Autonomous Formation of Concepts and Communication

Edwin D. de Jong

Artificial Intelligence Laboratory

Vrije Universiteit Brussel

June 2000

Promotiecommissie:

Promotor:	Prof. dr. L. Steels, Vrije Universiteit Brussel
Voorzitter:	Prof. dr. B. Manderick, Vrije Universiteit Brussel
Secretaris:	Prof. dr. A.S. Nowé, Vrije Universiteit Brussel
Overige leden:	Dr. H. Jaeger, GMD Sankt Augustin
	Prof. dr. J.N. Kok, Universiteit Leiden
	Prof. dr. D. Vermeir, Vrije Universiteit Brussel

Cover: The front cover shows a stretched phase plot of a communication system that developed in one of the experiments described in this thesis. Each data point specifies the association strengths of two words with a meaning at some point in time. The lines connect positions at subsequent time steps. No ambiguity has arisen in the system, as two different words are never strongly associated with the same meaning; this can be seen from the lack of data points in the top right corner.

Autonomous Formation of Concepts and Communication

PROEFSCHRIFT

ter verkrijging van de graad van
doctor in de wetenschappen
aan de Vrije Universiteit Brussel,
op gezag van de Rector Magnificus Prof. dr. E. Witte,
in het openbaar te verdedigen op vrijdag 23 juni 2000 te 15:00 uur

door

Edwin Dirk DE JONG

geboren te Rhenen, Nederland

Acknowledgements

First, I want to thank my advisors. Walter Van de Velde provided me with precisely what I was looking for: an opportunity to do a Ph.D. in artificial intelligence. As an advisor he gave me much freedom to explore even if this sometimes went in directions that had no immediate relevance to the project in which he and I were involved. When Walter left, Luc Steels was willing to take over the role of advisor, for which I am very thankful. What I appreciate even more is Luc's talent to quickly see what my research is about and what paths could be promising to follow from there. In this way he has pointed me to ideas that seemed unrelated to me at first but, when I dived into them, proved to be more relevant than I had imagined.

I am also very grateful for the stimulating research environment Luc Steels has built up with the VUB AI Lab. The lab has provided a very comfortable environment for research and a pleasant place to work, not in the last place thanks to my colleagues, who have been helpful over the past years by cooperating in research, providing support, and, also very important, pleasant company; thanks to Tony Belpaeme, Karina Bergen, Andreas Birk, Bart de Boer, Sabine Geldof, Holger Kenn, Dominique Osier, Joris Van Looveren, Paul Vogt, and all other people I have gotten to know who worked at the lab for short or for long but are in other places now. Thanks also to the people at Sony CSL in Paris who, even if they were at a distance, formed part of the team: Frédéric Kaplan, Angus McIntyre, Aleix Martinez, Gert Westermann, and Jelle Zuidema.

Many thanks for their attention and careful work go to the people who gave comments and suggestions on the text: Bram Bakker, Tony Belpaeme, Andreas Birk, Gianluca Bontempi, Bart de Boer, Michiel de Jong, and Dick de Ridder.

One recurring event I always looked forward to were the Emergent(s)ia gatherings. Apart from the welcome opportunities for me to visit the Netherlands, I appreciated our lively discussions and debates, which were fun and even useful once in a while! It's amazing we kept it up for so long, and I hope you enjoyed the stimulating afternoons and enjoyable evenings as much as I did. Thanks to Bram Bakker, Martijn Brinkers, Michiel de Jong, Paul den Dulk, Dick de Ridder, Hendrik Jan Hoeve, Rens Kortmann, Sander Martens, and David Tax.

I want to thank these colleagues who enriched my working life or were helpful in some way: Hugues Bersini, Axel Cleeremans, Marco Dorigo, Dean Driebe, Yassine Faihe, Sevan Ficici, Robert French, Rob Guttman, Takashi Ikegami, Leslie Kaelbling, Joost Kok, Tomas Landelius, Bill Langdon, Han La Poutré, Will Lowe, Bruno Marchal, Eric Postma, Christopher Ronnewinkel, Alessandro Saffiotti, Cees van Kemenade, Lourens Waldorp, and Marco Wiering. Special thanks to Rich Sutton for his inspiring decidedness to solve the AI problem and for making time for a visitor in the summer of '97 who had little to offer but much to learn.

There is also life outside work, and I thank my friends in Brussels for reminding me of this when I needed it: Elisabetta, Mark, Rosita, Ana Carla, Janet, Francesco, Regina, Maria, Marie-Ange, Cecile, Steffen, Loek, and Babak. David and Monique: thanks a lot for the pleasurable hours we spent playing Piazzolla, and for putting up with my variable preparation for them during the final stages of my Ph.D. Thanks also those friends not mentioned yet: Emile, Harm Jan, Laura, Olwijn, and Signe.

There are a few people who had great influence on my decision to pursue a PhD, and without whom I would probably have followed a different path. Hans Keuken, Bram, and Michiel: thank you.

Last but certainly not least, I want to thank my family. Jurgen and Michiel, of course, but especially my parents, for everything and for arousing my curiosity and stimulating me to explore; I dedicate this to you.

List of Figures

List of Tables

Chapter 1

Introduction

Language provides humans with the ability to convey information using a set of shared symbols, called words. The intriguing question of how the relation between words and meaning is achieved is not well understood, e.g. (Kripke, 1972; Putnam, 1975, 1988; van Orman Quine, 1975; Marconi, 1997). As the title indicates, the study in this thesis concerns autonomous agents. An autonomous agent is a system, situated in an environment, that receives sensor information, selects actions, and receives evaluative feedback expressing the appropriateness of its behavior. A recent direction in artificial intelligence research studies the question of how such autonomous agents can function and adapt their behavior to the requirements posed by their environment, see e.g. (Agre & Chapman, 1987; Maes, 1989; Beer, Chiel, & Sterling, 1990; Sutton, 1990; Wilson, 1990; Beer & Gallagher, 1992; Steels, 1993; Clancey, 1997). An autonomous agent can be a software system, in which case the environment is simulated, or a hardware system, in which case the environment is the physical world. It is called autonomous because it receives no instructions from outside but has to decide on itself what actions it will produce. Setting the particular properties of human language aside, the above question may be formulated more generally as the question of how a population of autonomous agents may develop a system for communicating information about their environment. The research in this thesis addresses this question, and delivers several contributions relating to it.

Insight into this general question may not only explain aspects of existing systems of communication, but can also be put to use in the design of autonomous agents that develop communication about their environment. This is important when the possible situations that may arise in the environment of these agents are partly unknown at the time of design. In such cases, the capacity of developing a system for communication that can adapt to the requirements posed by the particular environment in which the agents find themselves is essential. The importance of being able to communicate derives from the advantage in coordinating behavior that this ability brings.

Many different forms and aspects of the research question can and have been investigated. In order to specify which aspects are of interest to the current investigation, the different choices that can been made and corresponding research will first be discussed.

1.1 Animal Communication, Evolved Communication, and Learned Communication

A reliable way to learn something about a phenomenon is to study existing examples of it. Thus, it is no surprise that much research has investigated the use of communication by animals, as will be discussed first. Most animal communication systems have genetically evolved. Much of the early simulation research also dealt with evolved communication, but the present investigation will consider learned communication.

1.1.1 Animal Communication

A large amount of research exists that studies communication in the animal world. An excellent overview of such research has been given by Marc Hauser (1997). This research belongs to the realm of biology, and is primarily concerned with the particular behavior that animals display. In other work, the focus is on the investigation of general laws and principles that apply to the phenomena studied in biology and ethology. An example is the work by Maynard Smith applying game theory to evolution, e.g. (1982). The relevance of theoretical biology to biology is that it may explain why and under what circumstances particular behaviors or phenomena may be observed. For example, if evolution takes place in a competitive environment, as is the case in natural evolution, an important question is how reliability of signals is achieved (Maynard Smith & Harper, 1995), and the related question arises of when the altruism required for sharing information evolves, see e.g. (Ackley & Littman, 1994). One explanation that has gained momentum is that a reliable signal must be costly (Zahavi, 1975, 1977), as the handicap involved in displaying such a signal can only be afforded by highly fit individuals. The feasibility of this idea has been demonstrated in theory as well as in autonomous agent simulations, see (Wheeler & de Bourcier, 1995).

The use of agent based simulation is an interesting extension of the methods of theoretical biology; another good example of such work is provided by Jason Noble's thesis (1998). Noble considers simulation research performed in the fields of artificial life and adaptive behavior as a tool for testing models in biology (or other sciences). Indeed, an important problem with artificial life research is that the number of artificial models that can be constructed and subsequently explored is infinite, and many of those possible models are of no interest whatsoever; the condition that the model should have something to say about the real world, for example about biology, is a useful criterion to address this problem. However, the criterion is needlessly restrictive, as will be discussed in section 1.10; it excludes research that tries to develop methods that are not present in the world yet but nonetheless useful. Both computer science and mathematics abound with examples of such research, and also artificial intelligence is not only interested in understanding human intelligence but at least as much in possible architectures of other intelligent systems. Thus, experiments concerning the development of communication may not only serve to improve our scientific models of human and animal communication, but also to gain knowledge about how artificial constructs may develop communication. An example of where such knowledge could be usefully applied is the development of multi-agent systems (of robots or in software) where new agents are allowed to enter the system and where the imposing the restricting condition that the agents must speak a fixed, given language is not acceptable. For these reasons, the research here is not restricted to learning about the behavior of actual animals.

1.1.2 Evolved vs. Learned Communication

A substantial amount of simulation work on communication addresses the question of how systems of communication can genetically *evolve*, that is, how can a population of agents that do not communicate initially evolve towards population that does communicate, see e.g. (Werner & Dyer, 1991; MacLennan, 1991; Di Paolo, 1997; Hurford, 1989; Noble, 1998). A different question is how an existing population can develop communication, in which case we speak of a *learned* communication system, or of *cultural* evolution. A crucial difference between evolved and learned communication is that the latter form of communication can adapt at a much smaller time scale, namely during the life of an agent, in response to requirements posed by the environment. As the ability of communication to adapt to the environment is highly desirable when agents are to coordinate their behavior in an environment unknown at the time they were designed, this research will address the question of how communication can be learned. Existing work addressing this question is discussed in sections 1.4 and 1.5.

1.2 Discrete vs. Analog Communication

Concerning the interpretation or description of signals, a distinction can be made between discrete and analog communication. Discrete communication uses a number of distinct signals, e.g. letters or words, to encode information. An example of such an encoding is given by the signals used in digital circuits, where two logical states are distinguished: zero and one. In contrast, in analog communication, signals are not distinct but have an infinite number of possible states, and small differences in the signal are meaningful. Analog signals can usually be represented by a real variable that changes as a function of time. Underneath digital systems there often is an analog system. In the example of the digital circuit, the two logical values are represented by different voltage levels. The physical states corresponding to the two states of a bit of information are only two possible values of an in principle continuous voltage level. Likewise, although the sound produced by the human voice is perceived as a sequence of distinct vowels and consonants, the underlying acoustical signal displays a variety that is far greater than the number of letters used to represent it. Although hybrid systems exist, it is useful to distinguish between discrete and analog signals.

The advantage of discrete communication over analog communication is in its robustness against noise. When a spoken A is correctly perceived as an A and subsequently reproduced, chances are that this reproduction will again be interpreted successfully, as a result of the tolerance in the production and interpretation process; since there is only a limited number of discrete values that has to be encoded in the analog signal, the analog values can be chosen such that they are not too easily confused. Likewise, the voltage levels corresponding to zero and one are sufficiently far apart not to be confused by noise or other variations under normal circumstances. Likewise, the sounds of the vowels of a language are distributed over the acoustic space such that small errors in production and perception are tolerated. In analog communication on the other hand, where even small distinctions between the possible real values of the signal may carry meaning, this fault tolerance and error correction is not present. Thus, a note at a random pitch sung by one person may, when perceived and subsequently reproduced by another person, be slightly lower or higher than the original note[1].

If analog information has to be replicated often, the small errors in the reproduction process accumulate, and information about the signal is lost. Thus, much of the power of discrete communication is in its capacity to retain information in the presence of noise. A property that often accompanies discrete communication is that the nature of the signal has no direct relation to its meaning, but is in a sense arbitrary. Such arbitrary signals are called symbols (Eco, 1976; Maynard Smith & Harper, 1995). An example of a communication system where there *is* a direct relation between a signal and its meaning, as is common in the animal world, is that of the European Blackbird. The territorial males of this species vary the intensity of their songs in relation with the amount of aggression evoked by intruders (Hauser, 1997)[2]. The present investigation will be limited to discrete communication; apart from the above considerations, an additional advantage is that the investigation of this form of communication is more straightforward given the symbolic nature of the experimental platform provided by computers.

1.3 Grammar

The research question focuses on how information about the environment may be communicated, as this particular subject of communication is most relevant to the goal of coordinating the behavior of agents by sharing information. The most basic case that can be studied is that where a message of a single symbol, called *word* or *signal* in this thesis, yields information about aspects of the environment, and this case will be studied here. A more sophisticated case arises when multiple words may be used in a mes-

[1] The example assumes that no musical context is present; if it were, the pitch would not be random, but would be interpreted as an approximation to one of the tones in the musical scale, in which case the system would be properly modeled as discrete.

[2] Hauser presents the European Blackbird system as an example of graded communication, which he notes Marler defines as communication where signal variation is continuous, lacking clear acoustic boundaries for demarcating one signal type from another. However, in the description of this system, three categories of signals are distinguished (low-intensity, high-intensity and scrambled song). The non-arbitrariness of the Blackbird example is perhaps a clearer aspect than its gradedness.

sage describing the situation, especially when the order of the words affects the meaning of the message. In the latter case, information can be encoded in a more efficient way, and the issue of grammar begins to play a role, see e.g. /citebatali:98,kirby:99,steels:1998a,.

1.4 Concepts and their Formation

The aforementioned aspects of the environment about which communication yields information are called *features*. Features are used in pattern recognition (Duda & Hart, 1973) to denote information that can be used for classification, and as a concept may be viewed as a pattern, concepts are often seen as consisting of features (Bala, De Jong, Huang, Vafaie, & Wechsler, 1996; Schyns, Goldstone, & Thibaut, 1998). Typical examples of features used in pattern recognition systems are color features (*redness*, or the quality of being *green*), texture features (*rough*, *smooth*), orientation features, etc., but in principle there is no limit to the complexity of the patterns a feature detects.

The meaning of a word is determined by the features with which it is associated. The complex of features that determines to what extent a word is applicable will be called a *concept* or *meaning*. Since concept formation is performed by individual agents, concepts are entities internal to the agent. The publicly observable aspects of the environment to which they refer are called *referents*. This section considers the role of concepts in the development of communication.

1.4.1 Universal Concepts?

The notion of a concept appears to reduce the problem of establishing a relation between words and sets of features to that of establishing relations between words and concepts. However, this raises the question of where concepts come from. A possible answer to this question is that agents have universal concepts at their disposal. This means that the concepts a situation gives rise to in different agents are the same, in the sense that they are evoked in the same situations. In this case, the only task faced by agents is to establish a relation between shared concepts and words. This explanation will be called the *universal concepts hypothesis*. Although the idea that concepts are universal may appear difficult to maintain, it does exist, for instance as part of the idea that language is a timeless, unchangeable, objective structure, see (Katz, 1981).

To make the discussion easier to envisage, it will be useful to consider the case of human language. The universal concepts hypothesis appears to provide an accurate account of the lexicons used in human languages. Dictionaries give a description for each word, and although a word may have several meanings and several words may share a meaning, an underlying assumption behind the very existence of a dictionary is that the meaning of a word is independent of the person using it. However, the hypothesis also raises serious questions. A particularly striking question is how the hypothesized identical concepts end up in the heads of people. A purely nativist explanation must be ruled out by the immediate observation that the meaning of at least certain words can only be learned during a person's life, for the very reason that many new words did not exist when the person was born (e.g. words for new technological inventions). But even slightly more sophisticated models that assume people simply need to learn a number of truth conditions in order to acquire *the* meaning of a concept implicitly make the very strong assumption that there is a single concept that is shared by every speaker of a language. The immediate observation that different people, when asked to define a word, come up with different features as defining properties raises doubts concerning the realism of this assumption.

In (Fodor, 1998), several conditions for a theory of concepts are proposed. The conditions result from the viewpoint of the Representational Theory of Mind, which assumes that laws can be formulated specifying causal relations between beliefs, desires, and intentions. Among these is the condition that concepts should be *public*, and shared by people. The research that is presented here and the previously mentioned work by Steels explores another possibility. Whereas Fodor states that there must be (infinitely) many primitive concepts, another possibility is that concepts can be *formed*, as will be discussed in the following sections.

Research into the development of communication can be grouped according to whether it assumes that concepts are already present or not. In the first case, the problem to be addressed is how a relation between the existing concepts of agents and public signals or words can come about. An example is the work of Oliphant (1997), where each agent possesses a number of concepts that abstractly represent situations, and perfect coherence regarding these concepts is assumed in the sense that all agents will always observe the same situation. Thus, the concepts are public, in that the current situation is always known to each agent, or, in other words, a one-to-one correspondence between referents and concepts is given. Oliphant's thesis presents experiments on both evolved and learned communication. The latter experiments concern Bayesian learners that attempt to communicate optimally with the current population, biased by a normalization procedure that assumes a one-to-one relation between concepts and words.

The given presence of concepts does not imply that those concepts are public however. MacLennan, in one of the first papers presenting simulations concerning the evolution of communication (1991), has considered experiments where each agent has access to a private part of the state of the environment. The evolution of communication allows agents to share this private information by establishing a correspondence between concepts and signals. The interesting question of how these private concepts arise is not addressed.

Research on communication that assumes the concepts are already present leaves a substantial part of the question of how communication may develop unaddressed, viz. the question of how these concepts are acquired. We will now turn to this question.

1.4.2 The Acquisition of Concepts

It has been seen that the universal concepts hypothesis incorporates a rather strong assumption and has consequences that are difficult to explain. In this thesis, an alternative explanation for the coherent relation between situations and words is explored. While the idea of universal concepts is difficult to defend, the idea that concepts are constructed by each individual based on interaction with the world and the people one meets raises a new question that is perhaps as difficult: how can the relations between words and the concepts formed by individuals develop such that these individuals use the same words in approximately the same circumstances? It is clear that such a situation can only come about if there is a strong coherence between the concepts people maintain.

Two main sources for this coherence can be identified. First, people may construct concepts because they are *useful*, independent of communication. For an excellent account of the present-day view in artificial intelligence that individuals develop internal representations in the process of coordinating their interaction with the environment, see e.g. (Clancey, 1997). To give an example, a fisherman may come to develop accurate distinctions between cloud patterns that are correlated with the advent of rain and cloud patterns that are not. Likewise, lifelong experience may allow a music teacher to develop a concept of a promising young player, determined by all kinds of different features of which the teacher herself may not even be aware, that indicates whether a five year old pupil is apt to become a concert pianist or not. Although the association with the word 'promising' is determined by the fact that the teacher has heard this word being used for promising children or events, the structure of the concept is primarily determined by the personal experiences of the teacher. In the case of the fisherman, this is even more clear, as he probably has not even associated the concept of a rainy cloud pattern with a word. If there would be a need for such a word, for instance because fishermen want to refer to the cloud patterns on such a regular basis that they tire of using long sequences such as 'the cloud pattern that is almost always followed by rain', they might coin a new word for it. If the word catches on, it may spread, first through the local population of fishermen, but perhaps further. To give a final example illustrating the power of the principle, the fact that high school children can be taught the concept of force as developed by Newton allows Newtonian mechanics to spread to much larger parts of the population than would be the case if teachers would simply wait for each student to invent the concept him- or herself; the latter would only work in a world of geniuses.

This brings us to the second principle that may induce coherence in the concepts people maintain. If a new fisherman overhears two colleagues using the new word for rain inducing cloud patterns, he will not know its meaning initially. However, after a while, perhaps merely by observing the context in which the word is used, he may come to develop an idea of what his colleagues mean with the word. The development of this idea allows him in turn to distinguish between different cloud patterns, and hence the new concept has spread. The latter principle is especially powerful in that it can cause concepts to spread over populations of language users, but will not be investigated here. It involves concept learning, which is discussed below, but, as will be seen, concept learning alone cannot account for the principle.

1.4.3 Concept Learning

Concept learning as it is studied in the machine learning community refers to the acquisition of a mapping from an input space to an output space, where this acquisition is based on known combinations of input examples and output examples. Thus, it is a form of supervised or instructive learning. From a cognitive perspective, this raises the question of where the concepts that are thus learned originally come from; if a set of labeled examples is required to learn a concept, this presupposes the existence of the concept, and therefore mechanisms of this type can not explain the initial *formation* of concepts. Examples of concept learning include decision tree learning, see e.g. (Quinlan, 1990; Michalski, Carbonell, & Mitchell, 1986), where learning consists of building a tree with distinctions in its node that divide the input space into subsets.

