication* **4**, 119–144.

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

Clark, E. V. (1987), The principle of contrast: a constraint on language acquisition, *in* B. MacWhinney, ed., 'Mechanisms of language acquisition', Erlbaum, London, pp. 264–293.

Clark, E. V. (2003), *First language acquisition*, Cambridge University Press, Cambridge.

Corballis, M. C. (2003), From hand to mouth: the gestural origins of language, *in* M. H. Christiansen & S. Kirby, eds, 'Language evolution', Oxford University Press, Oxford, pp. 201–218.

Croft, W. (1990), Possible verbs and the structure of events, *in* S. L. Tsonatzidis, ed., 'Meanings and prototypes', Routledge, London, pp. 48–73.

Croft, W. & Cruse, D. A. (2004), *Cognitive linguistics*, Cambridge University Press, Cambridge.

Davidson, D. (2001), *Inquiries into truth and interpretation*, 2nd edn, Oxford University Press, Oxford.

Davidson, I. (2003), The archaeological evidence of language origins: states of art, *in* M. H. Christiansen & S. Kirby, eds, 'Language evolution', Oxford University Press, Oxford, pp. 140–157.

De Beule, J. & Bergen, B. K. (2006), On the emergence of compositionality, *in* A. Cangelosi, A. D. M. Smith & K. Smith, eds, 'Proceedings of the 6th evolution of language conference', World Scientific, London, pp. 35–42.

de Boer, B. (2000), 'Self-organization in vowel systems', *Journal of Phonetics* **28**, 441–465.

de Jong, E. D. (1997), An accumulative exploration method for reinforcement learning, *in* S. Sen, ed., 'Proceedings of the AAAI'97 workshop on multiagent learning', AAAI Press, Menlo Park, California, pp. 19–24.

de Jong, E. & Vogt, P. (1998), How should a robot discriminate between objects?, *in* 'proceedings of the fifth international conference of the society for adaptive behavior', MIT Press, Cambridge, MA, pp. 86–91.

Depy, D., Fagot, J. & Vauclair, J. (1997), 'Categorisation of three-dimensional stimuli by humans and baboons: search for prototype effects', *Behavioural Processes* **39**, 299–306.

Divina, F. & Vogt, P. (forthcoming), A hybrid model for learning word-meaning mappings, *in* P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv, eds, 'Proceedings of emergence and evolution of linguistic communication III', Springer.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996), *Rethinking innateness: a connectionist perspective on development*, MIT Press, Cambridge, MA.

Fagot, J., ed. (2000), *Picture perception in animals*, Psychology Press, Philadelphia, USA.

Fodor, J. A. (1998), *Concepts: where cognitive science went wrong*, Clarendon Press, Oxford.

Garner, W. R. (1974), *The processing of information and structure*, Wiley, New York.

Gong, T., Ke, J., Minett, J. W. & Wang, W. S.-Y. (2004), A computational framework to simulate the co-evolution of language and social structure, *in* J. Pollack, M. Bedau, P. Husbands, T. Ikegami & R. A. Watson, eds, 'Artificial life IX proceedings of the ninth international conference on the simulation and synthesis of living systems', MIT Press, Cambridge, MA, pp. 214–219.

Gopnik, A. & Choi, S. (1990), 'Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development', *First Language* **10**, 199–215.

Gumperz, J. & Levinson, S. C., eds (1996), *Rethinking linguistic relativity*, Cambridge University Press, Cambridge.

Halliday, M. A. K. (1975), *Learning how to mean: exploration in the development of language*, Edward Arnold, London.

Harnad, S. (1990), 'The symbol grounding problem', *Physica D* **42**, 335–346.

Hassoun, M. (1995), *Fundamentals of artificial neural networks*, MIT Press, Cambridge, MA.

Hauser, M. D. (1996), *The evolution of communication*, MIT Press, Cambridge, MA.

Hayes, B. K. & Taplin, J. E. (1993), 'Developmental differences in the use of prototype and exemplar-specific information', *Journal of Experimental Child Psychology* **55**, 329–352.

Herrnstein, R. (1991), Levels of categorization, *in* G. Edelman, W. Gall & W. Cowan, eds, 'Signal and sense', Wiley-Liss, Somerset, NJ, pp. 385–413.

Hurford, J. (1989), 'Biological evolution of the saussurean sign as a component of the language acquisition device', *Lingua* **77**, 187–222.