Aude Billard (1997, 1999) has investigated the development of a mapping between words and situations in the environment of a mobile robot. The concepts in this experiment are learned, and consist of sensory motor patterns. The relation between situations and words is determined by a teacher robot. The learner robot learns this relation by associating words received from the teacher with its own sensory motor patterns using a variant of the Willshaw net (Willshaw, Buneman, & Longuet-Higgins, 1969) that employs Hebbian learning. Other work where a learner robot learns a relation between words and concepts from a teacher robot is that by Yanco and Stein (1993). There, the following learner robot receives the teacher's signal but no sensor input, and has to learn a mapping between these signals and its actions based on reinforcement. The communicative behavior of the teacher is not fixed however, but also learned based on reinforcement learning, where success is determined by whether the follower robot executes the appropriate action based on the signal. Thus, the work also addresses the question of how the teacher learns its language. The teacher is an example of a central control force determining the language that develops, and does not address the question of how a distributed population of agents may develop a communication system.

An important question in the development of a communication system however, is which of the words will spread through the population and which will not. This question is not addressed by concept learning itself, but depends on other factors that determine whether a word is likely to be learned and produced by agents. For this reason, the presence of a teacher that decides how words are used is required. This is characteristic for concept learning, and distinguishes it from concept *formation*, discussed in the following section.

1.4.4 Concept Formation and Bias

It has become clear that the formation of concepts requires criteria determining which concepts an agent should form; without such criteria, an agent has no means to select a finite number of concepts from the endless amount of possible concepts it could construct. It is important to realize the hardness of this problem. The more information an agent receives, the more concepts it could form in principle. Let us consider a simple sim-

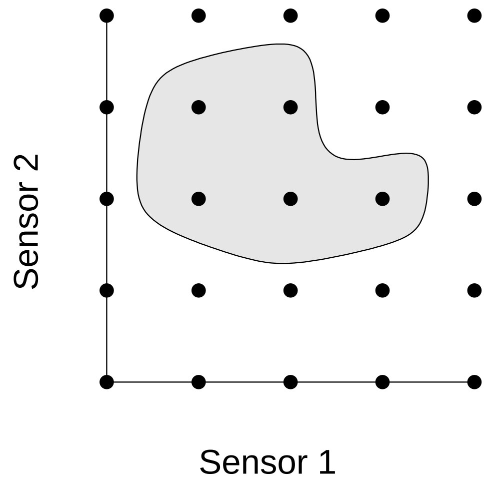


Figure 1.1: Example of a concept in a (discrete) sensor space. Each point in a sensor space represents a combination of sensor values, one value for each sensor. The concept distinguishes five such combinations from the other possible combinations.

ulated agent possessing only two sensors that determine its view on the world, where both sensors yield values between one and five. Then the number of possible concepts is already $2^{25} = 33,554,432$. This number is found by considering all single sensor values of each sensor, all combinations of two sensor values (e.g. sensor one equals 1 or 5, or both sensors equal 3), all combinations of three, and so on. The two sensors can be depicted in a diagram by using an axis for each sensor, see figure 1.1 for an example. The sensor information at one point in time can then be represented by a point of which the two coordinates equal the two sensor values. This two-dimensional space is called the sensor space, and a concept distinguishes one or more points in this space from the rest of the space. However, this is only the beginning. We have so far only considered the current sensor values; if a concept may also capture information about the sensor information at previous time steps, extra axes are required to represent those previous sensor values. With each axis, or dimension, that is added, the number of possible concepts grows extremely fast[3]. By adding not only the last, but all previous sensor information, and in addition the actions taken

[3]As with each added axis the exponent is multiplied by the number of values along the new dimension, the exponent grows exponentially, and thus the number of concepts grows super-exponentially.

by the agent and the subsequent rewards, the space of possible interaction histories is obtained. It is clear that this complete space corresponds to such a vast number of concepts that it becomes infeasible to consider every possible single concept. Moreover, we have so far only considered sensors with discrete values; as soon as continuous variables are considered, the number of possible concepts becomes infinite.

These considerations make clear that strong restrictions have to be posed on the possible concepts that will be considered by an agent. Two forms of restrictions can be distinguished.

First, restrictions may be imposed on the shape of the concepts, i.e. on the set of concepts that can be represented. Such a restriction cannot be made without loss of generality; if the agent may encounter any theoretically possible environment, there is no principled reason to prefer a simple concept, e.g. one that distinguishes between high values and low values of a sensor, over a concept distinguishing a completely random set of data points from all other possible sets of data points. This point relates to the No Free Lunch theorem (Wolpert & Macready, 1995), which basically states that if the set of possible problems one may encounter is unrestricted, no search method can do better than random search. When constructing a concept is viewed as searching in the space of possible concepts, this type of restriction could be called a *representational bias* in analogy with (Gordon & desJardins, 1995), as it limits the solutions that may be represented.

The second type of restriction on the concepts that are constructed does not exclude any concepts, but favors certain concepts over others in that some concepts occur earlier in the search than other concepts, i.e. the order of the search is changed. Since the time available for the search for concepts is not unlimited, this in effect restricts the concepts that can be found. In the framework of (Gordon & desJardins, 1995), this form of restriction would be called a *procedural bias*, as it influences the order of the search. In the presence of procedural bias, certain concepts are highly unlikely to be found, as only abnormal data or a very long search for concepts would produce them.

1.4.5 Criteria for the Formation of Concepts

This section discusses various criteria on which the formation of concepts may be based. In discussing such criteria, it is useful to consider the form of feedback they assume. The feedback of a learning system can be:

- *instructive*, in which case the learner receives the output it *should* have given or, equivalently, the error vector between the possibly multidimensional output the agent gave and some desired answer. This form of learning is called supervised learning.

- *evaluative*, in which case the learner receives a numerical evaluation of its behavior, for example the magnitude of the error between its output or action and the optimal action. This form of learning is called reinforcement learning.

- absent, in which case the learning system learns properties of the data, e.g. its density distribution. This form of learning is called unsupervised learning.

As supervised feedback assumes the presence of a teacher providing the feedback, using this information as a criterion would result in concept learning as opposed to concept formation. As will be argued in chapter 2, evaluative feedback is well suited to be used as a guiding principle for concept formation by autonomous agents. The assumption that autonomous agents have such feedback at their disposal does not imply that it is *provided* by the environment from some mysterious source; rather, evaluative feedback on behavior may be viewed as a judgement by the agent itself of its current situation. The ability to judge one's own behavior is a powerful and flexible mechanism, of which human sensation of pleasure and pain may be viewed as instantiations.

The question of how evaluative feedback can be used for concept formation was considered by Wrobel (1991). There, failure of the learning system to predict the reward following an action results in the addition of perceptual distinctions. The resulting sensor intervals were used as symbols in a concept learning tree. Although the idea is powerful, the model had several limitations. It requires binary rewards, but moreover requires direct rewards, and thus it is not suitable for environments that have multiple states[4], which excludes all interesting environments. These issues *are* addressed by Andrew McCallum's U-Tree method, which builds a tree of distinctions regarding current and past information available to the agent in order to estimate the value of the different actions an agent may take.

[4]As soon as actions influence state transitions, the value of an action is not only determined by its direct reinforcement, but also by its effect on the state due to possible changes in the value (discounted sum of expected future rewards) caused by the state transition.

Apart from concept formation methods based on evaluative feedback, methods have also been proposed that use no feedback at all; these are called unsupervised methods. An overview of several such methods is given in (Gennari, Langley, & Fisher, 1989). For examples of unsupervised concept formation methods, including Fisher's CobWeb system, Feigenbaum's EPAM system, and Cheeseman's Autoclass system, see (Shavlik & Dietterich, 1990).

In methods that use evaluative feedback, this feedback can be used to guide concept formation such that it assists the production of useful behavior. In methods based on unsupervised learning, no such feedback is present, and only general considerations can guide concept formation process. These include error minimization, entropy maximization, feature mapping, clustering, and density estimation (Fritzke, 1997). The maps that are found in mammalian cortex, e.g. retinotopic maps, tonotopic maps, and Kohonen's self-organizing map (SOM) (1995), are good examples of feature maps. Their existence suggests that the principles guiding unsupervised learning may have biological relevance.

A general principle that underlies concept formation is the aim of discriminating between various entities (e.g. objects, situations). This principle is investigated in the *discrimination game*, introduced in (Steels, 1996b). In a discrimination game, an agent receives information about the objects in its environment, called the *context*, and randomly selects one object to be the *topic*. The information about an object consists of a value for each of a number of *sensory channels*. It then determines whether it can distinguish each object from the topic using its current segmentation of the sensory channels. If not, the segmentation has to be refined. If it can, the topic gives rise to a *meaning*, a feature set that distinguishes the topic from the other objects in the context.

In the standard typology of learning systems, the discrimination game would be considered a form of unsupervised learning; the agent only receives sensor information, but selects no action. However, the criterion that guides the concept formation *is*, as in reinforcement learning problems, given by the agent's performance on a task. This task is to discriminate between the objects in the context. As the agent possesses all necessary information to determine whether it can perform its task (discrimination), the agent *itself* can determine whether it is successful or not. Apart from its value as an approach to concept formation, the discrimination game is a good demonstration of the point that evaluative feedback may well be available to an agent even though it is not 'sent' by the environment or by a teacher.

Unsupervised learning has also been applied to data resulting from the interaction of a robot with its environment, e.g. (Oates, Schmill, & Cohen, 1999). Dynamic time warping (Sankoff & Kruskal, 1983) is used to measure similarity between time series. The similarity measure is used in a subsequent clustering method, and is used to find distinctions between time series. Although the unsupervised approach excludes the use of goal directed criteria for these distinctions, the idea of looking for patterns in the interaction history is a notable aspect. Oates and Cohen also make a connection to language learning. In (Oates, Eyler-Walker, & Cohen, 1999), human subjects were asked to generate unrestricted natural language utterances to describe the behavior of a robot. Word clustering was then used to learn the relation between these utterances and the behavior of the robot.

A limitation of many concept formation methods, including the one investigated here, is that they yield crisp, rigid representations, whereas the production of adequate behavior in the real world often requires fluidity. In (MacDorman, 2000), partition nets (MacDorman, 1999) are used to learn the effect of the motor commands of a robot in sensor space. Although crisp distinctions are used to build a KD-tree, a method using blends of the experiences stored in the nodes or, when more accurate, a neural network approximation, yields smooth variable resolution function approximation that also learns quickly, both in terms of time and experience required.

Another approach, also based on the predictive patterns in time series data, is Jun Tani's Recurrent Neural Network (RNN) approach (1998). By letting RNN modules compete to predict the stream of sensory motor data and using a gating mechanism to allow the most appropriate network to learn, the modules become specialized in particular forms of behavior.

An interesting mechanism for concept formation, inspired by the constructivist theories of Piaget, has been given by Gary Drescher (Drescher, 1991). The computational model he describes, called the *schema mechanism*, constructs schemas, actions, and items to represent the state of the world, discover regularities in the world, and organize action sequences in pursuit of the goals of the agent. A schema represents a prediction of what the result will be when a particular action is taken in a particular context, where the context and the result are specified by sets of conditions. Drescher's approach employs increased predictability of the environment as a criterion for concepts.

In a number of the above methods (Drescher, 1991; Wrobel, 1991; McCallum, 1996; Tani & Nolfi, 1998; Oates et al., 1999; MacDorman, 2000), concept formation involves searching for patterns in the history of the interaction between the agent and the environment, and/or the predictive aspect of concepts is used as a criterion. These two ideas, especially when used in combination, lead to powerful methods for concept formation, even though the methods are not seen as concept formation method in all of the above cases. The combination of the two aspects mentioned yields methods that search for patterns in the interaction history that have a predictive aspect. Since this is a rather general principle, it may be useful to have a term for such concepts. They will be called *situation concepts* here, as the meaning of the word *situation* captures both the aspect of relating to the current state of the environment and the aspect of being relevant with respect to the future course of development of that state, possibly conditional on the actions the agent will select.

1.4.6 Situation Concepts

At each point in time, the total information an agent may possess about the state of its environment is called the *interaction history.* The interaction history contains the information the agent received through its sensors, but also incorporates the actions the agent itself selected. Furthermore, it will be assumed that the agent receives feedback evaluating the quality or appropriateness of its actions. As the interaction history represents the complete information an agent has about its environment, the features it uses to distinguish between different possible states of the environment can be defined over the interaction history, i.e. no information other than the interaction history is required to determine the degree to which a feature is present. Another, equivalent way to think about features is that they restrict the possible interaction histories.

Any system that uses words to transmit information about its environment may be described as representing a relation between features in the interaction history and words. This can be seen as follows. If a word conveys information about the state of the environment, it can not be equally applicable to all possible interaction histories. According to the view that features restrict the possible interaction histories then, the word can be viewed as corresponding to a set of features. Regardless of how these feature are implemented, they are characterized by the restrictions they pose on the possible interaction histories. The central question can then be formulated as the question of how relations between features and words can come about in agents such that the production of a word by one agent restricts, and hence yields information about, the possible states of the environment for an agent that receives this word.

To capture information about the environment, concepts may be formed that are made up of features defined over the interaction history of an agent with its environment. However, not all information about the environment is equally relevant to an agent that needs to choose actions. Of particular relevance are those concepts that predict aspects of the future behavior of the environment. Such concepts will be called situation concepts. The predictive aspect may simply concern information about future states of the environment, whatever actions the agent chooses. For example, the presence of low flying swallows may indicate the advent of a rain shower. A slightly more sophisticated form of situation concepts considers how the environment will react to the actions of an agent. For example, knowledge that the floor is slippery influences the effect that different forms of walking may have, and may be reason to place one's steps more carefully than usual. Or, for another example, knowledge that the fire extinguisher is empty may cause one to adapt one's expectation of the effects of using it in case of fire. In all examples, the presence of the concept, i.e. its relevance to the current state of the environment, may be known to some agent from observations in its interaction history (e.g. trying to walk on the slippery floor, or having observed someone attempting to use the empty fire extinguisher), and capture information about future states of the environment, or about the effect of actions on these states.

The definition of situation concepts is very general, and can be instantiated in a large number of ways. The examples illustrate how complex the concepts that have been defined may be; distinguishing visual scenes containing a book from all book-less scenes is, although a relatively easy task for most humans, already a highly nontrivial task for a computer vision system. They are mentioned to illustrate the generality of the description given so far, as the actual concepts that will figure in the experiments are much more primitive. In order to determine whether the approach is feasible, one of the most basic form of situation concepts will be investigated. If the findings of these experiments are positive, the road will have been paved to consider more elaborate situation concepts in the investigation of communication.

Situation concepts are interesting because of their special relevance to the problem of coordinating behavior with the environment. As the current state of the environment determines its future behavior, including the actions an agent may take and the way it responds to them, knowledge of the situation is crucial in determining what actions to take. The research investigates the simplest case of communicating about the situation, namely the case where information about the situation is expressed using a single word.

Biases of the Adaptive Subspace Method for Situation Concept Formation

In the particular method for the construction of situation concepts that will be considered here, called the Adaptive Subspace Method, both previously mentioned types of bias will be applied. Representational bias is used by limiting concepts to solid regions in sensor space. Thus, the features constituting a concept directly correspond to the current values of the sensors, and a concept defines an unbroken range of possible values for each of the sensors[5]. Furthermore, the concepts that are formed do not overlap.

The procedural bias is a more interesting case. It is determined by the criterion that guides the choice of these regions in sensor space. This choice is based on the central idea that communication must provide useful information about the environment. As mentioned, the behavior of an agent is evaluated by the rewards it receives after its actions. Therefore, knowledge about the environment is useful if it allows an agent to estimate in advance what rewards an action will return. In general, the action not only determines the evaluative feedback that follows it, but may also influence the next state of the environment and, by consequence, affect its capacity of obtaining rewards in the future. As will be seen, the combined influence of these two effects of an action on future rewards can be estimated at once by considering the *value* of the action as opposed to merely the reward following it. The proposed method uses the ability to estimate the value of a possible action as a criterion for the quality of a concept.

A further procedural bias is caused by the way concepts are formed. This happens by means of an increasing number of *distinctions*. Initially, no distinctions are made between interaction histories. Then, gradually, distinctions are introduced based on to what extent the action value estimation differ in different regions of the interaction history space. This approach brings with it a procedural bias that favors concepts that can be constructed with a small number of distinctions, as constructing concepts requiring many distinctions necessarily involves producing the other kind of concepts first. Whether this form of procedural bias is helpful depends, as

[5]Note that this use of the term 'subspace' (i.e. as hyperrectangles) differs from the perhaps more common way Oja (1983) uses the term, namely as a lower-dimensional manifold embedded within the space.

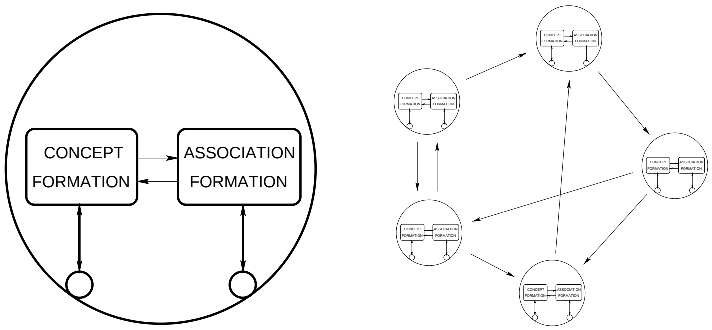


Figure 1.2: The coupled dynamics involved in the development of communication, as it is also investigated in the work of Steels. To the left, an individual agent is shown. The concept formation process forms concepts based on experience with the environment. The association formation process forms associations between these concepts and words. As the diagram shows, the concept formation process can also influence or be influenced by the association formation process.

with all biases, on the problem. However, it has an important advantage. The distinctions it makes only break up a single existing concept into two refined concepts. As a result, it produces a very small number of concepts, even if the dimensionality of the interaction history space is high; with each new distinction, only a single new concept is introduced. A small number of concepts is desirable because it reduces the amount of experience that is required to associate properties with a concept, in particular the values of the actions in the situation it represents, and the word with which the concept will be indicated in communication.

1.5 Associating Concepts with Words: Language as a Dynamical System

The view that concepts are formed in the heads of individual agents complicates the already difficult question of how a coherent relation between concepts and words may come about. The experiments in this thesis address this question by showing one of the simplest cases where this is achieved. Part of the contribution of these experiments is therefore in providing evidence for the viability of a particular view on the development of communication. This view is that a population of language users, consisting of multiple agents that interact by communicating, can be viewed as a *dy-*

namical system. A diagram depicting the interactions in such a system is given in figure 1.2. The words an agent uses depend on the concepts it has constructed (the right arrow in the diagram), but communication can also influence concept formation (the left arrow). Whereas the external (vertical) arrows of the concept formation process represent interaction with the sensors and actuators of the agent itself, the association formation process interacts with other agents by means of communication. Thus, the dynamics of concept formation and association formation are coupled. The interactions between different agents are shown to the right. Together, the coupled interactions determine how communication develops.

Mathematically, a dynamical system is a system that changes over time, see e.g. (Strogatz, 1994). Such systems can be used as models for all sorts of systems in our physical world. Particularly interesting forms of dynamical systems are those which consist of many interacting elements. The number of interactions between the elements in such a system may be so large that considering them individually becomes infeasible. Surprisingly, this does not imply that such systems cannot display interesting observable behavior. The development of ordered patterns in such systems results from the many interactions between the elements of the system, not by some central organizing force that imposes order. Although the above is not a strict definition, the principle occurs in such a variety of systems and is often so clear that it is useful to have a term for it, namely *self-organization.* A good example of self-organizatoin is provided by ant colonies, see e.g. (Bonabeau, Theraulaz, Deneubourg, Aron, & Camazine, 1997). When ants walk, they leave a pheromone trail. Since ants are attracted by this substance, they are more likely to walk over paths that have been used by other ants. As this in turn causes more pheromone to be placed on those paths, the result is that gradually the ants start using the same paths. The process of path formation has several interesting aspects, such as a preference for short paths and paths that lead to food, but these need not be discussed for the basic principle to be clear. The lesson is that although no central control force, i.e. no single ant, decides on the choice of the paths, the result of the many interactions between the ants and their environment (placing and following pheromone) is that the ants follow the same paths instead of randomly crossing the area, and thus these local interactions lead to a global form of order.

Globally, there are two ways of taking a dynamical systems perspective on language and communication. First, the system that learns a language can be viewed as a dynamical system. Following this approach, Elman

(1995) used Simple Recurrent Networks that had to learn to predict sequences. The dynamic aspect of this task is a result of the necessity for the network to build up a form of internal state, based on the parts of the sequence observed so far, in order to predict its continuation. Furthermore, Pollack introduced the dynamical recognizer (1991), a variant of which induced languages consistent with all Tomita (1982) data sets. Successful induction corresponded to a phase transition in the weights of the learning network, and an analysis is provided in (Blair & Pollack, 1997).

The other approach is to view language *itself* as a dynamical system. In this view, the interactions between the members of a population determine the language they develop as whole. This approach has been advocated by Steels, see (1997a) for an overview of work from this perspective and (Steels, 1996c, 1996d, 1996a; De Boer, 1999) for examples of this approach. As we are interested in the development of communication in a population of agents, as opposed to learning an existing language, this approach will be adopted here.

In the view of communication development as a dynamical system, the order that needs to be explained is how agents may come to use the same words in the same situations. The interactions between the agents are the production and interpretation of words. In contrast with most physical dynamical systems, simulation experiments allow an investigator to have complete information about what determines these interactions. In the view of communication development as the formation of associations between concepts and words, these are the *variables* that govern the production and interpretation behavior of communicating agents. Therefore, an exceptional opportunity presents itself, in that the behavior of this dynamical system may be investigated mathematically. In the thesis, a (deterministic) communication system will be demonstrated mathematically to have *attractors* that correspond to perfect communication.

Approaches to learned communication, i.e. where agents adapt associations between meanings and words during their lifetime, can be classified bas on various criteria. First, there is the question of whether concepts determine the language that may be developed, or whether communication also influences the conceptual systems of agents. These two forms of interactions respectively correspond to the right and left arrows in figure 1.2. The current investigation has been limited to the right arrow, which greatly facilitates the analysis of the concept formation processes. On the other hand, it implies that the powerful idea of concepts spreading through the population can not be examined. Work that does address this question has been done by Steels in an experimental setup called the Naming Game, introduced in (Steels, 1996c).

The naming game investigates how concepts, formed e.g. in a discrimination game, can be associated with words such that different agents refer to different referents with the same word. In the naming game, the speaker produces the word that is most strongly associated with the meaning this agent has found for the topic. This word is received by the hearer, which interprets the word to yield a meaning, and uses this meaning to identify one of the objects in its context. If this results in the same object, the speaker has successfully communicated the intended object to the hearer, and the game is a success. This success information is then used to adapt the associations between the meaning and word of both the speaker and the hearer. If it is unsuccessful, the underlying discrimination module is triggered to form new distinctions, and thus communication can influence the concept formation process.

A second distinction concerns the availability of feedback on communication. Oliphant (1997) describes 'observational learning', where no feedback on communication is given. The motivation is illustrated by this quote from (Pulliam & Dunford, 1980): "the obvious problem with trial-and-error learning is error". However, a possible role of evaluative feedback could be to improve upon information available from the observation of communicative behavior by taking into account the value of the choices involved in communication. If the agent has a way of determining which ways of using communication (its production or perception) are successful, the extra information thus provided can be used to refine knowledge obtained by pure observation.

In several approaches to communication, e.g. in (Yanco & Stein, 1993) and in the naming game mentioned above, success is communicated as nonverbal communication and used by the agents to adapt their communicative behavior. Similar to the way agents can determine their own success in the discrimination game, this is also possible when associating concepts with words. In the experiments that will be described, words represent situations. Since a situation must have a predictive aspect, the agent can test whether the subsequent development of the environmental state conforms to its expectation. The outcome of this test indicates whether the determination of the situation was correct. If this determination was based on communication, the agent can determine its own success regarding the interpretation of words. This setup is another demonstration of the idea that evaluative feedback is not necessarily provided by an external source, which would raise the questions about the plausibility of this feedback being available, but can be determined by the agent itself. A related question, which will also be addressed, is whether evaluative feedback is required for the development of communication in the experimental setup that is used.

A final distinction, although many more distinctions are possible, concerns the information received by agents other than feedback on communication. In order for communication as it is viewed here to develop, this should minimally include the production of words or signals by other agents. In the observational learning paradigm used by Oliphant (1997), agents additionally receive information about the *interpretation* behavior of other agents. The rationale behind this appears to be that communication corresponds to situations, situations correspond to actions, and since actions can be observed, the corresponding situation as interpreted by the receiver can be inferred[6].

One of the commitments of the current research is that meanings are not passed from one agent to another. Thus, even if there would be a one to one correspondence between the actions and internal meanings of other agents, this fact could not be used to observe the interpretation behavior of other agents.

1.6 Measuring the Development of Concepts and Communication

In any experimental investigation of the development of concepts or communication, it is valuable for the experimenter to have information about the state of this development. The thesis provides measures expressing this information for both the development of concepts and the development of communication. In order to determine whether a concept or word represents useful information about the environment, one has to know what information it is that needs to be represented. The subjects of communication that are to be distinguished, here situations in the environment, are the *referents*. This use of the term referent differs slightly from its use in linguistics, where referents are the entities and states of affairs designated by linguistic expressions in particular utterances (Lambrecht, 1994); the difference is that whereas the linguistic notion of a referent is that to which language refers, here the term is used to denote the entities that *should* be referred to by the words in a system of communication.

[6]The interpretation in terms of situations and actions is purely abstract; no concept formation or action selection is involved in the experiments

The measures presented in thesis assume the presence of a set of referents. They can be used whenever the number of occasions on which a referent and meaning or a referent and word occur together, called the co-occurrence frequencies, can be counted. Using the co-occurrences, the conditional probability matrices on which the measures are based can be calculated.

For both the relation between referents and meanings and that between referents and words, two aspects are desirable. First, a meaning or word should capture information. In the case of words, let us assume that the receiver of a word initially has no knowledge of the referent intended by the speaker. Then the word captures information if, after receiving the word, the receiver has gained information about which referent the speaker intended. Ideally, the receiver's initial uncertainty as to the nature of the referent would be taken away completely by reception of the word.

The uncertainty about the identity of an entity can be measured in a principled way using *entropy*. Let us consider a variable with a limited number of outcomes, each of which is associated with a probability. Whereas the probability of the most likely outcome P_m determines the probability of guessing wrong $(1-P_m)$, it does not capture any information about the way the probabilities over the less likely outcomes are distributed. Shannon's entropy measure does capture this information, and provides a principled measure of uncertainty, see (Shannon, 1948).

The degree to which a word has the capacity to take away uncertainty about the referent can be measured as the *decrease of uncertainty*, relative to the maximum possible decrease of uncertainty, which equals the initial uncertainty. The measure in which this results is called *specificity*, and indicates to what extent a word identifies a referent. By combining the specificity of all words into a single measure, the specificity of an agent is obtained.

The capacity of words to bring information is not the only criterion for an optimal communication system. A second desirable criterion is that for each referent an agent consistently uses the same word. The former criterion of specificity could perfectly be satisfied by using a new word each time an agent wants to communicate a referent. Although all words would perfectly identify the referent they corresponded to, such a state of affairs is not desirable, as it would be difficult or, in this extreme case, impossible, to learn the language these agents speak. This second requirement, which will be called *consistency*, can be measured using the very same principle as the first measure, but reversed; that is, by considering to what extent a referent identifies the word an agent produces to express it.

If both specificity and consistency are achieved by a system of communication, there is a strong correspondence between referents and the words an agent uses to refer to them; a referent identifies a word, and a word identifies a referent. The only remaining requirement for perfect communication is that the agents not only consistently use a word for each referent, but that different agents use the *same* words for those referents. This is expressed in the *coherence* measure, which is calculated as the maximum fraction of agents that have the same preferred word (i.e. the word they use most often) for a referent.

Having described ways to measure the satisfaction of requirements that may be posed for a system of communication, conceptual systems can now be considered. Interestingly, the requirements for a conceptual system are similar to those used to evaluate the communicative behavior of a single agent; whereas words should capture information about the environment to represent it in communication with other agents, the meanings inside an agent's head must capture information about the environment to represent it to the agent itself. Thus, a meaning should identify a referent, and the degree to which it succeeds in doing so for the meanings of an agent is expressed by the *distinctiveness* measure, named so because it indicates whether the agent can distinguish between the various referents. Analogously, the extent to which referents identify meanings is called the *parsimony* measure, and expresses whether the agent has developed a parsimonious set of meanings. In contrast with the case for communication systems, the coherence between the conceptual systems of different agents is not measured; as long as words can be associated with meanings such that coherent communication develops, there is no need to demand coherence between the meanings these agents use internally to achieve this coherence.

1.7 Commitments

In order to determine the relevance of research results and to appreciate their significance, it is important to know the commitments made in the design of the experiments[7]. A central tenet is that agents autonomously construct the concepts they use to communicate about their environment.

[7] This criterion has been proposed as a feature that distinguishes different cognitive architectures (Van de Velde, 1995).

This implies that agents may not be assumed to possess identical concepts[8]. This point of view has earlier been taken in the research of Luc Steels, see e.g. (Steels, 1996c, 1996d, 1996a), which has been of great influence on the ideas in this work.

Although concepts are formed independently by each agent, it is clear that communication can only be functional if there is a close match between the concepts constructed by the different agents, even if these are not identical. For this to be possible, two additional assumptions are necessary given that the agents share the same environment:

- The concept formation mechanisms of the agents are similar.
- The perceptual apparatuses of the agents are similar. If they would be different, the agents would view the world in different ways. A good example is given by the paper 'What is it like to be a bat?' by Thomas Nagel (Nagel, 1974), which asks the reader to imagine what the world would be like when 'viewed' using echo-location.

In human language, a complementary, and probably more influential source of coherence between concepts is that language influences the concepts people form. This relates to the Principle of Linguistic Relativity, see e.g. (Foley, 1997), and is discussed in some more detail in chapter 2.

Furthermore, the following commitments are made:

- Meanings are not passed directly from one agent to another. To some extent, this is already implied by the previous point; if agents form concepts themselves, this means they do not receive them from some other source. There are two reasons for making this commitment. First, it poses less restrictions on the concept formation mechanisms that agents may employ; if agents *were* to share concepts, some format for specifying these concepts would be required, i.e. a common language for defining concepts. It may be very difficult to define such a language for the interesting and complex forms of concepts that lean on associations with experiences. Investigating how communication can be formed without sharing meanings explicitly leaves open the

[8] A possible objection to this is that the generalization capacities of agents may be such that, in combination with the regularities of the world, they come to construct identical concepts. Although this can certainly be achieved in simulation, it is only likely to happen when the shapes of the concepts agents are allowed to form are restricted. In order not to introduce such restrictions, it is necessary to assume that agents may possess different concepts.

possibility of discovering communication formation mechanisms that can be used by populations of agents with different internal structures. Such mechanisms can be required for agents that develop concepts based on their specific experiences, and suited to their specific competences and requirements, as such concepts would not function in the same way when used by an agent with different experiences. Second, the commitment increases the chances of finding results that apply to other communication systems where agents can not look 'inside the head' of other agents. Communication between humans is an example of such a system.

- There is no central control to guide the formation of concepts or communication. This is not to say that the development of communication using a central control is not an interesting problem, merely that it will not be investigated here. This has implications for the possible interactions between agents; for example, it means that it is not possible to let one agent choose a system of communication and allow all other agents to learn it. Coordination mechanisms that do not require central control are of interest because they can be used in situations where agents are more or less equivalent, in that they can all influence the course of affairs. Such systems can be more robust than centralized systems; the presence or absence of a number of agents does not change the capacity of the group to achieve coordination, as can be seen in the development of human languages or food trails in ant societies.

- Agents are not appointed fixed roles; there is no separation between teaching and learning agents, and every population member at times acts as a speaker and at times as a hearer. Like the previous point, this allows all agents to influence the system of communication that develops.

- No direct feedback on the quality of the communication system is available. The only way in which an agent can notice whether it interprets received signals correctly, is to consider the feedback on its actions. This assumption is necessary because of our interest in the *autonomous* development of communication; since the agents determine their own feedback, direct information about the complete communication system (as determined by the communicative and interpretative behavior of the collected agents) is not available. Rather, the fact that communication can positively influence the behavior and the experiences of an agent is one of the mechanisms that provide feedback to communication.

1.8 Restrictions

The question of how agents can autonomously form concepts about their environment and arrive at a system of communication is a very general one. In addressing this question many restrictions have been made to limit the scope of the investigation. First there are the commitments that have been described; in a way these are also restrictions. However, the commitments differ from the remaining restrictions in that the restrictions are primarily made to limit the scope of the investigation; in subsequent research, extensions would rather consider a different choice of restrictions than a different set commitments, as the latter have been made with particular goals in mind, as has been described.

Both the idea of using the interaction history to construct situation concepts and the idea of adapting relations between concepts and words to arrive at a system of communication, are of a general nature, and can be instantiated in many different ways. To use these ideas in experimental research, a particular instantiation has to be chosen. A good idea in such situations is to select the simplest instantiation possible; in this way, the behavior of the system that is investigated is as uncomplicated as possible, which best allows one to analyze why the system works or why it doesn't work. As more knowledge is gained about the ideas, more complex and powerful instantiations can then be investigated in further research. In the above, apart from the commitments, several restricting choices have been mentioned that were made to keep these first experiments on communication about autonomously formed situation concepts simple. Here they are reiterated:

- The investigation concerns *learned* as opposed to *evolved* communication
- Only *discrete* communication is considered, as opposed to *analog* communication.
- Messages used in communication consist of single words; thus, grammar plays no role.
- The concepts are situation concepts, defined by the criterion that they can be determined from the history of interaction between an

agent and its environment, and have predictive value regarding some aspect of the future state evolution of that environment, possibly conditioned on the actions the agent may select. In the particular method for forming situation concepts that has been used, called the Adaptive Subspace method, concepts correspond to unbroken regions in sensor space. This implies that the features that constitute a concept are limited to the sensor values at the current point in time. Possible extensions would be to include information about previous time steps. Another possible extension is to use more sophisticated features as a basis for concepts, e.g. if an agent would receive video images as sensor input, filters are probably better suited to provide input to the concept formation mechanism than the direct sensor data. Finally, the shapes of the concepts could be extended. The shape of the regions does not have to be restricted to hyperrectangles; concepts could also consist of conjunctions of regions, or curved multidimensional volumes, and they might be allowed to overlap. Also, the relation between a point in sensor space and a concept does not have to be binary, but could for example be based on fuzzy sets, in which case each point is associated with a value that expresses the degree to which the point is a member of the concept.

- Concepts are formed by all agents on the basis of their own experiences. Thus, communication has no direct influence on the concepts that are formed, and the left arrow in figure 1.2 is not functional. This important restriction simplifies the analysis of the concept formation process. However, it also is a limitation. The influence of communication on concept formation represents the powerful idea that concepts can spread through a population, as in the example of the fishermen.

The strongest restrictions have been described. Further choices made in instantiating the ideas to arrive at an implementable mechanism are documented as part of the description of the concept formation and association formation mechanisms, as describing them here would require going into too much detail.

1.9 Contributions

The contributions of this thesis are the following:

- A contribution to the dynamical systems perspective on the development of language and communication is delivered by describing a mechanism that, when followed by individual agents to update their associations between concepts and words, leads to coherent communication. For a deterministic version of this mechanism, it is proven mathematically that systems of ideal communication are attractors of the dynamical system determined by the associations. Furthermore, it is demonstrated experimentally that the standard, stochastic mechanism, which more reliably results in communication, has pseudo-attractors that play a role similar to the attractors in the deterministic system.

- A class of concepts called *situation concepts* is described. This class of concepts is particularly relevant to the problem of communicating information about the environment. An algorithm for the formation of one of the most basic forms of situation concepts is described, called the *adaptive subspace* method. This method performs generalization in continuous sensor action spaces. By keeping sensor and action distinctions separated, concepts are obtained that do not require information about the action to be selected, but are identified by the current sensor information. Because of this property, the concepts can be used by agents to communicate about the current situation before selecting an action.

- An algorithm for updating the associations between words and the concepts of an individual agent is described. The algorithm is specifically suited to be used in combination with situation concepts, as it provides a way for an agent to determine whether it interpreted a word it received correctly. This function is based on matching the subsequent behavior of the environment to the expectation of the agent. The mechanism is a demonstration of how evaluative feedback may be involved in language learning without relying on nonlinguistic communication or other sources that evaluate communicative behavior. In an additional experiment, it is demonstrated that communication can even be achieved without this information, called *success information*, and that the specificity of words required for unambiguous communication can also be obtained as a result of lateral inhibition alone.

- Measures are introduced that express the quality of conceptual systems and communication systems. The measures are principled in that they are based on the decrease of uncertainty achieved by these systems, and complete in that for both conceptual and communica-

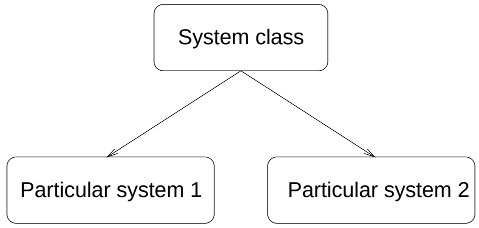


Figure 1.3: Diagram illustrating a methodological danger in artificial intelligence research (see text).

tion systems, optimal values for all measures of a system implies an optimal system.

1.10 Methodology

The methodology of this research is somewhat complex. The primary purpose of the experiments reported here is to determine whether particular mechanisms are successful in the formation of concepts and communication. This purpose is given in by the underlying objective of developing artificial systems that are able to communicate about their environment and use this facility to coordinate their behavior. However, at the same time, an aim is to investigate the general question of how agents can develop communication in such a way that findings may apply to the specific case of human language. Although these aims may sometimes conflict, they are not contradictory. An example of a commitment that is made partly in view of the second aim, is the commitment that agents do not pass meanings to each other, although additional reasons for the commitment have been outlined.

The above methodology is not uncommon in artificial intelligence research, which has both the goal of constructing novel systems that display intelligence and of improving our understanding of living examples of intelligent systems, although this is not always explicitly pronounced. The approach requires careful consideration in the interpretation of experimental results. A clear danger is that results are interpreted beyond their scope. Consider for example figure 1.3. Although both particular system 1 (e.g. a human population) and particular system 2 (e.g. a particular brand of autonomous agents) are particular members of a certain class of systems (populations of autonomous agents), this does not imply that properties of particular system 2 (e.g. conditions under which communication develops) also hold for particular system 1 (and vice versa). However, this methodological danger does *not* imply that knowledge about a particular system can only be gained by investigating that particular system. The abstraction from a particular system to a more general class of systems (autonomous agents, or populations thereof) can lead to two forms of solid results.

First, general properties of the system class must, if the class correctly represents the particular systems under study, by definition hold for all particular systems. Some examples of such properties are the second law of thermodynamics, which says that the entropy of a closed system can only remain constant or increase, or the property that a signal must be minimally be sampled at the Nyquist frequency in order not to lose information, or the property that at least $^{2}log\ n$ bits of information are required to encode n equiprobable outcomes, etc.

Second, the investigation of a particular system may be used to investigate possible mechanisms that produce a phenomenon of interest. The demonstration that a proposed mechanisms produces a certain phenomenon in particular system 2 can not in itself prove that this mechanisms must be the explanation for a phenomenon observed in particular system 1. However, it *can* suggest a mechanism that *may* be responsible for such a phenomenon. The role of such experiments is therefore primarily in testing whether hypothesized mechanisms may work, and in generating new hypotheses when they do not. The value of such a research strategy, which is adopted here, is therefore in its potential to gain insight into proposed mechanisms. When such research reliably demonstrates a mechanism to produce a certain phenomenon, the possible role of the mechanism in other particular systems displaying the same phenomenon can be hypothesized and subsequently investigated.

1.11 Relevance and Motivation

The two central topics of this thesis, concept formation and the development of communication, are both relevant to artificial intelligence. To appreciate this, some familiarity with the bottom-up route to artificial intelligence and the notion of shared ontologies is required.

1.11.1 The Bottom Up Route Versus the Top Down Approach

The top-down approach to artificial intelligence begins by analyzing the requirements of an intelligent systems and gradually refining these, down to a level where all required components can be implemented. A typical area where this approach is followed is the design of expert systems. The approach has the advantage that high level constructs can be used right from the start. However, it involves a serious risk. The risk is that there may be no path from the high level model leading down to the lowest level where a system can be grounded by coupling it to the environment. This is a possible consequence of assuming a structure for high level concepts. If the actual structure of high level concepts is different from the one hypothesized, then the approach is likely to fail. This problem is associated with that of *brittleness* (Holland, 1983) in expert systems. Brittleness refers to the notion that when a system is used for goals slightly different from those held in mind by the designer, the resulting behavior of the system may be useless. An issue related to this is the *grounding problem* (Harnad, 1990). This signifies the problem that when a the meanings in a system, represented as symbols, are only defined in terms of other symbols, the system is not linked to the environment.

These ideas have been an important influence in the rise of a different approach to synthesizing intelligence: the bottom-up approach, see e.g. (Steels & Brooks, 1995). According to the bottom-up approach, the research and development of intelligent systems should start at the lowest level, e.g. the level of sensory-motor interaction in the case of robots. An important advantage of this approach is that it produces models along the way that can be tested to see if they *work*. Within artificial intelligence, much current research follows the bottom-up approach, for example research into neural networks, behavior based robotics, and dynamical systems. It is assumed that gaining knowledge about these low levels of operation clears the way towards higher level behavior and, eventually, full-fledged intelligent systems. Research into communication along these line addresses the symbol grounding problem by developing relations between the environment and the symbols used in communication via the perception of the agent as provided by its sensors. Examples where symbols are grounded in the physical environment as opposed to simulated environments are provided by research into communication development on robots, see e.g. (Vogt, 1998; Steels & Vogt, 1997; De Jong & Vogt, 1998; Billard & Hayes, 1999; Oates et al., 1999)

1.11.2 Shared Ontologies

In recent years, the notion of a multi-agent system has become influential in computer science, see e.g. (Weiss, 1999). The idea of these systems is that instead of determining a single sequence of operations that will solve a problem, it may sometimes be more effective to analyze a number of concerns that need to be taken into account, and let each of these concerns be represented by one or more agents from a collection that makes up the multi-agent system. The possible advantage of this approach is that instead of having to specify a sequence of operations that coordinates the different concerns, the required coordination may result from interaction between the different agents, each guarding their own concern. If a suitable coordination mechanism is found, this reduces the task of the designer to implementing the individual agents, thus relieving him or her of the need to worry about the interaction between the different activities.