Hurford, J. R. (1999), The evolution of language and languages, *in* R. Dunbar, C. Knight & C. Power, eds, 'The evolution of culture', Edinburgh University Press, Edinburgh, pp. 173–193.

Hurford, J. R. (forthcoming), 'Language in the Light of Evolution: Volume 1, Meaning'.

Jackendoff, R. (1999), 'Possible stages in the evolution of the language capacity', *Trends in Cognitive Sciences* **3**, 272–279.

Kaminski, J., Call, J. & Fischer, J. (2004), 'Word learning in a domestic dog: evidence for "fast mapping"', *Science* **304**, 1682–1683.

Kaplan, F. (2000), L'émergence d'un lexique dans une population d'agents autonomes, PhD thesis, Paris VI University.

Keil, F. C. (1992), *Concepts, kinds and cognitive development*, MIT Press, Cambridge, MA.

Kirby, S. (2000), Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners, *in* J. Knight, J. R. Hurford & M. Studdert-Kennedy, eds, 'The evolutionary emergence of language: social function and the origins of linguistic form', Cambridge University Press, Cambridge, pp. 303–323.

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

Lakoff, G. (1987), *Women, fire and dangerous things: what categories reveal about the mind*, University of Chicago Press, Chicago.

Landau, B., Smith, L. B. & Jones, S. S. (1988), 'The importance of shape in early lexical learning', *Cognitive Development* **3**, 299–321.

Lanyon, S. (2006), A saltationist approach for the evolution of human cognition and language, *in* A. Cangelosi, A. D. M. Smith & K. Smith, eds, 'Proceedings of the 6th evolution of language conference', World Scientific, London, pp. 176–183.

Lenormand, T. (2002), 'Gene flow and the limits to natural selection', *Trends in Ecology and Evolution* **17**, 183–189.

Lupyan, G. (2005), Carving nature at its joints and carving joints into nature: how labels augment category representations, *in* 'Modelling language, cognition and action: proceedings of the 9th neural computation and psychology workshop', World Scientific, Singapore.

Macnamara, J. (1972), 'The cognitive basis of language learning in infants', *Psychological Review* **79**, 1–13.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M. & Levinson, S. C. (2004), 'Can language restructure cognition? The case for space', *Trends in Cognitive Sciences* **8**, 108–114.

Mandler, J. M. (2004), 'Thought before language', *Trends in Cognitive Sciences* **8**, 508–513.

Mandler, J. M., Bauer, P. J. & McDonough, L. (1991), 'Separating the sheep from the goats: differentiating global categories', *Cognitive Psychology* **23**, 263–298.

Markman, E. M. (1989), *Categorization and naming in children: problems of induction*, MIT Press, Cambridge, MA.

Markman, E. M. & Hutchinson, J. E. (1984), 'Children's sensitivity to constraints on word meaning: taxonomic vs. thematic relations', *Cognitive Psychology* **16**, 1–27.

Medin, D. L. & Schaffer, M. M. (1978), 'Context theory of classification learning', *Psychological Review* **85**, 207–238.

Mervis, C. B. & Rosch, E. (1981), 'Categorization of natural objects', *Annual Review of Psychology* **32**, 89–115.

Murphy, G. L. (2002), *The big book of concepts*, MIT Press, Cambridge, MA.

Murphy, G. L. & Medin, D. L. (1985), 'The role of theories in conceptual coherence', *Psychological Review* **92**, 289–316.

Noble, J. (1998), The evolution of animal communication systems: questions of function examined through simulation, PhD thesis, University of Sussex.

Nosofsky, R. M. & Palmeri, T. J. (1997), 'An exemplar-based random walk model of classification learning', *Psychological Review* **104**, 266–300.

Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. (1994), 'Rule-plus-exception model of classification learning', *Psychological Review* **101**, 53–79.

Oakes, L. M., Coppage, D. J. & Dingel, A. (1997), 'By land or by sea: the role of perceptual similarity in infants' categorization of animals', *Developmental Psychology* **33**, 396–407.

Oudeyer, P.-Y. (2006), *Self-organization in the evolution of speech*, Oxford University Press, Oxford.

Pepperberg, I. M. (1999), *The Alex studies: cognitive and communicative abilities of grey parrots*, Harvard University Press, Cambridge, MA.

Pierce, C. S. (1955), Logic as semiotic: the theory of signs, *in* H. Buchler, ed., 'Philosophical writings of Pierce', Dover Publications, New York, pp. 98–119.