The metaphor of a multi-agent system is based on human organizations, where the possible benefits of this form of coordination is most clear. A difficulty with computer systems however is that current command languages require a formal and exact notation of the task that is to be performed. Thus, when large multi-agent systems are designed, consisting of entities that are contributed by different sources (persons, organizations), it is necessary to specify exactly the meaning of the concepts used by agents in communication. This problem has lead researchers with an interest in multi-agent systems to propose formal specifications for agent communication, such as KQML (Finin, Fritzson, McKay, & McEntire, 1994) and FIPA (fipa, 1999)[9]. However, as noted in (Steels, 1998), this may not be the best way to proceed. When all concepts that are used by agents in a multi-agent system need to be defined beforehand, a large potential for flexibility is lost; the introduction of new agents to the system with capabilities that were not available before can not be benefited from unless these new capabilities happen to be expressible in the existing language.

1.11.3 Relevance and Motivation

Now that the notions of a bottom-up approach and shared ontologies have been described, the relevance of the research in this thesis to computer science can be discussed. With respect to the bottom-up approach, an important open research question is how concepts may be formed based on experience. The mechanism used in the concept formation experiments takes experiences as input and forms concepts, based on patterns in this data. These concepts allow the agents to adapt their behavior such that it becomes more appropriate with respect to the environment, as evaluated by the increasing success of their actions. Although the concepts that are

[9] Although the FIPA specification includes a protocol for the exchange of rule-based code, the language for these exchanges still presupposes a public, shared semantics.

formed are of a very basic nature, the experiments are interesting in that they demonstrate the potential for evaluative feedback to serve as a guiding principle for concept formation.

Coordination requires a common language. This is a problem in multi-agent system design, since part of the potential power of multi-agent systems is that they can adapt by incorporating new agents. However, if the requirements for agents are too tight, this limits the class of new agents that may enter the system. In (De Jong, 1997b), a coordination method for multi-agent systems was proposed that limited communication between agents to the transmission of numerical evaluations. The evaluations are sent from one agent to another, and indicate to what extent the sender appreciates the behavior of the receiver of the evaluation. Although this approach addresses the problem of shared ontologies by reducing the complexity of the common ontology to a minimum, it also seriously limits the complexity of communication.

In the subsequent research that is reported in this thesis, a different approach has been explored. The language used by the agents to communicate about their environment is formed by the agents themselves. It is based on concepts that are formed by the agents and grounded in interaction with a shared environment. Even though, due to different experiences, the agents may form different concepts, the language that is developed in the experiments is such that agents use the same words for the same situations, and may thus be said to be a shared language. To overcome the fundamental limitations connected with fixed ontologies, coordination methods along these lines may be necessary in the long run.

1.12 Nomenclature and Notation

Following Steels, see e.g. (Steels, 1996c), the internal representations that are formed by the concept formation process are called *meanings*, while the objective aspects of the environmental state these meanings correspond to will be called referents.

In the simulation experiments, the referents are the optimal concepts, i.e. the concepts that distinguish between states of the environment where different actions are required, but that make no unnecessary distinctions. It is only due to the uncomplicated nature of the experiments that these referents can be determined. In more complicated environments, such as the real world, this will often not be possible. Each individual agent only has access to the concepts it formed itself, as described in the introduction.

Therefore, referents only play a role in the *analysis* of experiments, not in the experiments themselves.

Meanings, the categories of sensory experiences formed by agents, are also referred to as *concepts*. Although this is not uncommon in the machine learning community, see e.g. (Mitchell, 1997), it may raise eyebrows of people with a different background, such as psychology, linguistics or cognitive science. For these readers, it is important to state clearly that the concepts formed by the agents in the experiments are not viewed as models of human concepts; they are far simpler. However, the particular structure of these concepts does not imply the assumption that all concepts have similar properties. What the research does assume about concepts is they are entities that relate possible contexts to a word. This has no direct implications for the *structure* of concepts; even highly abstract concepts such as *democracy* or *reflexivity* are relevant in certain contexts and not in others; the assumption here is that a concept allows a speaker to distinguish between these classes of contexts[10]. In this thesis, the focus is on a particular kind of concepts, called *situation* concepts. A situation concept captures aspects of the environment that determine its expected future behavior.

Matrix notation is used to describe the relations between referents, concepts, and words. Conceptual systems are characterized by the conditional probabilities of a referent ρ activating a meaning μ in the matrix $P(\mu|\rho)$, and of the probabilities of a referent being present given the activation of a meaning in the matrix $P(\rho|\mu)$. Concerning communication, analogous matrices are used, but here there is an additional distinction, viz. whether the probabilities concern production (speaking) or interpretation (hearing). The production behavior of an agent is represented by probabilities relating meanings to signals or words σ in the matrix $P_{prod}(\sigma|\mu)$, while interpretation is characterized by the conditional probabilities in the matrix $P_{int}(\mu|\sigma)$. Matrix multiplication can be used to determine an objective representation of an agent's communicative behavior as provided by the matrices $P_{prod}(\sigma|\rho)$ and $P_{int}(\rho|\sigma)$. The objectivity of these matrices derives from the fact that no meanings but only publicly observable entities are involved, and is useful in comparing the communicative behavior of different agents. Measures based on these matrices are given where they are introduced in the text.

Much has been written about the problem of defining communication, see e.g. (Hauser, 1997; Noble, 1998; Di Paolo, 1997). As communication oc-

[10]The applicability of a concept to a context can also be gradual. In the experiments here however, it will be binary.

curs in a wide variety of contexts, and its study likewise concerns different aspects and forms of communication, it appears unlikely that a single definition can be given that covers all instances of it without being over-inclusive. The form of communication that will be studied here is characterized by a sender, a symbolic message, and a receiver. The defining requirement that will be posed before this process can be called communication is that by the act of sending the message, the sender provides information to the receiver. Here information refers to the notion as it is used in information theory, where information is seen as uncertainty, see (Shannon, 1948). The uncertainty of the receiver concerns its knowledge about the state of its *environment*. Thus, the information of a message is determined by the degree to which it improves the receiver's knowledge about its environment. This definition has the consequence that the information of a message can not be considered separate from the context and the receiver.

The definition of communication used here addresses a possible question concerning the subject matter of the thesis, namely to what extent communication *develops* in the experiments; if an explicit mechanism for adapting associations between meanings and words is present, are agents not already communicating from the very beginning? Although providing such a mechanism implies that the development of this mechanism can not be the subject of experiments, the answer to the question is that communication *does* develop. This can be seen by considering an experiment. At the beginning of each experiment, agents have not yet received any words yet and hence they are unable to extract information from any word that is sent to them. This implies that according to the above definition, no communication can occur at this point. As will be seen, this situation changes in the course of the experiment; interaction between the agents brings about a situation where the same words are used for the same referents. In this situation, receiving a word increases the knowledge of an agent about its environment when this knowledge was incomplete, and hence communication has developed.

A final note concerns the use of the term *language*. Here, this term will be used to denote any system relating symbols to meanings. Thus, it is a technical term here (cf. programming languages, formal language), and refers to systems of a whole different order of complexity than human language. Accordingly, the systems encountered in the investigation should not be viewed as models human language or animal communication. Rather, the aim is to investigate issues concerning the formation of concepts and communication for autonomous agents in general. As argued, this approach may lead to knowledge about principles involved in the development of communication, and such knowledge *can* be relevant to the study of human or animal communication.

1.13 Outline

In chapter 2 evaluative feedback is described as a possible guiding principle for concept formation. According to this perspective, certain generalization methods used in reinforcement learning can be used for concept formation. One such mechanism, based on incrementally partitioning the sensor-action space(the space where each point represents a (possibly multidimensional) action and the values of the sensors at the moment the action was selected), is described. A matrix notation is used to describe the result of concept formation as conditional probabilities specifying the relation between meanings μ, subjective entities formed during concept formation, and referents ρ, objective entities determined by the structure of the environment, leading to the matrices $P(\mu|\rho)$ and $P(\rho|\mu)$. The entropy in these matrices corresponds to the uncertainty in determining a meaning or a referent. This leads to a principled definition of two measures indicating the quality of a conceptual system. The effects of sensor and evaluation noise on concept formation are investigated.

The formation of concepts can serve as the basis for mechanisms governing the association between privately formed meanings and words. In chapter 3 a mechanism with special relevance to the case where communication concerns states of the environment is described. It is experimentally demonstrated that when the agents in a population use this adaptation mechanism, their initially random associations between meanings and words converge towards a shared system of communication. Analogous to conceptual systems, communication systems can be captured by specifying the production of words σ in the conditional probability matrix $P(\sigma|\mu)$ and the interpretation behavior in the matrix $P(\mu|\sigma)$. These subjective matrices can be combined with the concept formation matrices to yield the objective production and interpretation matrices $P(\sigma|\rho)$ and $P(\rho|\sigma)$. These matrices give rise to entropy based measures that quantify the quality of communication.

In the same chapter, the necessity of the different components of the algorithm is assessed by removing each component and comparing the development of communication of the modified system to that in the standard system. The comparisons show that each component when removed leads to a statistically significant degradation of the quality of communication. Analysis of the component responsible for maintaining information about the success of communication reveals that not the success information itself, but rather the lateral inhibition performed by the component is essential for the development of communication.

In further experiments, differences between the privately formed meanings are induced by varying the noise on the sensor inputs of different agents. These experiments demonstrate how interactions based on different conceptual systems can lead to a public, shared language.

The strengths of associations between meanings and words of the different agents determine a high dimensional dynamical system. In chapter 4 it is shown, both mathematically and experimentally, that a deterministic version of the system converges to point attractors that correspond to ideal systems of communication. Based on these results, it is shown that the stochastic systems has points in its phase space that play a role similar to the attractors in the deterministic communication system. Stochasticity is found to be a useful ingredient for the development of communication in that it encourages exploration; the introduction of stochasticity improves both the stability and range of conditions under which communication is developed. Finally, the behavior of the system in phase space as a function of *temperature*, an important control parameter, is examined, providing insight into the effects of this parameter. The analysis in this chapter demonstrates that it can be fruitful to view the development of communication as the behavior of a dynamical system.

Finally, chapter 5 presents conclusions and briefly looks ahead.

Chapter 2

Autonomous Formation of Concepts

2.1 Introduction

This chapter describes a way for autonomous agents to form concepts about their environment based on their experiences. As described in the previous chapter, the only assumption adhered to here concerning concepts is that they relate possible contexts to a word. Although the knowledge of the particular set of contexts in which a given concept is appropriate does not define or identify that concept, this assumption is minimal in the sense that any agent wishing to use a concept in communication has to be able to determine this relation.

A further aspect of concepts as they are viewed here is that they partially determine the behavior its holder can produce; the concepts an agent has at its disposal determine the way this agent views the world. The aim of this research is not to present a model for concepts that matches the concepts humans possess as closely as possible. Rather, it is to investigate one of the principles that may guide the formation of concepts. Even though the concepts in the experiments will be of a much simpler form than the concepts humans employ and only concern a particular *type* of concepts (those that capture information about the situation), the principle may also play a role in human concept formation. The principle is that concepts should improve the ability of an agent to produce successful behavior.

The approach requires the presence of evaluative feedback. Arguments are given for the potential of adaptation based on such feedback, suggesting this assumption is not implausible. Furthermore, a matrix notation is introduced that describes the result of concept formation, and principled measures to evaluate the quality of a conceptual system will be derived.

The first section discusses the influence of feedback on behavior, and argues for the potential of evaluative feedback as a criterion guiding various forms of adaptation. Two such forms are *selectionism* and *reinforcement learning*. Selectionism is discussed in section 2.3. In section 2.4, reinforcement learning is discussed. Next, in section 2.6, experiments in concept formation are presented. The presentation includes the definition of a matrix notation capturing the result of concept formation, and the definition of measures that monitor the process of concept formation. The chapter ends with conclusions in section 2.7.

2.2 Feedback and Adaptation

2.2.1 Concepts Influence Behavior and Communication

The influence of concepts on behavior is twofold: first, concepts determine the way an agent views the world, and the actions of an agent therefore depend on the concepts that are evoked in the agent. The other influence is that on communication; the lexicon used in communication is a set of words that are associated with concepts, and thus the set of concepts an agents has at its disposal confines the set of words it may use in communication. The latter influence is a special case of the former. The realization that concepts influence behavior leads to a principled research paradigm for investigating the formation of concepts, which is laid out in this introduction.

Before continuing though, it is important to notice that there is not only an influence of an individual's concepts on communication, but that there can also be an influence the other way around. To appreciate this, it will be useful to consider the example of human communication. The words a person encounters during his or her life determine what concepts will be learned and hence the ways in which the person thinks and views the world.

Different versions of this idea have been described by several researchers. Foley (1997) describes its history as follows. Boas pointed out the function of language in organizing our experience of the sensible world, and emphasized its classificatory function. The difference with earlier researchers considering this idea is that Boas appreciated the necessity of grounding the ideas in empirical work. Sapir (1949) described the idea that the world is to a large extent unconsciously built up on the language habits of the group. This is known as the Principle of Linguistic Relativity. This principle was not viewed as an experimental hypothesis, but rather as an axiom, especially in the case of Whorf.

Different versions of this idea are known as the Sapir-Whorf hypothesis, which has been surrounded by controversy. This may be partly due to the many interpretations that can be given to the idea that language influences thought; if these are exaggerated, they can lead to the idea that someone's native language determines his or her potential for thought. Unmindful reports of the evidence may have been another factor; a good example of this is the number of words Eskimos use for snow, which has varied from 3 to (at least) 400 in written reports (Pullum, 1991).

For scientific knowledge about the issue, empirical evidence is required. A recent example of such evidence has been reported by Davidoff. In (Davidoff, Davies, & Roberson, 1999), he reported experimental evidence that people with native languages containing different sets of color terms show differences in remembering colors. More specifically, color categorization tasks were learned more rapidly if the distinctions were consonant with those made in the language of the subject. This appears to be a clear example of the influence of language on memory, an aspect of cognition, although other explanations of the findings are still possible (e.g. that the differences are not a result of the different color terms in the languages, but of environmental differences).

Perhaps some of the controversy can be removed by considering the influence of language on a single person. This results in the more general idea that the words and corresponding concepts a person possesses influence his or her thought. The verity of this idea can almost be seen a priori; whenever one comes to learn a new word and its meaning, e.g. *serendipity*, instances of this meaning can henceforth be recognized and therefore influence one's behavior.

Although the influence of language on concepts is an interesting and important phenomenon, it will not be studied here; rather, the subject of study will be the influence of the concepts formed by agents on their behavior and on the language they develop.

2.2.2 Feedback on Behavior may Guide Concept Formation

Although the realization that concepts influence behavior may seem obvious, it is important in that it leads to a principled research paradigm for investigating the formation of concepts. For this to follow, an assumption has to made, viz. that the concepts an agent forms are *useful* in the way they may influence that agent's behavior. Given this assumption, the formation of concepts may be investigated by considering any concept formation method that yields concepts which improve the behavior of the agent. But before continuing, a possible objection needs to be addressed.

2.2.3 Where Does Feedback on Behavior Come From?

The assumption requires that the agent can somehow determine what is useful behavior and what is not. Ultimately, useful behavior is behavior that ensures survival. Survival provides environmental feedback by selecting individuals with the capacity to survive long enough to reproduce. However, the forces of Darwinian evolution operate on the population level and can not provide feedback on behavior at the smaller time scale of an agent's lifetime. Indirectly though, they *do* provide a possible explanation. The solution is as follows. Whether an individual survives depends on factors such as its ability to maintain homeostasis (Cannon, 1932) (which involves keeping properties such as body temperature, oxygen level in the blood, hydration, etc. all within certain ranges), and to avoid dangers (fire, fast moving objects, falling from a high tree, etc.). From the principles of Darwinian evolution, it may be expected that evolution favors maintenance of homeostasis and recognition and avoidance of threats. The most direct way for evolution to induce particular behaviors in agents is to construct organisms in such a way that they will produce these behaviors under any conditions that are normally encountered in the environment of the organisms; the behaviors could be said to have been *hard-wired* into the organism. Whenever possible, this way of ensuring behavior is most likely to be successful, and it is responsible for many primitive forms of behavior, such as reflexes. Examples include *breathing*, a reflex activated by chemoreceptors detecting the oxygen level in arteries near the heart, and the *heart beat*, which is regulated by reflexes acting in response to activity of mechanoreceptors detecting high and low brood pressure.

Although hard-wiring is a good way to ensure some behaviors, it lacks the flexibility that is needed for more complex tasks. Consider for example the problem of obtaining food. For basic life forms, e.g. bacteria, it may be sufficient to have a mechanism that retains food once it enters the organism. To obtain the diverse diet that humans need is rather more complex. At the very least this involves recognizing appropriate types of food. After the early phases in a person's life, some degree of sensory motor control will be required in order to successfully take in the recognized food items. Moreover, it is typically not sufficient to recognize food whenever it comes along; from the hunting and gathering societies to the world of today, assuring food requires a form of interaction with the environment that cannot be captured in any small set of responses to be executed upon reception of stimuli. Rather, people adapt to their environment using trial and error. After several crop failures, a different piece of land will be tried. If more fish are landed in the morning than during the afternoon, a fisherman will make it a habit to rise early. If merely displaying vegetables in a market place does not attract the customers necessary for a stall-holder to earn her living, she may decide that some shouting is called for. In these examples, variation of behavior leads to the discovery of a better practice. It is important to notice however that this variation is not random; some form of model of the world is used to limit the production of inappropriate behavior, and in this manner errors can be limited. This is an important argument against the idea that "the obvious problem with trial-and-error learning is error" (Pulliam & Dunford, 1980), quoted in (Oliphant, 1997); using basic knowledge about the world to avoid unnecessary mistakes does not prevent anyone from evaluating his or her own behavior and using this information to learn and improve that behavior.

A common factor of the examples is that this adaptation of behavior only requires evaluative information about the behavior that was produced. The agent uses its own evaluation as qualitative feedback on its behavior. Considered from an evolutionary perspective, a genetic adaptation that equips an organism with the ability to evaluate its own behavior and use it as a guiding principle for adapting it is incomparably more powerful and flexible than a repertoire of fixed routines and reflexes. Concluding, the answer to the question posed in this section is that evaluative feedback probably comes from the need for flexible, adaptive behavior.

2.2.4 Using Feedback to Guide Behavior

The idea that organisms adapt their behavior according to some evaluative criterion was first formulated by Edward Thorndike in his Law of Effect (1911):

> Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have

> their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.

Around the same time, the behaviorist school of psychology gained momentum. The central tenet of behaviorism is that the study of psychology should be based on objective observations, and that any reference to internal, mentalistic notions should be avoided. Pavlov investigated classical conditioning, the association of a random stimulus with a natural reaction, such as a dog's salivating in response to the presence of food in its mouth; when the bell sounds every time the dog receives food, the salivation reflex will be triggered by the sound of a bell. Skinner built boxes to study the behavior of animals, which became known as Skinner boxes. One of the phenomena he investigated is that a pigeon can learn to perform a random action (pressing a lever) in response to a random condition (a flashing light) when this action is followed by a reward (food); this principle, where the learned behavior is a new one, is known as operant or instrumental conditioning.

Both classical and operant conditioning are forms of Stimulus-Response learning, a very simple type of learning where some action always follows a particular condition. They are perfect examples of Thorndike's Law of Effect. However, it is these types of very basic adaptive behavior that are often associated with adaptation based on evaluative feedback. What is often not realized is that evaluative feedback can be used in an indefinite number of ways to guide behavior. One of these is selectionism, where the principle is that new, (partially) random structures are generated, and an evaluative criterion guides the selection of structures that remain, while other structures disappear. Another one is reinforcement learning, where the usual approach is to learn the value of different behaviors. The theory of optimal control is still another field where behavior is based on evaluative (and other forms of) feedback, the difference being that here a model of the system is usually present.

When humans are viewed as having a model of their environment, it becomes clear that evaluative feedback can in principle guide sophisticated behavior, such as planning. As the examples that were given earlier purport to show, there is no reason not to consider more sophisticated forms of adaptive behavior guided by evaluative feedback. Since models or experiments requiring evaluative feedback are not considered plausible by some, it may be useful to note that neurophysiology has indicated the presence in primates of neurons that produce signals related to the prediction of future salient and rewarding events (Schultz, Dayan, & Read Montague, 1997). More completely, there is substantial evidence that dopaminergic neurons, located in the pars compacta of the substantia nigra and in the ventral tegmental area, play an essential role in both the primary reinforcement of behavior and in guiding preparatory behavior on the basis of the likelihood that the animal will subsequently receive reinforcement (Houk, Adams, & Barto, 1995), although dopamine is not the only reward transmitter, and the dopamine system probably also relates to other aspects of motivation (Wise & Rompre, 1989).

Naturally, the presence of neurons correlated with rewards on itself does not imply that this presence is innate in the sense that it invariably develops regardless of the lifetime experience of the animal; however, finding such neurons consistently in the members of a species does imply that evolution has developed individuals such that these neurons develop under normal circumstances.

In summary, it appears useful from an evolutionary perspective to equip organisms with the ability to produce evaluative feedback on their own behavior, and evolution indeed has developed such functionality. The work in this thesis, and much other work in artificial intelligence, is built on the assumption that some form of evaluative feedback is available to the agent.

Using Evaluative Feedback to Guide Concept Formation

At the beginning of this chapter, the role of concepts in guiding behavior was mentioned. In combination with the idea of evaluative feedback guiding behavior, this naturally leads to the idea of using evaluative feedback to guide concept formation. It is this approach that will be followed here.

The idea of using behavioral success to guide concept formation was also used in a proposal for concept formation by Stefan Wrobel (Wrobel, 1991). The model was somewhat ahead of its time, in that the quick growth of the field of reinforcement learning was yet to come, and indeed Wrobel mentions that the emphasis in the field was on action selection, rather than representation development. Naturally, the primary goal of reinforcement learning methods is to learn to produce appropriate behavior, but meanwhile there exists a substantial amount of work in reinforcement learning where internal representations are constructed as part of this. A good example is Andrew McCallum's U-Tree algorithm, which will be discussed in more detail later on in this chapter.

In Wrobel's model, preprocessing elements segment the sensor ranges, and the resulting intervals, which he refers to as the symbolic vocabulary of the system, are used to learn a tree (conceptual hierarchy). The nodes in this tree represent situations in the environment, and in this sense the method is similar to the one that will be described here. A limitation of Wrobel's model is that it assumes binary rewards, and does not take the issue of delayed rewards into account; hence, it is not applicable to the general reinforcement learning problem. The methods that will be described later on in this chapter overcome this problem.

Using Supervised Feedback to Guide Concept Formation

There is no reason to assume that the feedback available to agents is limited to the evaluative sort. Although this is the most basic form of feedback that can be imagined, some forms of learning in human appear to benefit from richer forms of feedback, that might be referred to with the term *supervised feedback* as it is used in computer science.

Supervised feedback on behavior yields information about what other behavior would have been more appropriate than that produced by the agent. Thus, it tells the agent what it should do, and for this reason this form of feedback is sometimes called *instructive* feedback. Adaptive mechanisms based on supervised feedback can use input-output examples (sensor-action examples) to learn a mapping from inputs to outputs. This description fits the fields of pattern recognition (see e.g. (Duda & Hart, 1973)), where the goal is to construct a system that, based on features of objects, groups objects into classes. For this reason, the problem of pattern recognition is also known as *classification*. The classes, and a set of examples for which the classes are known, have to be known in advance. Concept formation based on decision trees, see e.g. (Quinlan, 1990), is a particular type of pattern recognition.

An example of human learning that appears to be based on supervised feedback is practicing the mastery of musical instruments. If a particular sequence of motions does not produce the right sound when playing the guitar, the player would need to apply more or less random changes to this motion and judge their effect if only evaluative feedback were present. This is not the case, or at least not always; rather, since the player has a model of the instrument and his own hands, he has an idea of what changes might yield the desired sound. This idea is a form of supervised feedback, since it specifies not only that the initial behavior was incorrect, but also *how* it should change. For another example, consider a person who is learning to play tennis, and notices his last ball bounced somewhat to the left of the intended spot, e.g. due to side wind; in this case, the tennis player could use his internal model of hitting a ball to determine a motion that would have caused the ball to go more to the right, and adapt his play accordingly.

Although the above shows that evolution can result in organisms that generate supervised feedback on their own behavior, the computation of supervised feedback is generally more complicated than that of evaluative feedback. Whereas the generation of evaluative feedback only requires recognition of desired situations, supervised feedback is necessarily linked to the mechanism that produced the behavior, since the desired behavior must be specified in terms proper to that mechanism. For this reason, research in artificial intelligence, including the work reported in this thesis, often limits the form of feedback that is assumed to be available to evaluative feedback. The argument given here however indicates that this limitation is not mandatory.

2.3 Selectionism

Selectionism is a powerful mechanism that performs search in a very broad sense, its domain ranging from well-defined problems such as optimizing a function to the construction of organisms that adapt themselves to changing environments.

A large literature exists on computational methods inspired by evolution. This research is known under various names, including evolutionary computation, genetic computation, and evolutionary strategies.

In Darwinian evolution, changes become manifest after spreading though many generations of individuals; hence, it does not (directly) account for the adaptation of an individual during its life. Individuals are elements of variation in this process. *Cultural evolution* on the other hand, see e.g. (Steels, 1997b; Steels & Kaplan, 1999), describes a process of change where individuals adapt their behavior during lifetime. Thus, changes may spread from one person to another, through parts of the population that are related by interaction, and yield patterns specific to an interacting group of individuals, such as culture. The elements of variation are structures within the individual.

There is neurological evidence suggesting that selectionist principles are involved in the development of the human brain. In (Huttenlocher, 1990), measurements of the volume of the visual cortex and synaptic density have been combined to obtain estimates of the total number of synapses in the visual cortex as a function of age. This revealed that the number of synapses has its maximum around the age of 8 months, and starts to decrease to arrive at 50-60% of that amount at the age of 11 years. When related to the number of neurons, the average number of synapses per neuron in visual cortex can be estimated. This ratio peaks at over 15,000, also at 8 months, and decreases to around 10,000 in adults. Apart from this overall decrease of synaptic contacts, it may well be possible that synapses constantly form and disappear; since the data only provides estimates at fixed points in time resulting from anatomic countings, this would not be visible in the experimental setup used.

This data can be interpreted as support for the idea that cognitive development is the result of selective pruning of initially superfluous synaptic contacts, where selection is activity dependent. The main proponents of this idea are Changeux (Changeux, 1985), and Edelman (Edelman, 1987). Changeux hypothesizes that at the stage of transient redundancy, embryonic synapses can be labile, stable, and degenerated, and that labile connections can become stable (stabilization) or degenerated (regression), depending on the activity of the post synaptic cell (the cell receiving the signal). Also, stable connections may become labile (labilization). According to Changeux, "To learn is to stabilize preestablished synaptic combinations, and to *eliminate* the surplus" (emphasis in the original).

A related theory is expounded in (Edelman, 1987). Edelman has put forward the hypothesis that selection forms neuronal groups whose exact composition differs among individuals. A second selective process then acts on these groups, the criteria for selection being dependent on correlation of activity with signals arising from adaptive behavior.

2.3.1 Generation and Selection of Sensory Channels

An example of the use of selectionism as an adaptive mechanism based on evaluative feedback is the experiment reported in (De Jong & Steels, 1999), where sensory channels were generated randomly and selected according to their capacity to discriminate between different geometrical figures. Figures were represented symbolically, with representations consisting of the type of figure (triangle, circle, trapezium, square, or rectangle) and data concerning the shape of the figure, depending on its type. The primitives that were available to build sensory channels (see table 2.1) were operators that act on these representations to extract the curvatures, lengths, and angles of the curve segments that make up a figure. These operators all yield a lists of numbers, to which sequence operators (first element, n^{th} element, n^{th} subsequence of length m) and arithmetic operators (sum, difference,

Table 2.1: Functions used for generating sensory channels and their applicability

Input type	Function	Output type
figure	`break-figure`	list of curves
list of curves	`curvatures`	list of numbers
list of curves	`lengths`	list of numbers
list of curves	`angles`	list of numbers
list of numbers	`nth subsequence of length m`	list of numbers
list of numbers	`sum`	number
list of numbers	`difference`	number
list of numbers	`average`	number
list of numbers	`first`	number
list of numbers	`nth element`	number

average), could be applied. The result of generation and selection was a set of these sensory channels, which was evaluated on its capacity to distinguish between different figures; if and only if a sensory channel yields substantially different values for two figure, they can be discriminated from each other.

Although the search space in this experiment is not large, an interesting phenomenon was found. When selection operated at the level of single sensory channels, reasonable but not outstanding solutions were found. When selection was applied to complete sets of sensory channels however, much better solutions were found, even though by selecting for complete groups, information about individual sensory channels is lost. The explanation is that the group selection favors diverse solutions, containing several complementary channels. By selecting for successful individual channels on the other hand, cooperation between the sensory channels that constitute a solution is not ensured. Indeed, the difference boils down to that between selecting the best group of channels and a group of the best channels, the latter of which is less likely to be sufficiently diverse for distinguishing between a variety of figures.

The idea that perceptual mechanisms may be formed by selectionist principles is further investigated in research by Tony Belpaeme, see e.g. (Belpaeme, 1999). Marc Ebner used genetic programming to evolve an image operator yielding points that can be used to determine the optical flow in images (Ebner & Zell, 1999).

2.4 Reinforcement Learning

Another domain where evaluative feedback drives adaptation is that of *reinforcement learning*. According to (Sutton & Barto, 1998), reinforcement learning characterizes a learning *problem*, rather than a set of learning methods. Specifically, the problem is that of an *agent* in an *environment* that has to achieve a *goal*. The section begins with a brief explanation of reinforcement learning. This background is necessary to explain the *adaptive subspace* method[1], that is used for concept formation in the experiments. The adaptive subspace method is an algorithm that allows agents to develop *situation concepts*. As their name indicates, situation concepts represent situations in which an agent may find itself. From the agent's perspective, a situation is determined by the recent interactions between the agent and its environment. However, since the agent has no complete view of the environment, it may not always be able to determine its own current situation. Knowledge of the current situation is useful because it implicitly brings information about the near future.

Reinforcement learning can be described as the problem an agent in an environment faces if it wants to obtain delayed rewards. Such problems can be addressed using methods and techniques that have been developed in the reinforcement learning field; examples include learning to ride a bicycle (Randløv & Alstrøm, 1998), elevator scheduling (Crites & Barto, 1996), job shop scheduling (Zhang & Dietterich, 1996), network routing (Littman & Boyan, 1993), and all sorts of games (e.g. Backgammon (Tesauro, 1995), checkers (Samuel, 1959)). An excellent introduction to the subject is provided in (Sutton & Barto, 1998). For a shorter overview, see (Kaelbling, Littman, & Moore, 1996).

A cycle or step in reinforcement learning proceeds as follows. Through its sensors, the agent receives information about the state of the environment. After receiving information about the environmental state, the agent has to select an *action*. This action may cause the state of the environment to change. At this point, the agent receives its evaluative feedback in the form of a numerical reward.

A formal model commonly used to describe reinforcement learning problems is the finite Markov Decision Process (MDP). An MDP is defined by a finite set of states the environment can assume, a finite number of ac-

[1] In this thesis, the word subspace is used to refer to n-dimensional regions, in particular those of hyperrectangular shape, that lie within a confined n-dimensional space. In the literature, see e.g. (Oja, 1983), the term often refers to spaces of *lower* dimensionality than the original space.

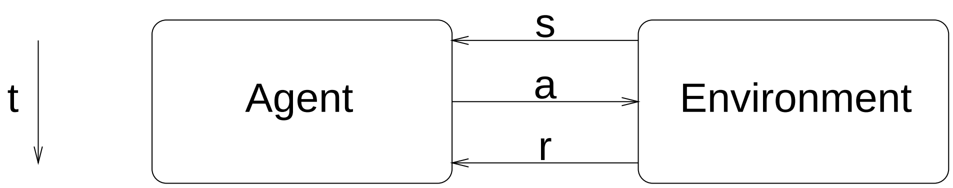


Figure 2.1: Reinforcement learning: an agent in an environment tries to obtain delayed rewards.

tions the agent can select, and a state transition function $\mathcal{P}^a_{s,s'}$ describing how the environment reacts to the actions of the agent. More formally, the function returns the probability that given an action a of the agent, the environment will transit from state s to state s':

$$\mathcal{P}^a_{s,s'} = \mathcal{P}(s_{t+1} = s' | s_t = s \cap a_t = a) \tag{2.1}$$

The process is further required to have the *Markov property*, which states that the next state is a function of only the current state and action, and depends on no other information such as previous states. Thus, in a Markov Decision Problem the state information received by the agent identifies the complete state of the environment; otherwise, the non-observable elements of the environmental state would violate the Markov property. When the information the agent receives is not guaranteed to identify the state of the environment, the problem belongs to the class of Partially Observable MDPs (POMDPs), which require different methods.

The task of the agent is to find a good *policy* determining its behavior. A policy $\pi(s, a)$ is specified by giving the probabilities of taking action a in state s. If the state transition function of equation 2.1 is known, the methods of dynamic programming (Bellman, 1957) can be used to determine the optimal policy such as value iteration or policy iteration, both of which approximate a value function over the states and use it to determine a policy. In reinforcement learning however, the state transition function is generally unknown. Thus, the state transition function can be learned, or the value function or policy can be approximated directly without using a transition function.

Although it is possible in principle to find a policy by randomly generating policies and selecting one that performs well, this approach would take a long time because evaluating the fitness of a solution requires a sufficiently large number of interactions with the environment.

Instead of varying a complete policy, the element of variation normally corresponds to a state, and often even to an action in a state. The level of adaptation of these three options respectively corresponds to policy space (where each point represents a distribution over the entire state-action space), state space (where a change affects all states which for some action have nonzero probability of transiting to it), and state-action space (where a change only affects the optimal policy in other states if the action becomes or ceases to be optimal in its state).

It is interesting to note a parallel with the experiments on sensory channel construction described in the previous section. One of the conclusions of that experiment was that it was better to evaluate complete solutions (sets of sensory channels) as opposed to using the fitness information of the lowest available level, i.e. that of the individual sensory channels. The analogy here would be that it might be better to evaluate complete policies instead of using each state's distribution of action probabilities. There are two reasons why this is not true in general. First, by varying policies instead of action probabilities, the size of the search space is increased by a factor that is exponential in the number of states, which is normally much larger than the size of a set of sensory channels in those experiments (5). But what's more important is that the elements of the solution are far less interdependent, making it easier to improve a solution by adapting its elements one at a time. The influence of a single channel on the quality of the set of channels strongly depends on the other channels within that set; for example, generating a channel that was already present in the set does not increase this quality. Changing the behavior of the agent in a single state however depends far less on the policy in other states (although the dependence is certainly important). For these reasons, most work in reinforcement learning attempts to benefit from the structural property reinforcement learning problems have of involving different states and actions.

A good way of using the division into states and actions to be able to vary policies at a finer grain size, is to approximate the *value* of each state. The value of a state s is determined by the future rewards that are to be expected given the information that s is the current state. If the lifetime of an agent is unknown, the influence of rewards far in the future needs to be discounted to avoid infinite returns. By multiplying each immediate reward R_t with a discount factor γ that decreases exponentially over time, a bounded estimate of the future rewards is obtained:

$$\sum_{t=T+1}^{t=\infty} \gamma^{t-(T+1)} R_t \tag{2.2}$$

The *value function* $V(s)$ returns the expected value of this discounted sum of future rewards for each state. The optimal value function $V^*(s)$ satisfies the *Bellman equation* (Bellman, 1957):

$$V^*(s) = \max_a \sum_{s'} \mathcal{P}^a_{ss'}[\mathcal{R}^a_{ss'} + \gamma V^*(s')] \tag{2.3}$$

The term $\mathcal{R}^a_{ss'}$ is the reward function, which gives the immediate reward given the transition from state s to s' as a result of the agent choosing action a. The equation, which is actually a system of equations, states that the optimal value of a state is the maximum over all possible actions a of the expected return of a. Given the optimal value function, an optimal policy is simply a policy which in each state s selects one of the actions with an expected return of $V^*(s)$.

Instead of approximating $V^*(s)$ with a value function $V(s)$, another possibility is to approximate the value of every action that may be selected in a state. This function is called the *action-value function*, and is denoted by $Q(s,a)$. This has the advantage that it is not necessary anymore to know the state transition function $\mathcal{P}^a_{ss'}$ in order to choose an action. A well known learning method for reinforcement learning based on this idea is called *Q-learning*, introduced in (Watkins, 1989). Q-learning is a *temporal difference* (Sutton, 1988) learning method. Temporal difference learning methods adapt estimates of whatever quantity is to be approximated to later approximations of that same quantity. With Q-learning, the later approximation is the approximated value of the immediate next state. More general temporal difference reinforcement learning methods also take later states into account, where the influence of the future decays with a parameter λ, hence their name: TD(λ) methods. The learning rule for Q-learning is as follows (Watkins, 1989):

$$Q_{t+1}(s_t, a_t) = (1-\alpha)Q_t(s_t, a_t) + \alpha(R_t + \gamma max_a Q_t(s_{t+1}, a)) \tag{2.4}$$

An interesting variation is SARSA (Rummery & Niranjan, 1994; Sutton, 1996). SARSA is similar to Q-learning, but instead of using the maximum action value of the successor state, the value of the *actual* action that is selected at the next time step is used. This makes SARSA an *on-policy* method, meaning that it behaves according to the policy it is learning.

$$Q_{t+1}(s_t, a_t) = (1 - \alpha)Q_t(s_t, a_t) + \alpha(R_t + \gamma Q_t(s_{t+1}, a_{t+1})) \qquad (2.5)$$

2.4.1 Exploration and Exploitation

Since reinforcement learning agents have to learn which actions to choose, the initial policy will normally not be optimal (otherwise there is nothing to learn!). This normally means that the quality of some actions is overestimated, whereas other actions are underestimated. The actions for which the value is overestimated will tend to be selected soon enough (sooner than desirable in fact), receive lower evaluations than expected, and their values will be adjusted accordingly. Although this is desirable, the actions for which values are underestimated are a more difficult case. If the agent acts according to a *greedy* policy, i.e. it always chooses actions with the highest estimated values, most of these actions will never be selected[2]. Since these actions may include optimal actions, the agent may never learn a good policy. Thus, it is necessary to *explore*. But since the goal of reinforcement learning is to achieve good behavior even while learning, there is a tradeoff between exploration, which improves the agent's current knowledge, and *exploitation*, which benefits from the acquired knowledge by choosing good actions.

In the experiments in this section, exploration is ensured by selecting a random action with a small probability ϵ, and the estimated optimal action with probability $1 - \epsilon$. This is a simple, and very common exploration method, and is known as the *epsilon-greedy* policy. ϵ decreases over time, and is determined as $\frac{1}{t^\alpha}$, where in the experiments $\alpha = 0.97$.

One of the most interesting exploration methods is the Interval Estimation algorithm developed by Leslie Kaelbling (Kaelbling, 1993), which uses a statistical method to determine the upper bounds of an action's value. The actions with the highest upper bounds are selected, so that values of overestimated actions will quickly decrease as more information becomes available. Actions which have not been selected so often (or recently, if statistics are decayed), have a high uncertainty and thus high upper bounds, which encourages their exploration. For an overview of exploration policies, see (Thrun, 1992).

[2] Exceptions are possible since actions with underestimated values may still have higher estimated values than the other actions

In (De Jong, 1997a), an exploration method is described that keeps an *exploration bucket* for each action. At each time step, the contents of this bucket are increased with a value composed of a term proportional to the estimation error of the action value when the action was last selected, and a term proportional to the average estimation error over all actions. Action selection is based on the sum of the estimated action value and the contents of the exploration bucket, and is followed by emptying the corresponding bucket. The method is simple, and has the property that exploration is particularly active when the approximation of the value function is or becomes inaccurate. The method is particularly suited for changing environments that are deterministic at each point in time; in stochastic environments, it will tend to explore too much. For such environments, a better approach is to consider the expected improvement in the reliability of future predictions that is gained by exploration, as done in Schmidhuber's Adaptive Curiosity method (Schmidhuber, 1991).

2.4.2 Generalization and Bias

It is a common idea that reinforcement learning is *tabula rasa* learning, i.e. that no a priori knowledge can be provided. This is a misconception; several ways exist. One way of providing a priori knowledge is to initialize the value function with an approximation based on that knowledge. Another possibility is to use *generalization.* This is the subject of the current section.

Although principled learning methods have been derived for problems with finite state-action spaces, these methods are slow for large finite state-action spaces since information has to be learned about every single combination of a state and an action. Furthermore, they cannot be directly applied to continuous state-action spaces. Generalization addresses these problems, and is a way of benefiting from similar experiences.

When generalization is used, the information gained by executing an action in a certain state is not only used to update the value of that particular state-action pair, but also the values of other state-action pairs that are assumed to be related. Thus, some sort of *bias* is introduced. For a lucid explanation of the tradeoff between errors due to bias and variance, see (Geman, Bienenstock, & Doursat, 1992). There, the mean squared error in regression problems is decomposed into a bias component and a variance component. In the paper, the corresponding terms in the formula are labeled "bias" and "variance", but it is important to keep a clear distinction between a bias itself and the component of the error caused by that bias.

A bias may be viewed as an assumption, e.g. concerning the shape of a function for instance, or concerning the likelihood of different parameter settings. This view is consistent with that of Geman et al., who discuss the crucial problem of designing a bias that is useful for a certain problems[3], and refer to 'a set of simple constraints on the architecture' as a bias. The error due to *bias* is the *error* that results when inappropriate assumptions are made.

When a bias is appropriate, using it can have very positive effects. The necessity of finding useful forms of bias in the context of generalization has been stressed by Mitchell (1980). There, bias is described as 'any basis for choosing one generalization over another, other than strict consistency with the observed training instances'. Thus, bias refers assumptions explicitly or implicitly made by a system, namely the assumptions that cause it to make the choices referred to by Mitchell.