Pika, S. & Liebal, K. (2006), Differences and similarities between the natural gestural communication of the great apes and human children, *in* A. Cangelosi, A. D. M. Smith

& K. Smith, eds, 'Proceedings of the 6th evolution of language conference', World Scientific, London, pp. 267–274.

Pinker, S. & Bloom, P. (1990), 'Natural language and natural selection', *Behavioral and Brain Sciences* **13**, 707–784.

Quine, W. V. O. (1960), *Word and object*, MIT Press, Cambridge, MA.

Rosch, E. (1975), 'Cognitive representations of semantic categories', *Journal of Experimental Psychology: General* **104**, 192–233.

Rosch, E. (1978), Principles of categorization, *in* E. Rosch & B. B. LLoyd, eds, 'Cognition and categorization', Lawrence Erlbaum, Hillsdale, NJ, pp. 27–48.

Rosch, E. & Mervis, C. B. (1975), 'Family resemblances: studies in the internal structure of categories', *Cognitive Psychology* **7**, 573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976), 'Basic objects in natural categories', *Cognitive Psychology* **8**, 382–439.

Rosch, E., Simpson, C. & Miller, R. S. (1976), 'Structural bases of typicality effects', *Journal of Experimental Psychology: Human Perception and Performance* **2**, 491–502.

Rubenstein, A. J., Kalakanis, L. & Langlois, J. H. (1999), 'Infant preferences for attractive faces: a cognitive explanation', *Developmental Psychology* **35**, 848–855.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L. & Rumbaugh, D. M. (1993), 'Language comprehension in ape and child', *Monographs of the Society for Research in Child Development* **58**(3–4).

Savage-Rumbaugh, S., Sevcik, R. A., Brakke, K. E. & Rumbaugh, D. M. (1990), Symbols: their communicative use, comprehension, and combination by bonobos (pan paniscus), *in* L. P. Lipsitt & C. Rovee-Collier, eds, 'Advances in infancy research', Ablex Publishing Co., Norwood, NJ, pp. 221–278.

Schyns, P. G., Goldstone, R. L. & Thibaut, J. P. (1998), 'The development of features in object concepts', *Brain and Behavioral Sciences* **21**, 17–54.

Seyfarth, R. & Cheney, D. (2006), Primate social cognition and the cognitive precursors of language, *in* A. Cangelosi, A. D. M. Smith & K. Smith, eds, 'Proceedings of the 6th evolution of language conference', World Scientific, London, p. 442.

Seyfarth, R. M. & Cheney, D. L. (1980), 'The ontogeny of vervet monkey alarm calling behavior: a preliminary report', *Zeitschrift fur Teirpsychologie* .

Seyfarth, R. M. & Cheney, D. L. (1986), 'Vocal development in vervet monkeys', *Animal Behaviour* **34**, 1640–1658.

Seyfarth, R. M. & Cheney, D. L. (1990), *How monkeys see the world*, University of Chicago Press, Chicago.

Shepard, R. N. (1987), 'Toward a universal law of generalization for psychological science', *Science* **237**, 1317–1323.

Shepard, R. N. (1994), 'Perceptual-cognitive universals as reflections of the world', *Psychonomic Bulletin & Review* **1**, 2–28.

Smith, A. D. M. (2003*a*), Evolving communication through the inference of meaning, PhD thesis, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

Smith, A. D. M. (2003*b*), 'Intelligent meaning creation in a clumpy world helps communication', *Artificial Life* **9**, 559–574.

Smith, E. E. & Medin, D. L. (1981), *The psychology of conceptual processes*, Harvard University Press, Cambridge, MA.

Smith, E. E., Osherson, D. N., Rips, L. J. & Keane, M. (1988), 'Combining prototypes: a selective modification model', *Cognitive Science* **12**, 485–527.

Smith, J. D. & Minda, J. P. (1998), 'Prototypes in the mist: the early epochs of category learning', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **24**, 1411–1436.

Smith, K. (2002), 'The cultural evolution of communication in a population of neural networks', *Connection Science* **14**, 65–84.

Smith, K., Kirby, S. & Brighton, S. (2003), 'Iterative learning: a framework for the emergence of language', *Artificial Life* **9**, 371–386.

Smith, K., Smith, A. D. M., Blythe, R. A. & Vogt, P. (forthcoming), Cross-situational learning: a mathematical approach, *in* P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv, eds, 'Proceedings of emergence and evolution of linguistic communication III', Springer.