When no assumptions are made at all, a method is unbiased. Analogous to (Gordon & desJardins, 1995), the assumptions constituting a bias can be viewed as factors influencing search. If no assumptions are made, there is no bias, and hence all states of the search space are equally likely to be visited, which amounts to random search. *Representational bias* reduces the search space by simply not considering certain parts of it; this is equivalent to the assumption that the solution is not located in those parts of the search space. *Procedural bias* on the other hand changes the *order* of the search, so that during a search of limited duration some states will more likely be encountered than others. In this case, the assumption that is made is that some states are more likely to result in a (good) solution or in more information that may lead to a (good) solution than others, and hence should be visited first.

If the assumptions that are made are consistent with the problem (the environment, in this case), it will be easier to learn appropriate behavior. A strong bias carries a risk though; if the assumptions are inconsistent with the problem, it becomes more difficult, or even impossible to learn appropriate behavior. If on the other hand one chooses to use a very general model, the bias-variance tradeoff shows: the error due to bias will be low, but errors due to variance will be high because many parameters have to be learned, which is problematic given a limited amount of training data. Examples of using knowledge of the problem to create a useful bias by introducing generalization are the famous Boxes system for learning to

[3]By its nature, a bias can not be appropriate for all problems, and thus introducing a bias always implies restricting oneself to a limited set of problems.

control a cart-and-pole (Michie & Chambers, 1968), and experiments with control problems in (Santamaría, Sutton, & Ram, 1998).

Even if no very specific domain knowledge is available for constructing a useful bias, one may introduce bias that is expected to be appropriate for the learning tasks that will be addressed. For example, Tomas Landelius (1997) argues that locality and continuity are useful biases for many reinforcement learning problems. In the following section, a generalization method is introduced which assumes locality; experiences that are close in sensor-action space are expected to have similar values. Other generalization methods include neural networks (Williams, 1990), the G-algorithm, which builds a tree by selecting bits of the inputs that are correlated to the reward (Chapman & Kaelbling, 1991), variable resolution dynamic programming (Moore, 1991), the Parti-game algorithm, which refines the regions in state space the agent visits (Moore & Atkeson, 1995), and CMAC, a linear combination of a fixed set of features (Albus, 1981; Sutton, 1996).

2.4.3 Adaptive Resolution Methods

Adaptive resolution generalization methods divide a multidimensional space into regions based on some criterion. An early implementation of an adaptive resolution method for discrete problems is the G algorithm (Chapman & Kaelbling, 1991). It uses the Student's t-test to determine which bits of a binary state variable are relevant. Andrew Moore's Parti-Game Algorithm (1995) is an adaptive resolution method for continuous reinforcement learning with very good performance, but requires the availability of a controller that moves the system between states.

The *Adaptive Subspace Method*, which is used for the concept formation experiments in this thesis, was introduced in (De Jong & Vogt, 1998). It initially considers the complete state-action space to be a single region. As experiences are obtained though interaction with the environment, this region is split into subregions recursively. In the most straightforward case, splits are parallel to the axes of the coordinate system. The decision of whether to split and in what dimension to split is based on some *split criterion*, which depends on the application of the method. The recursive splitting process terminates when no region is left for which distinctions should be introduced according to the split criterion. An important point is that a split only introduces a distinction in a *single* dimension. Thus, a split only increases the number of regions with 1. Because of this property, the algorithm is economic in the number of states it generates. In k-dimensional versions of the two-dimensional quad-tree (Finkel & Bentley, 1974) and the three-dimensional oct-tree (Jackins & Tanimoto, 1980), which have been developed to represent spatial occupancies, each split increases the number of regions with $2^k - 1$. Splits are made in the middle of existing intervals. This choice is based on a trade-off against computation; whereas it would be possible to select a split location that better fits the data, for instance by considering a split between every two data points, this would involve more comparisons, and would complicate the algorithm. Furthermore, if increased precision is required, this can always be obtained by introducing one or more extra splits. When operating in a k-dimensional space, the tree generated by adaptive subspace methods is called a k-d tree and has efficient methods for storage and retrieval (Bentley, 1975) and finding nearest neighbors (Friedman, Bentley, & Finkel, 1977). A k-d tree is a binary tree. Its root represents the whole k-dimensional space, and each node splits a region of the space into two subregions. Therefore, the leaves of the tree are non-overlapping regions of the space, and their union equals the complete space again.

To apply the adaptive subspace method to reinforcement learning, a useful split criterion is the difference in distribution of the values of the experiences in the two potential subregions, where the value of an experience $(s_t, a_t, r_t, s_{t+1}, a_{t+1})$ is based on SARSA and defined as follows:

$$V = R_t + \gamma Q(s_{t+1}, a_{t+1}) \tag{2.6}$$

An adaptive resolution generalization method similar to the adaptive subspace method is Andrew McCallum's U-Tree algorithm (McCallum, 1996), which also constructs a k-d tree. The U-Tree algorithm recursively partitions a discrete state-action space based on the cumulative distribution of future discounted rewards of the experiences in a region. It employs *fringes*, provisional distinctions applied to a short term memory of past experiences to store information at higher resolution than the current one, allowing to determine which splits are useful.

The U-Tree algorithm uses a Utile Distinction Test (McCallum, 1996) to determine a history of experiences that allows for accurate prediction of the values of experiences in a region of the state-action space. This principle augments the state information the agent receives by taking into account the extra information of previous experiences when necessary, and can overcome some degree of partial observability of the environment. On the other hand, if the expected value of actions is similar within some region of the state-action space, this region is represented by a single node of the tree, thus accomplishing generalization.

The U-Tree method uses the Kolmogorov-Smirnov Test to determine whether the distribution of values in two adjacent regions of the state-action space is constant or not, and in the latter case decides to separate the two regions. Although this statistical test determines *whether* these values are different, it does not tell *how* different they are; thus, possible splits for which the difference in expected value is large cannot be distinguished from those for which it is marginally small. In the discrete spaces for which the U-Tree method was designed this is not necessarily a problem, since the number of possible distinctions can be finite. However, the criterion can not be directly applied to continuous spaces. There, gradual changes in the value function can cause split criteria based on distribution difference to keep splitting indefinitely, reacting to minute changes in the approximated values of different regions. The idea of taking the importance of the difference into account, for instance by using the *area* between the distributions as done here, solves this problem.

Some differences between the U-Tree method and the adaptive subspace method are that the former has variable length history which allows it to overcome a certain form of hidden state, and that it uses fringes, extensions of the split hierarchy that collect information about candidate distinctions. The latter simultaneously performs state generalization and action generalization, and is suited for continuous state-action spaces. It takes into account the importance of the distinction between regions considered for a split by using a statistical test based on the area between the two distributions.

The result of concept formation using the adaptive subspace method is a tree representing a subdivision of the complete state-action space into regions. In general, nodes representing sensor distinctions and action distinctions can occur at any place in the tree. Here however, the splits of these types are kept separate such that the state space distinctions always come first in the tree. In other words, once an action distinction is encountered, it may be assumed only action distinction will follow. This organization has the advantage that the resulting tree will contain a set of intermediate nodes that correspond to aggregated states. These states are basic forms of *situations*, since they imply information about the future behavior of the environment. These concepts are formed based on the criterion of whether the *value* of different experiences is different. A more general notion of *situation concepts* is obtained by not only considering this value, but also (or alternatively) other aspects of the expected future state of the environment. Furthermore, the interaction history taken into account, here the current experience, can be extended to form arbitrarily long interaction histories, as in the case of the U-Tree method. Situation concepts are described in more detail below. The point of describing this more general version of situation concepts is that it may clarify how similar principles could play a role in human concept formation.

2.5 Situation Concepts

A *situation concept*, introduced in (De Jong, 1999), is a pattern in the interaction history between an agent and its environment with the property that knowing to which situation concept the actual history of interaction corresponds, allows the agent to predict some aspect of the future. It represents an agent specific view on the complete state of the environment. A common aspect of situations is their predictive power. The computational definition of situation concepts given here is based on this idea. The principles behind situation concepts may play a role in human concepts representing situations. The situations people distinguish can be characterized by a huge variety of properties, but they are not random.

As an example, consider the advent of a thunder storm. Both seeing a flash of lightning and hearing a roaring sound of thunder are indicators that in a few moments, it may start to rain. Thus, these observations may be grouped together to form a situation which has the property that a shower is likely to arrive within short, whereas this possible future event will be less likely in the case of a bright blue sky. In this example, the situation is based on observations in the recent past, and the prediction concerns future observations. Actions of the agent or evaluative feedback played no role.

Other examples of situation concepts are the presence of sunshine after rain (which announces the possible advent of a rainbow), the presence of ice (which indicate that if you are incautious, you may fall), and the condition of being in a restaurant and having ordered lobster (which significantly increases the probability that one will be eating lobster soon). When a situation occurs, i.e. a situation concept applies, this yields information about the likelihood of the environment being in certain state(s). As the examples make clear, the predictive aspect of a situation may lie in the environmental states the environment will adopt in the near future, but also in the effects of the actions the agent can take.

Situation concepts can be developed by agents by observing regularities in the interaction history with the environment. The interaction history is a sequence of inputs from the environment, actions of the agent, and subsequent evaluative feedback. A particular interaction history can be view as a point in *interaction space*, the space representing all possible interaction histories, where each dimension represents a sensor value, action component, or evaluative feedback at a particular distance in time relative to the current moment.

In principle, all adaptive resolution methods can be used as a basis for situation concept formation, e.g. the U-Tree method (McCallum, 1996) and the adaptive subspace method, since they build discrete states that represent particular situations. Another method that can be used is the *schema* mechanism described in (Drescher, 1991), where the context and an agent's action are used to predict the result of the action. In that framework, a context is specified as a set of conditions and can be viewed as an instantiation of situation concepts, since it defines a subset of the possible histories of interaction (viz. the current input) and has predictive value. However, most other generalization methods for reinforcement learning are not suitable, either because the representation of the interaction space is not adapted to the learning problem (e.g. plain discretization, CMAC (Albus, 1981)), or because no distinct representations are constructed (e.g. neural networks).

In (Zwaan & Radvansky, 1998), an overview of literature on *situation models* is given. In that article, a situation is a *particular* situation that occurred at some point, whereas situation concepts as they are used here by definition have some *general* validity that is responsible for their predictive power. In the article, Zwaan et al. argue that the time has come for researchers to address *multidimensionality* in situation models. Although the situation models in that article are more complex than the model of situation concepts, the latter are defined with much more detail, making implementation possible, and describes *how* concepts may be generated that cover multiple dimensions.

2.6 Experimental Investigation of Concept Formation

In this section, an experiment will be described that makes clear how similar situation concepts can be formed by individual agents with differing experiences. In contrast with earlier reports of this research, the problem description will be kept abstract here in order to emphasize the general nature of the mechanisms.

situation	action 0	action 1	action 2
-1	high	high	high
0	high	low	low
1	low	high	low
2	low	low	high

Table 2.2: Success of the different actions given the situation.

2.6.1 Basic Setup of the Experiments

In the experiments, the adaptive subspace method will be used to form situation concepts. The interaction history will consist of the current input from the environment. The situation concepts are chosen such as to allow prediction of the subsequent reward given an action the agent may choose. This is achieved by using the degree of difference in distribution of the values in a region as the split criterion.

The experiment concerns a number of agents that are present in the same environment. Although a shared environment is viewed here as an important force guiding the development of similar concepts, the experiences of different agents in a shared environment are never *exactly* the same. In the experiments in chapter 3, the situation sensor is masked 10% of the time. This causes uncertainty in the knowledge of agents about the environment that can only be overcome by means of communication.

When perception of the agents is local, or when it is global but incomplete, different agents will receive different sensor inputs. Furthermore, these differences in perception and stochasticity in the action selection mechanisms of different agents are two factors that produce different behavior in those agents, which in turn influences the evaluative feedback that agents may receive.

In the experimental environment, a number of different *situations* may occur. The cause of these situations is not important here; what matters is how the agents might detect these situations, and what predictive potential the detection of a situation brings.

The sensor input has $d_i = 3$ dimensions for each of which the agent receives a scalar value. The agent can then select a $d_a = 2$-dimensional action, which yields a scalar evaluation. One of the perceptual dimensions will indicate the situation. Thus, a useful conceptual system will allow the agent to distinguish between the possible values this sensor can take, and any other distinctions may be harmless but do not have any positive effect.

s_t			a_t		r_t	s_{t+1}			a_{t+1}	
-1	5	0	-1	1	1.361	-1	4	1	-1	0
-1	4	1	-1	0	1.214	-1	3	0	1	1
-1	3	0	1	1	1.174	-1	4	1	-1	0
-1	4	1	-1	0	0.736	0	3	0	1	1
0	3	0	1	1	0.001	0	4	1	-1	2
0	4	1	-1	2	0.173	0	3	2	0	0
0	3	2	0	0	1.008	0	3	0	0	2
0	3	0	0	2	-0.293	0	3	2	1	2
0	3	2	1	2	0.200	0	4	2	1	0
0	4	2	1	0	0.776	0	5	0	-1	0

Table 2.3: Experiences of the agent in a region that is considered for splitting.

In some situations, situation -1 in this experiment, the actions will determine the behavior of the agent, but not the evaluations it receives. For other situations however, there is only a single appropriate action which will be evaluated highly; all of the other actions will return a low value, as shown in table 2.2. The high and low rewards are distributed normally around the values of zero and one.

The agent is not provided with any knowledge about the structure of its inputs. Therefore, it doesn't know which of the inputs indicates situations in the environment. In fact, it doesn't even know that this information is contained in a single dimension, but will look for general k-dimensional subspaces in sensor space.

Interpretation of the Experiment

First, it should be stressed that the experiment is *not* intended to model an existing communication system. However, as the above description is rather abstract, it may be useful for the reader to relate the experiment to a known problem. The communication system that served as inspiration for the design of this experiment is the alarm call system of Vervet monkeys, see e.g. (Hauser, 1997).

Convincing experiments have shown that these animals have a warning system with specific calls for different kinds of predators (Seyfarth, Cheney, & Marler, 1980): birds of prey, large mammals, and snakes. In the experiments, the calls produced by these monkeys were played back using a tape recorder. Following the playback of these signals, the monkeys responded as if the corresponding predator were present, by displaying a corresponding flight behavior. This demonstrates the principle of how situation concepts can be useful when communicated. When a monkey does not detect the threat of an approaching predator, successful interpretation of the alarm calls produced by other community members may make it aware of its perilous situation. Put in abstract terms, the signals an agent receives from the other agents allow it to deduce that its situation is different from what it had observed using ordinary perception.

The three sensors of the agents correspond to a predator sensor, a sensor indicating the agent's own horizontal position, and a sensor indicating the agent's vertical position. Actions correspond to moving zero or one step to the left or right, and selecting a new vertical position. In the standard situation, no predator is present, and all actions yield a high reward. Periodically however, a random predator arrives. For each predator, there is only a single safe vertical position. However, if the predator is too far from an agent, determined by their horizontal locations, the agent's predator sensor will not detect it. In the standard experiment, this happens in 10% of the cases. The visibility of a new predator is determined at its time of arrival for each agent and remains so during its existence, even if the horizontal position changes again. Due to this partial perception, the sensors do not provide sufficient information to produce optimal behavior, and hence optimal behavior indicates that communication has been used.

2.6.2 Operation of the Adaptive Subspace Method

The adaptive subspace method for concept formation will now be described in detail. To visualize the workings of the mechanism, the description will make use of an example. Table 2.3 shows some (10) of the 250 experiences that have been obtained in a particular region of the sensor-action space. Each experience consists of the three sensor values and two action values at the previous time step, the reward following the action, and the sensor and action values at the next time step. The first of the three sensors indicates the situation, but this is unknown to the agent, who has to discover this pattern by looking for relations between the values of each of the sensors and the effects of its actions.

The experiences are ordered chronologically, and hence the right half of each experience equals the left half of its successor. Using formula 2.6, the values of the experiences can be calculated. Since the actions of the agent do not affect its future possibilities for obtaining rewards in this particular problem, the history factor γ can be set to zero and the value equals r_t, shown in the middle column of the table. This information is the ground on which the decision of whether to split, and in which dimension, will be taken.

The decision of whether to split is considered for the region in which the current experience is located with a certain frequency. It is made as follows. For every dimension of the sensor-action space, i.e. for every sensory channel and action component, a split in the middle of the region is considered, which would result into two subregions, both half the size of the original region. Since the initial shape of the space is a box and each split divides a box-shaped region into two smaller boxes, all regions will be boxes (hyperrectangles).

The actual split criterion is a parameter of the adaptive subspace method. Here, this criterion is chosen to measure whether there is a difference between the distributions of the values of the experiences in the two subregions. The test that will be used here is related to the *area* between the empirical distributions, and calculates the integrated squared distance between the distributions.

Figure 2.2 shows the values of the complete set of experiences obtained for the region we are examining, which includes the data in table 2.3. Each point is a six-dimensional vector, corresponding to the first six entries of a row in the matrix of which the table is a sample. Since six-dimensional vectors are difficult to visualize the most important dimension, the one representing the value, has been combined with each of the other dimensions, yielding five different projections of the same data. The graphs also show the plane that would separate the two subregions if a split would be decided for.

For each graph in figure 2.2, figure 2.3 shows a corresponding graph. Whereas 2.2 shows the values of the experiences, 2.3 shows the distributions of those values. The two distributions concern the points to the left and to the right of the splitting plane. The points lying *on* the splitting plane in the first graph fall within the right region, which is called the upper subregion because its coordinates in that dimension are higher than those in the region to the left.

A common aspect in all of the graphs of figure 2.3 is that a group of experiences with values around zero causes the distribution to rise somewhat, then no values are encountered in the middle region, leaving the distribution at a plateau, and finally there is another cluster of experiences with values around 1. This is a consequence of the structure of the environment.

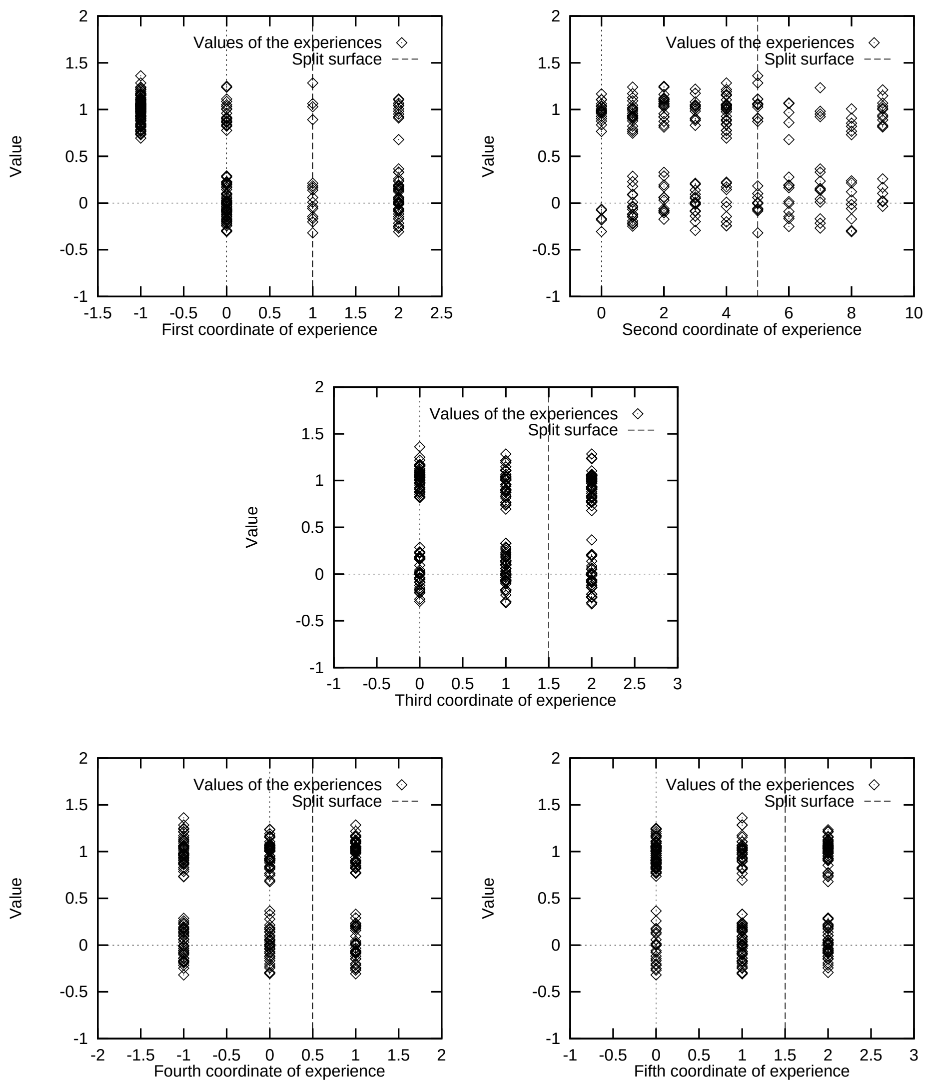


Figure 2.2: Projection of the values of experiences (3 sensor and 2 action dimensions) in a region of the sensor-action space onto planes perpendicular to the axes. All of the projections show the same set of experiences, but in each graph their values are plotted against a different sensor- or action-dimension.

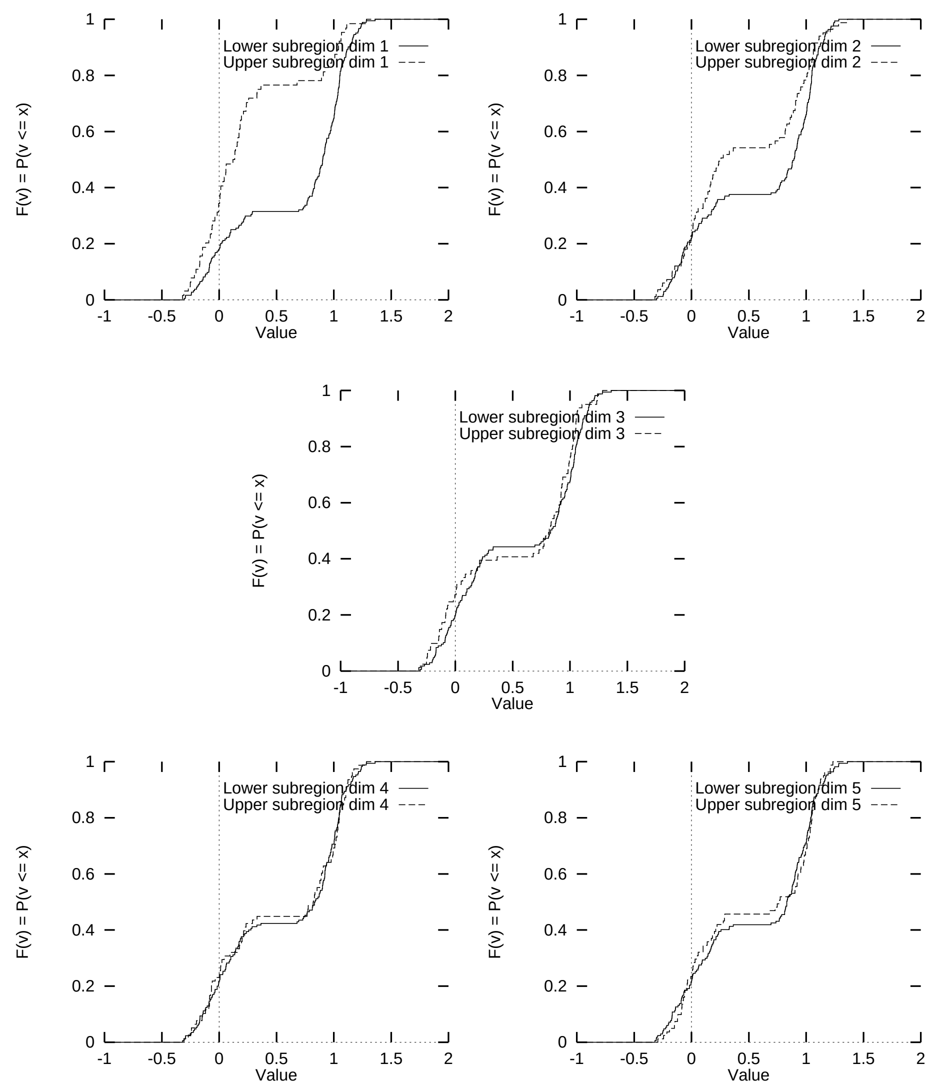


Figure 2.3: Empirical distributions of the values of experiences in the two regions that would remain after splitting in the first dimension.

As described earlier in this section, the rewards are distributed normally around zero and one, which explains the shape of the distributions.

The decision criterion is determined by the area between the distributions. In the first dimension, this area is larger than in any of the other graphs. It indicates that for the given set of experiences, the lower subregion contained more experiences with high evaluations than the upper subregion, since its distribution function rises most around high values of the horizontal coordinate. Similarly, the graph in the middle shows that there is little difference between two sides of the splitting plane when the third sensor value is considered; hence, it would not be very useful to split in that dimension. The integrated squared distance between the distributions is 9.17% of the maximum squared area, which was higher than the threshold value of 2%, and hence the final result of the procedure was that the region was split in the first sensor dimension.

Results of the Splitting Process

The initial representation of the state-action space is the root of a tree. With each split, two new nodes are added to the tree, one for each new subregion. These nodes are the two children of the node that represented the region before splitting. Thus, the finest regions are always represented by leaves of the tree.

Figure 2.4 shows the tree of an agent in the experiment after concept formation has stabilized, i.e. when the tree is not growing anymore. As the figure shows, the splits stored in the nodes are not ordered randomly; all sensor distinctions occur *before* action distinctions. This is required to ensure that each situation is represented by single node; if action distinctions were allowed to precede perception distinctions, the nodes corresponding to different action possibilities would need to be considered. This property is important when communication starts to play a role, as will be see in the next chapter.

To maintain the property that sensor splits precede action splits, a node that has already made one or more action split and is about to be split in a perceptual dimension takes the subtree starting at the rightmost perceptual distinction and splits it, attaching a copy of that subtree to both of the resulting nodes. This is necessary in order to maintain all action distinctions for the region. The experiences stored in the subtrees are redistributed over the two subtrees according to the new perceptual split. This is done in order to maintain correspondence between the regions represented by the nodes and the experiences stored in them.

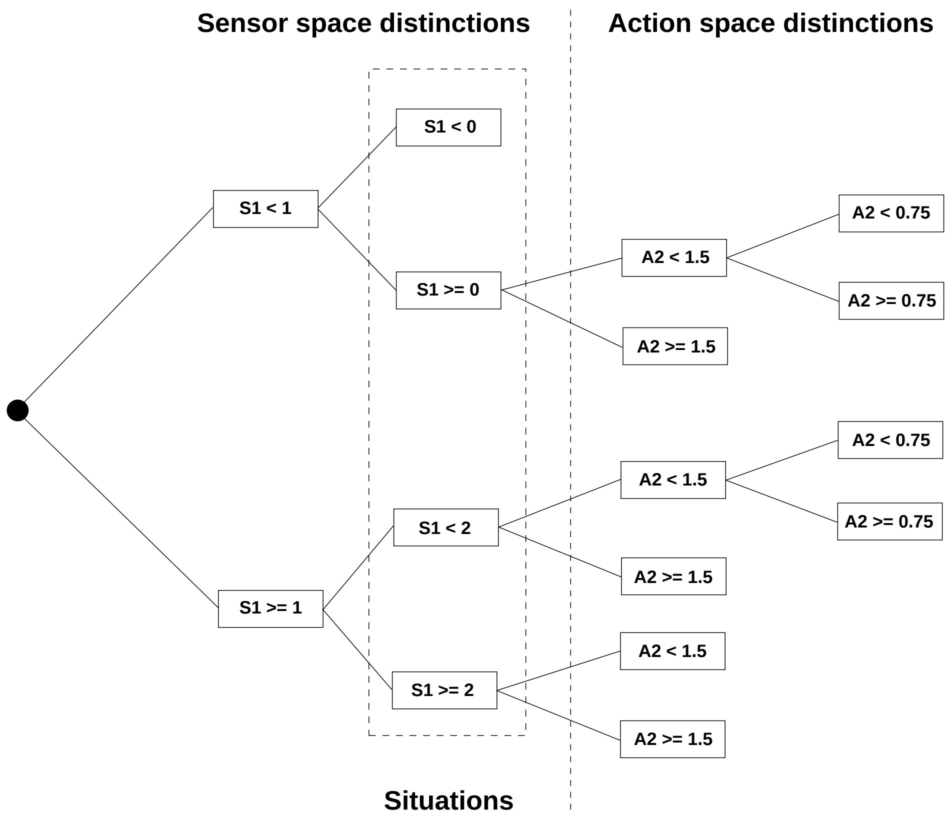


Figure 2.4: The resulting tree for agent 1 after forming situation concepts. Boxes referring to Sx constrain the space in sensor dimension x, while boxes referring to Ay constrain the action of the sensor-action space in action dimension y. The agent distinguishes four situations; they are located within the dotted frame.

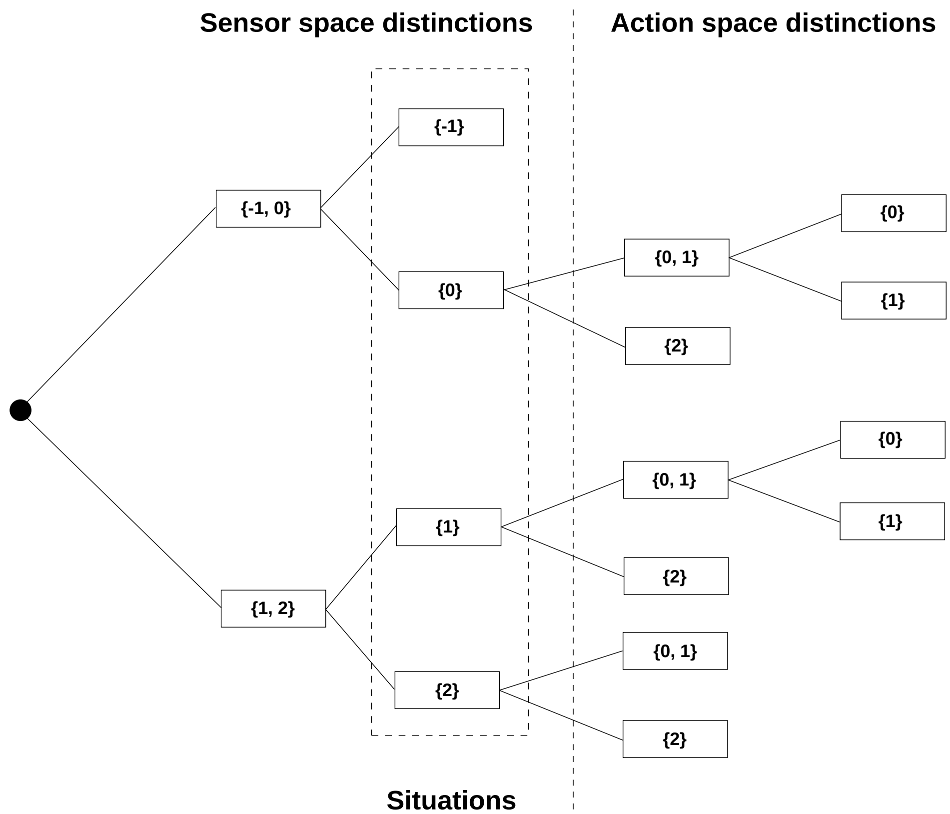


Figure 2.5: Intervals of the tree for agent 1. Each node corresponds to a node in figure 2.4, and shows the interval of a sensor or action determined by that node.

The question of which actions the agent should distinguish between depends on the situation. As table 2.2 showed, situations 0, 1 and 2 each have a single good action, viz. the action with the same index. In situation -1 however, the evaluation the agent receives does not depend on the action. In the tree, sensor distinctions in dimension i are shown as S_i, and action distinctions as A_i. As the tree shows, the agent benefits from this property by not making any distinction regarding its actions in situation -1. For each of the other situations, it has created distinctions that allow it to differentiate the successful action from the other actions, but not any other distinctions, which would indeed be useless. This is illustrated in figure 2.5, which represents the constraints in the tree of figure 2.4 in a different format; here, each node shows the set of possible values given the corresponding to the node and its ancestors.

2.6.3 Analyzing the Result of Concept Formation

A principled way of investigating is in terms of necessary and sufficient distinctions. Whenever a distinction is necessary in order to distinguish one situation from another, will be called a *necessary* distinction. If all situations can be distinguished from each other, the set of distinctions is called *sufficient*. Concept formation should at least yield a sufficient set of distinctions, and should not contain too many unnecessary distinctions. These notions are expressed by the concepts of *distinctiveness* and *parsimony*. Distinctiveness expresses whether concept formation has made all the necessary distinctions, and thus resulted in a sufficient set of distinctions. Parsimony indicates whether no more distinctions than necessary have been made; unnecessary distinctions slow down learning and decrease the chances of building up useful communication. Distinctiveness and parsimony will be examined in an example. After that, both concepts will be captured formally in two corresponding *measures*. First however, a matrix notation describing the results of concept formation will be introduced.

Matrix Notation for Concept Formation

A matrix notation describing the results of concept formation will now be introduced. This representation will not only serve to describe the relationships between the private meanings or concepts of an agent and public referents, but will later prove to be useful in describing systems of communication. At this point, it is important to clarify the use of *referents*. An aspect of the viewpoint that lead to this research is that it is not necessary to assume that referents in the real world can be accurately defined. One might be led to think this property should apply to the experimental world as well. However, there is an important difference between the simulated world of experiments and the real world. In the real world, it is impossible to measure the complete state of the world accurately and monitor it over time. In the simulated world however, all information that one might wish to know can be extracted. Since in addition the criteria used by agents to form concepts are known, i.e. they try to distinguish between situations that require different actions, the ideal conceptual system that allows agents to produce successful behavior is known exactly. Thus, knowledge of the simulation world can be used to determine the exact referents about which the agents want to talk. Given the referents, the relation between meanings and referents that define a conceptual system can be expressed in matrix notation.

Meanings will be represented as μ_i, where i is an index. Likewise, referents (situations) are referred to with ρ_i. The relationship between an agent's conceptual system and the world can be investigated by measuring co-occurrence, i.e. how often meanings and referents occur together. The co-occurrence between a meaning and a referent determines the conditional probabilities of the referent given the meaning ($P(\rho|\mu)$) and the meaning given the referent ($P(\mu|\rho)$). If both of these probabilities are 1 for a combination of μ and ρ, then the agent activates meaning μ if and only if referent ρ is present. Examples of the two matrices are given in table 2.4.

Two Criteria for Success: Distinctiveness and Parsimony

The matrix representation provides all information necessary to calculate whether an agent has made sufficient and only necessary distinctions. Perfect distinctiveness has been achieved if for each meaning μ there is *at most* one referent ρ such that $P(\rho|\mu) > 0$. Such meanings allow the agent to determine unambiguously, noise permitting, whether referent ρ is present in the environment, i.e. whether the agent's situation is ρ. Furthermore, if every state of the environment corresponds to some referent, there has to be *at least* one referent to which a meaning corresponds. Thus, by combining these properties, it follows that sufficient distinctions have been made if for each meaning μ there is exactly one referent ρ such that $P(\rho|\mu) > 0$. Finally, since both referents and meanings cover the complete state-actions state, the sum of $P(\rho|\mu)$ over all referents for a given meaning μ equals 1. Thus, a conceptual systems makes sufficient distinctions if for each meaning μ there is precisely one referent ρ for which $P(\rho|\mu)$ is one, while this

	ρ_1	ρ_2	ρ_3	ρ_4
μ_1	1	0	0	0
μ_2	0	1	0	0
μ_3	0	0	1	0
μ_4	0	0	0	1

	μ_1	μ_2	μ_3	μ_4
ρ_1	1	0	0	0
ρ_2	0	1	0	0
ρ_3	0	0	1	0
ρ_4	0	0	0	1

Table 2.4: Conditional probabilities of referent given meaning $P(\rho|\mu)$ (left) and meaning given referent $P(\mu|\rho)$ (right). Since all rows contain a single one, both distinctiveness and parsimony have been achieved.

probability is zero for all other referents.

If only necessary distinctions have been made, then the number of meanings that can occur given the presence of a situation ρ is as small as possible. Ideally, $P(\mu|\rho)$ will be one for a single meaning μ and zero for all other meanings. Thus, whether perfect parsimony has been achieved can be determined from the matrix $P(\mu|\rho)$ in the same way distinctiveness can be determined from $P(\rho|\mu)$, i.e. by seeing whether each row contains a single one and has zeroes at the remaining entries. It is important to notice that although any good method for situation concept formation should be able to achieve high distinctiveness if the problem allows it, this does not hold for parsimony; whereas the structure of referents is in principle unrestricted, and may take arbitrary shapes in sensor-action space, the structure of meanings is restricted by the representational format of the concept formation method. To give an example, a referent might present itself to the agent such that one of its sensor values can be in either of two separate intervals. In this case, the adaptive subspace method would not be able to represent the referent in a single contiguous region, but would need at least two meanings to represent the referent, thus diminishing parsimony. Nonetheless, given an environment and a representation format, the objective of the agent will always be to achieve high parsimony, even if the maximum value it may obtain is bounded.

The situations in the tree of the experiment (see figure 2.4) all correspond to a single value of the sensor S1, and thus they all identify a single situation. This is readily seen by constructing the matrices $P(\rho|\mu)$ and $P(\mu|\rho)$, as shown in table 2.4, where $\rho_1 \ldots \rho_4$ represent the four possible situations of the environment and $\mu_1 \ldots \mu_4$ are the four meanings constructed by the agent in the form of nodes in the tree.

The *Distinctiveness* and *Parsimony* Measures

The concepts of distinctiveness and parsimony can both be captured in *measures.* All measures in this work take values between zero and one. When distinctiveness and parsimony are perfect, their corresponding measures should yield a value of one. The advantage of quantitative measures over binary ones is that when either of the criteria is not completely satisfied, they provide an evaluation of the *degree* to which they have are satisfied.

Intuitively, distinctiveness expresses to what degree a meaning identifies the referent, and parsimony expresses to what degree a referent gives rise to a unique meaning. Using the *entropy* concept from information theory, this intuition can be formalized in a principled way. For distinctiveness, what we would like to measure is the amount of information that is gained about the current referent by knowing which meaning applies. Likewise, parsimony refers to the extent to which the referent determines which is the current meaning. These can almost directly be calculated from the entropy.

Entropy can be viewed as the uncertainty about a set of elements; a high entropy means a chaotic or almost random distribution, whereas a low entropy indicates order. In information theory, entropy is defined as follows. Let X be a random variable with a set of possible outcomes $\mathcal{A}_X = \{a_1, \ldots, a_n\}$, having probabilities $\mathcal{P}_X = \{P_1, \ldots, P_n\}$, with $P(x = a_i) = P_i$, $P_i \leq 0$ and $\sum_{i=1}^{n} P_i = 1$. Then according to (Shannon, 1948), the entropy H of X is defined by:

$$H \equiv -\sum_{i=1}^{n} P_i \log P_i, \tag{2.7}$$

with the convention for $P_i = 0$ that $-0 \cdot \log 0 \equiv 0$. The information theoretic entropy used here should not be confused with entropy in physical systems, as in the second law of thermodynamics, although there is a relation in that both forms of entropy measure disorder.

What we want to measure is the degree to which knowledge of one entity X (meaning or referent) decreases the amount of uncertainty about the other Y. This can be calculated by taking the difference between the amount of uncertainty in Y before and after observing the entity X. This quantity is known as the *information gain.*

Since in the ideal case all uncertainty is removed, the maximum decrease in uncertainty equals the initial amount of uncertainty. Dividing the outcome by this maximum yields a measure between zero and one. Thus, the distinctiveness $\mathsf{dist}(\mathcal{A})$ of an agent $\mathcal{A}$ can be defined for meanings $\mu_1 \dots \mu_{n_m}$ and referents $\rho_1 \dots \rho_{n_r}$ as follows:

$$H(\rho|\mu_i) = \sum_{j=1}^{n_r} -P(\rho_j|\mu_i)\log P(\rho_j|\mu_i) \tag{2.8}$$

$$\mathsf{dist}(\mu_i) = \frac{H(\rho) - H(\rho|\mu_i)}{H(\rho)} = 1 - \frac{H(\rho|\mu_i)}{H(\rho)} \tag{2.9}$$

$$\mathsf{dist}(\mathcal{A}) = \frac{\sum_{i=1}^{n_m} \mathsf{dist}(\mu_i)}{n_m}, \tag{2.10}$$

where the initial uncertainty $H(\rho)$ equals $\log n_r$. Based on the $n_r \mathrm{x} n_m$ matrix $P(\mu|\rho)$, parsimony $\mathsf{pars}(\mathcal{A})$ can be defined likewise for referents $\rho_1 \dots \rho_{n_r}$ and meanings $\mu_1 \dots \mu_{n_m}$:

$$H(\mu|\rho_i) = \sum_{j=1}^{n_m} -P(\mu_j|\rho_i)\log P(\mu_j|\rho_i) \tag{2.11}$$

$$\mathsf{pars}(\rho_i) = 1 - \frac{H(\mu|\rho_i)}{H(\mu)} \tag{2.12}$$

$$\mathsf{pars}(\mathcal{A}) = \frac{\sum_{i=1}^{n_r} \mathsf{pars}(\rho_i)}{n_r}, \tag{2.13}$$

where $H(\mu) = \log n_m$. If there is only a single meaning, parsimony is defined to equal one.

In the following, the use of the newly defined measures will be examined in practice. Figure 2.6 shows the evolution of both measure over time. Furthermore, the number of meanings is displayed, scaled onto $[0, 1]$. The initial number of meanings is one, as is always the case when the adaptive subspace method is used, since the root of its empty tree represents the complete space. As is visible in the graph, the number of meanings increases three times, i.e. three splits are introduced in the sensor dimensions, yielding a total of four meanings. The meanings are equivalent to those earlier depicted (figures 2.4 and 2.5).

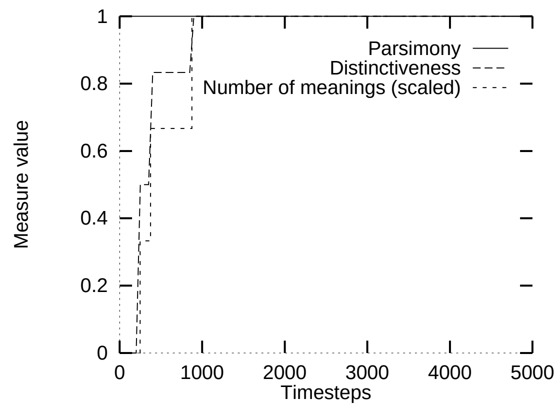


Figure 2.6: Distinctiveness and parsimony as a function of time.