Spelke, E. S. (1994), Initial knowledge: six suggestions, *in* J. Mehler & S. Franck, eds, 'Cognition on cognition', MIT Press, Cambridge, MA, pp. 433–448.

Steels, L. (1996*a*), Emergent adaptive lexicons, *in* P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack & S. W. Wilson, eds, 'From animals to animals 4: proceedings of the 4th international conference on simulation of adaptive behaviour', MIT Press, Cambridge, MA, pp. 562–567.

Steels, L. (1996*b*), Perceptually grounded meaning creation, *in* M. Tokoro, ed., 'Proceedings of the international conference on multi-agent systems', MIT Press, Cambridge, MA, pp. 338–344.

Steels, L. (1997), Constructing and sharing perceptual distinctions, *in* M. van Someren & G. Widmer, eds, 'Proceedings of the European conference on machine learning', Springer-Verlag, Berlin, pp. 4–13.

Steels, L. (2001), 'Language games for autonomous robots', *IEEE Intelligent Systems* **16**(5), 16–22.

Steels, L. & Belpaeme, T. (2005), 'Coordinating perceptually grounded categories through language: a case study for colour', *Behavioral and Brain Sciences* **28**, 469–529.

Steels, L. & Kaplan, F. (1998), Stochasticity as a source of innovation in language games, *in* G. Adami, R. Belew, H. Kitano & C. Taylor, eds, 'Proceedings of the conference on artificial life VI', MIT Press, Cambridge, MA, pp. 368–376.

Steels, L. & Kaplan, F. (2002), Bootstrapping grounded word semantics, *in* T. Briscoe, ed., 'Linguistic evolution through language acquisition: formal and computational models', Cambridge University Press, Cambridge, pp. 53–73.

Steels, L., Kaplan, F., McIntyre, A. & Van Looveren, J. (2002), Crucial factors in the origins of word-meaning, *in* A. Wray, ed., 'The transition to language', Oxford University Press, Oxford, pp. 252–271.

Taylor, J. R. (1995), *Linguistic categorisation: prototypes in linguistic theory*, 2nd edn, Clarendon Press, Oxford.

Tincoff, R. & Jusczyk, P. W. (1999), 'Some beginnings of word comprehension in six-month-olds', *Psychological Science* **10**, 172–175.

Tomasello, M. (2000), 'First steps toward a usage-based theory of language acquisition', *Cognitive Linguistics* **11-1/2**, 61–82.

Tomasello, M. & Call, J. (1997), *Primate cognition*, Oxford University Press, Oxford.

Tomasello, M. & Call, J. (2004), 'The role of humans in the cognitive development of apes revisited', *Animal Cognition* **7**, 213–215.

Tyack, P. L. & Sayigh, L. S. (1997), Vocal learning in cetaceans, *in* C. T. Snowdon & Hausberger, eds, 'Social influences on vocal development', Cambridge University Press, Cambridge, pp. 208–233.

Vallortigara, G., Regolin, L. & Marconato, F. (2005), 'Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns', *PLoS Biology* **3**(7), e208.

Van Looveren, J. (2005), Design and performance of pre-grammatical language games, PhD thesis, Vrje Universiteit Brussel.

Vogt, P. (2000), Lexicon grounding on mobile robots, PhD thesis, Vrije Universiteit Brussel.

Vogt, P. (in press), Language evolution and robotics: issues on symbol grounding and language acquisition, *in* A. Loula, R. Gudwin & J. Quieroz, eds, 'Artificial cognition systems', IDEA Group Inc.

Vogt, P. & Coumans, H. (2003), 'Investigating social interaction strategies for bootstrapping lexicon development', *Journal of Artificial Societies and Social Simulation* **6**(1). Published online at: http://jasss.soc.surrey.ac.uk/6/1/4.html.

Watanabe, S., Sakamoto, J. & Wakita, M. (1995), 'Pigeons' discrimination of paintings by Monet and Picasso', *Journal of the Experimental Analysis of Behavior* **63**, 165–174.

Webb, A. (2005), Modelling type-denoting concepts and words in a simulation of vocabulary development, Master's thesis, University of Otago.

White, K. G., Alsop, B. & Williams, L. (1993), 'Prototype identification and categorization of incomplete figures by pigeons', *Behavioural Processes* **30**, 253–258.

Wittgenstein, L. (1953), *Philosophical investigations*, Blackwell, Oxford.

Wray, A. (1998), 'Protolanguage as a holistic system for social interaction', *Language and Communication* **18**, 47–67.