The distinctiveness measure is equal to zero initially. This implies that knowledge of the meaning does not yield any information about the referent; since there is only a single meaning, this meaning was already known beforehand, and so observing that that meaning is the current meaning does not provide any extra information. With every subsequent split, distinctiveness increases. This implies that only necessary splits are made[4]. When splitting has stabilized, just before time step 1000, the distinctiveness measure increases to one, indicating that sufficient splits have been made, and thus complete distinctiveness is the result.

The parsimony measure equals one from the very beginning. It is one initially because there is a single meaning, hence no uncertainty about what the current meaning is, and thus the zero decrease in uncertainty after learning what the current referent is equals the maximum decrease in uncertainty, which by definition yields perfect parsimony. Parsimony does not decrease because of any of the splits. This is in line with the earlier observation that no unnecessary splits are made.

We now consider a case where the threshold that determines whether a difference is sufficient to cause a split is set lower than necessary (0.1%). The noise level in the feedback remains the same ($\sigma = 0.15$). Figure 2.7 shows the results. The distinctiveness measure behaves similarly. After the first few splits, it reaches its maximum of one and remains at that value. The only difference is that it takes the agent less time to reach complete distinctiveness. The fact that splits take place earlier is a direct consequence of lowering the split threshold; what's interesting is that the

[4]Although the implication holds for this problem, it does not hold in general.

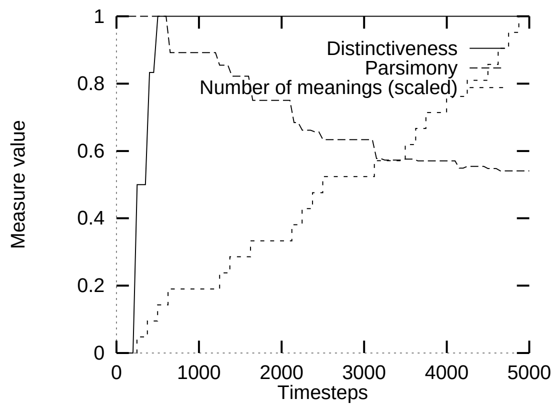


Figure 2.7: Distinctiveness and parsimony as a function of time.

split procedure still selects the right distinctions in the beginning.

Parsimony however behaves rather differently. Although remaining one during the first few splits, showing that indeed these first splits were necessary, parsimony then starts to decrease. The large number of splits caused by the low split threshold can readily seen by inspecting the parsimony measure.

An interesting phenomenon is that parsimony decreases, but not monotonically. The decreases are caused by unnecessary splits. When some region of the sensor-action space is split in a dimension in which the referents have no distinctions, then parsimony will drop, since the knowledge of the referent does not enable one to distinguish between the two new meanings, and thus uncertainty has increased. Such a distinction is superfluous and should cause the parsimony measure to decrease, as indeed it does.

At several points however (t = 3500 and t = 4250), *increases* in parsimony can be observed. Although the increases are very small, the effect is counterintuitive, and calls for an explanation. The effect has two manifestations, one causing substantial increases in parsimony, the other only small increases.

The larger increases are explained as follows. If the region of some meaning corresponds to multiple referents, a new split can yield two regions, both of which correspond to a unique referent. This increases the number of possible meanings associated with both of these referents without increasing the uncertainty $H(\mu|\rho)$. The increase in the number of possible meanings increases the initial uncertainty, which has a diminishing influence on the scaling fraction in formula 2.12. The effect is visualized in

$P(\mu\|\rho)$	μ_1	μ_2	μ_3
ρ_1	0.5	0.25	0.25
ρ_2	0.5	0.25	0.25
ρ_3	0.5	0.25	0.25
ρ_4	0.5	0.25	0.25

$P(\mu\|\rho)$	μ_1	μ_2	μ_3	μ_4
ρ_1	0	0.5	0.25	0.25
ρ_2	0	0.5	0.25	0.25
ρ_3	0.5	0	0.25	0.25
ρ_4	0.5	0	0.25	0.25

Table 2.5: A split that increases parsimony. While the amount of uncertainty remains constant, the a priori uncertainty, which depends solely on the number of meanings, increases. The relative uncertainty decreases, so that parsimony increases from $1 - \frac{1.5}{\log 3} \approx 0.0536054$ to $1 - \frac{1.5}{\log 4} = 0.25$.

table 2.5. Although it has been observed in other experiments, it does not explain the increases in parsimony here, since the splits did not distinguish between different referents; indeed, this would not be possible, since complete distinctiveness has already been reached just after the beginning of the experiment.

The above makes clear that extra distinctions can increase parsimony when $H(\mu|\rho_i)$ remains constant while $H(\mu)$ increases. Apart from this effect, smaller increases in parsimony are explained by a different but similar phenomenon. Introducing an unnecessary distinction in a meaning with a low contribution to $H(\mu|\rho_i)$ will increase the uncertainty in the numerator $H(\mu|\rho_i)$, but this increase may be smaller than the increase of the a priori uncertainty in the denominator $H(\mu)$. Thus, the net effect is that the fraction (i.e. the relative residual entropy) becomes smaller, and thus parsimony increases as well. This explains the increase in parsimony at time steps 3500 and 4250.

The Concept Fidelity Measure

The aim of concept formation is to achieve a reliable process of encoding a referent into a meaning and decoding a meaning back into a referent again. Distinctiveness measures whether every meaning can identify a referent, while parsimony measures to what extent every referent identifies a meaning. A slightly different question is the combined goal of developing meanings such that encoding a referent into a meaning and decoding it back into a referent again yields the referent that was encoded. In this question, the occurrence probabilities of meanings and referents play a role. This difference is illustrated in the following.

To some extent, the distinctiveness measure already indicates whether

$P(\mu\|\rho)$	μ_1	μ_2	μ_3	μ_4	μ_5
ρ_1	1	0	0	0	0
ρ_2	0	1	0	0	0
ρ_3	0	0	0.99	0	0.01
ρ_4	0	0	0	0.99	0.01

Table 2.6: An example where distinctiveness yields too high an estimate of concept fidelity. This table shows the conditional probabilities of obtaining a meaning μ given the presence of a referent ρ.

$P(\rho\|\mu)$	ρ_1	ρ_2	ρ_3	ρ_4
μ_1	1	0	0	0
μ_2	0	1	0	0
μ_3	0	0	1	0
μ_4	0	0	0	1
μ_5	0	0	0.5	0.5

Table 2.7: Example where distinctiveness yields too high an estimate of concept fidelity (continued). Although the uncertainty in meaning μ_5 is large, its probability of occurring is low, limiting the effect of the uncertainty.

the right referent is likely to be found; if the distinctiveness is perfect, then no matter what the parsimony is (i.e. no matter how many different meanings may be associated with the referents), the referent will always be determined correctly. However, if some of the referents are not associated with a single meaning but with multiple meanings, then distinctiveness may give too negative a picture. This is the case when some of those meanings have a very low probability of occurring and a low distinctiveness; for an example of such a conceptual system, see tables 2.6 and 2.7. Also, the picture can be too optimistic, e.g. when not all referents are represented by a meaning, although the concept formation method that will be investigated guarantees that every part of the space is covered by a meaning, which prevents this.

The measure that is described in this paragraph evaluates to what extent the aim of reliable encoding and decoding of referents is achieved. This measure will be called *concept fidelity*, since it expresses the fidelity with which a conceptual system encodes and decodes referents. Concept fidelity is expressed by the probability of obtaining the same referent after encoding and decoding. The matrix $P(\rho|\rho)$ contains exactly this information on its diagonal. To obtain a single measure, the concept fidelity measure is defined as the average of $P(\rho_i|\rho_i)$ over all referents i, reflecting the goal of reliably encoding and decoding *all* referents. This goal differs from that of successfully decoding the current referent as often as possible, which would be achieved by multiplying with the probabilities $P(\rho_i)$. Concept fidelity can be thus computed:

$$P(\rho|\rho) = P(\rho|\mu)P(\mu|\rho) \tag{2.14}$$

This can be seen as follows:

$$P(\rho_i|\rho_j) = \sum_{k=1}^{n_m} P(\rho_i|\mu_k)P(\mu_k|\rho_j) \tag{2.15}$$

$$\mathsf{fid}(\mathcal{A}) = \frac{\sum_{i=1}^{n_r} P(\rho_i|\rho_i)}{n_r} \tag{2.16}$$

Calculation of the Matrices

The parsimony and distinctiveness measures are based on matrices containing the conditional probabilities $P(\rho|\mu)$ and $P(\mu|\rho)$. There are multiple ways of obtaining values for these probabilities, each with their own advantages and shortcomings. In this section, the *direct* calculation method is described. In chapter 3 (section 3.4.2), a method based on sampling is described. The direct calculation method is so called because it directly transforms the conceptual system of an agent into one of the two matrices mentioned; no sampling over time-intervals is necessary. The conditional probability can be computed as follows:

$$P(\mu_j \mid \rho_i) = \frac{P(\mu_j \wedge \rho_i)}{P(\rho_i)} \tag{2.17}$$

The probability of a meaning or referent is equal to the probability of an experience being located in a region of interaction space that corresponds to that meaning or referent. Thus, what formula 2.17 tells us is that for a referent ρ_i and a meaning μ_j, the probability $P(\mu_j|\rho_i)$ is equal to the probability of an experience being located in a region of the interaction space corresponding to both ρ_i and μ_j divided by the probability of it being located in a region corresponding to ρ_i. The probability of an experience being located in a particular region of interaction space is obtained by integrating the probability density of experiences over that region.

In the experiments here the referent is encoded into a separate dimension and meanings are hyperrectangles parallel to the axes of the coordinate system. Furthermore, the actual distribution of referents is known; each referent corresponds to an integer. This implies that the projection of the meanings containing the integer value of the referent onto a plane perpendicular to the referent axis contains all necessary information. Here, it will be assumed that the distribution of experiences over these two dimensions is more or less homogeneous. The validity of this assumption is dependent on the problem however. The probability $P(\mu_j|\rho_i)$ is equal to the proportion of the complete projection taken up by μ_j. This value is equal to the product of the fractions of the ranges occupied by the region in each dimension:

$$\prod_{d \in D \setminus d_{ref}} \frac{max\{x_d|x \in \mu_j\} - min\{x_d|x \in \mu_j\}}{max\{x_d|x \in S\} - min\{x_d|x \in S\}} \tag{2.18}$$

where D is the set of dimensions, d_{ref} is the dimension encoding the referent, x_d is the coordinate of x in dimension d, and S is the complete state space or interaction space.

This calculation method has several advantages over methods based on sampling:

- **Accuracy** If the distribution of experiences over interaction space is known, the method yields exact probabilities, in contrast with sampling based methods.

- **Availability** The measure values are available at any point in time, since no time dependent information needs to be gathered.

- **Memory Requirements** It is not necessary to keep estimates or counts of occurrences of probabilities in memory. Although this is not an issue in the experiment described here, it may become an issue when the number of meanings or referents is large.

The main disadvantage of the method is its rigidity. When the formation of concepts is to be compared against a different ideal conceptual

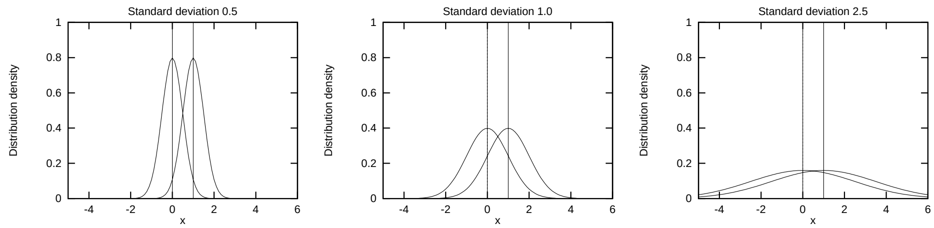


Figure 2.8: Distribution density for three different standard deviations (0.5, 1.0, and 2.5). A broad distribution obfuscates the source of the data (0 and 1), making it difficult to distinguish between the sources.

system, and the shapes of the referents change, the procedure needs to be adapted. Approximation of the matrices by sampling, described in chapter 3, overcomes this problem.

The Effect of Evaluation Noise on Concept Formation

The previous experiments demonstrate how situation concepts can be formed by the adaptive subspace method. Here, the effect of several changes to the environment on concept formation will be examined. The first change concerns the evaluations the agent receives. As described, the problem assigns two basic evaluations to the actions of the agent, which vary around zero and around one. In the following experiments, the amount of variation will be gradually increased, up to such an extent that the distributions have substantial overlap, thus making it difficult to determine whether an action was appropriate given the situation, and hence to determine what situations are to be distinguished. Evaluations are normally distributed with kernels centered around zero and one. Variation in uncertainty is obtained by changing the standard deviation of these normal distributions.

The effect of stochasticity in the evaluations on concept formation has been investigated by varying the standard deviation in the evaluations between $\sigma = 0$ and $\sigma = 2.5$ with steps of 0.25. As figure 2.8 shows, the distributions of the evaluations have some overlap for a standard deviation of 0.5 (leftmost graph), but the larger part of the area under each curve is separated from the area under the other curve. For higher standard deviations however, the distributions become increasingly similar, making it very difficult to distinguish between high and low evaluations. The surface under both curves is a measure for the separability of the distributions; the larger its share in the total area, the less separable the evaluations are.

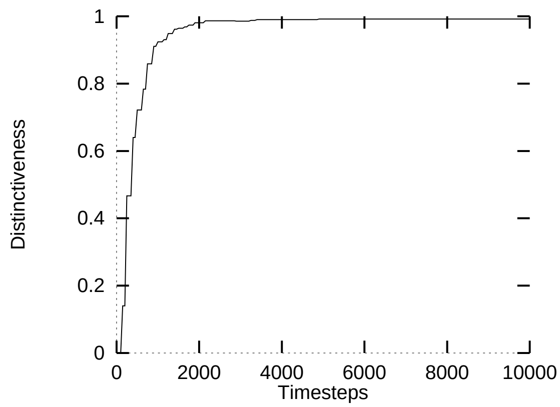


Figure 2.9: The evolution of distinctiveness over time, averaged over ten runs of the experiment. There is a stable trend towards high distinctiveness.

Figure 2.9 shows how distinctiveness develops over time. To get an idea of the variance of this measure across different runs of the experiment, the experiment has been performed 25 times with different initializations of the random generator.

The first effect effect of evaluation noise that is considered here is that on concept fidelity, i.e. the probability that a referent, when coded into a meaning and subsequently decoded again, yields the original referent. This effect is shown in the graphs of figure 2.10. The different graphs concern different settings of the split threshold. Starting in the top left corner, where the threshold is set to 0.1%, it gradually increases until, in the bottom right corner, a split threshold of 5% was used. Each graph shows the same experiment; the concept formation process is active for 10,000 time steps for different noise levels, varying from $\sigma = 0$ to $\sigma = 2.5$. For each combination split threshold and noise level, i.e. for each line on the page, 25 runs have been performed. A line represents the mean of its 25 runs.

The graphs show a clear pattern. For low split thresholds, the agent soon develops a conceptual systems that captures the situations of its environment. This is expressed by the high values of the measure. As the split threshold increases, causing a more conservative splitting behavior, the measure starts to drop for the higher noise levels. This process continues until, for a split threshold of 5%, only three of the original eleven lines are higher than the chance level of 25%, determined by the number of referents (4). Apparently, more sensitive detection is necessary when the patterns in the environment are augmented with evaluation noise.

Another aspect that is less obvious, but clear nonetheless, is the fact that for intermediate and low levels of evaluation noise, the lowest split threshold takes *longer* to reach high probabilities than all other thresholds. The explanation for this is that although the agent will earlier decide to split than in the other cases, these splits apparently are less useful, since they cause smaller increases in the probability that the original referent will be found.

The following set of graphs (figure 2.11) show the parsimony measure for the same experiments. Whereas from the previous graph it would seem that a low split threshold is desirable since it resulted in high probabilities of finding the right referent, these graphs learn that there is a tradeoff with parsimony. For the lowest threshold, parsimony drops in the course of the experiments, although when no noise is added the measure ends up rather higher than when the evaluation signal is noisy. For increasing split thresholds, parsimony increases considerably.

Figure 2.12 shows the evolution of distinctiveness. It is readily seen that the pattern observed here closely matches that of the graphs in figure 2.10. This should come as no surprise; as discussed before, the information conveyed by these measures is similar, to which these graph testify.

In figure 2.13, the number of meanings is displayed. To clearly show the influence of the split threshold, the scales of the graphs have been equalized. The cause for the low levels of parsimony for the lowest split threshold immediately become clear. Except for the case without noise (SD = 0), the number of meanings grows at a steady rate even at the end of the experiment with no clear signs of convergence. This is the price that is paid for eager splitting. Since the number of necessary meanings is four, an overwhelming majority of the meanings produced by the agent in the top left graph are superfluous.

The previous graphs have yielded insight into the influence of the split threshold on concept formation for different levels of evaluation noise. However, our initial purpose of investigating the effect of noise has been fulfilled to a lesser extent. In order to concentrate on this issue, a different form of presentation will now be introduced. The previous series of graphs should serve to clarify what the data represents. The following four graphs show the final value of each measure, i.e. the value of a measure for a fixed combination of parameters after 10,000 time steps, averaged over 25 runs. Since each experiment only needs a single point in these graphs, the different combinations of split threshold and noise level can be represented in one

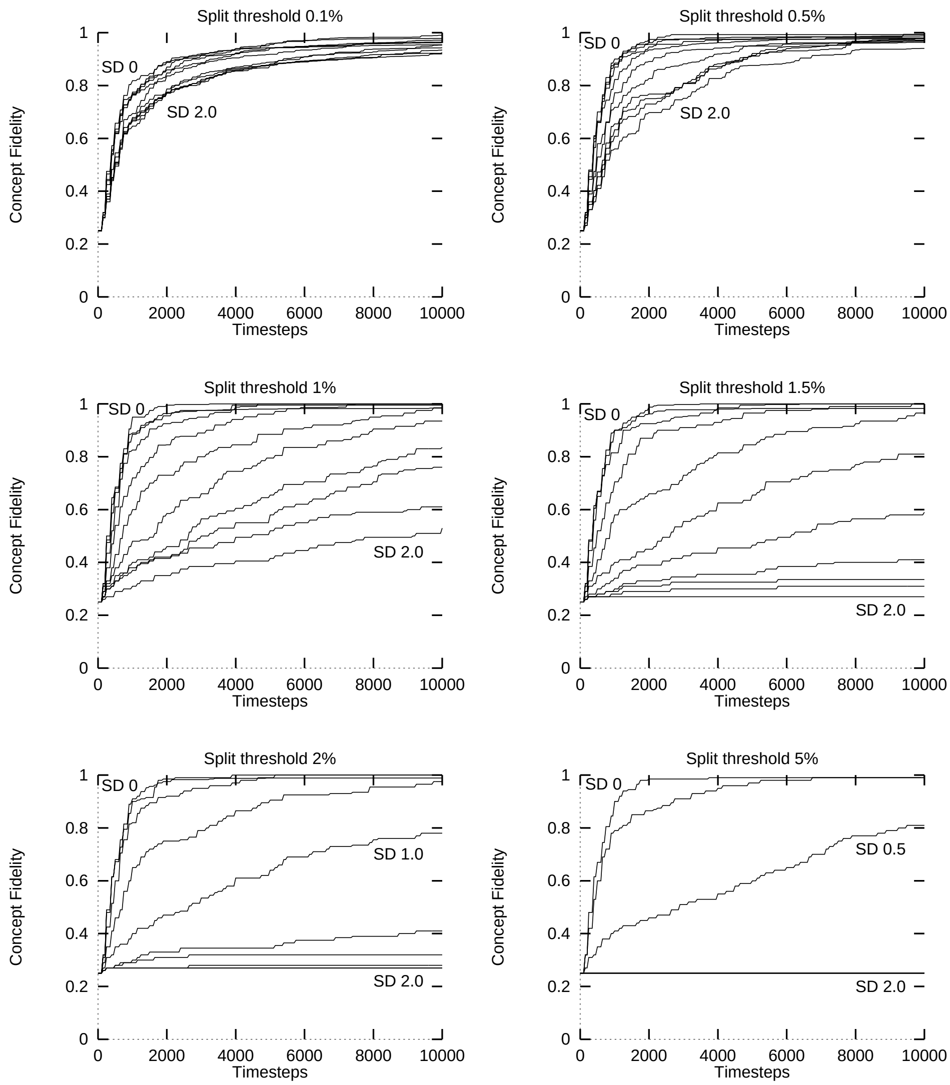


Figure 2.10: Concept fidelity over time for six different split thresholds. For low split thresholds, high concept fidelity is reached across a wide range of noise distributions. For increasing split thresholds, experiments with large amounts of noise are the first to display a decrease in concept fidelity. For the highest thresholds, this development also affects cases with small amounts of noise (starting with $\sigma = 0.5$ for a threshold of 5%).

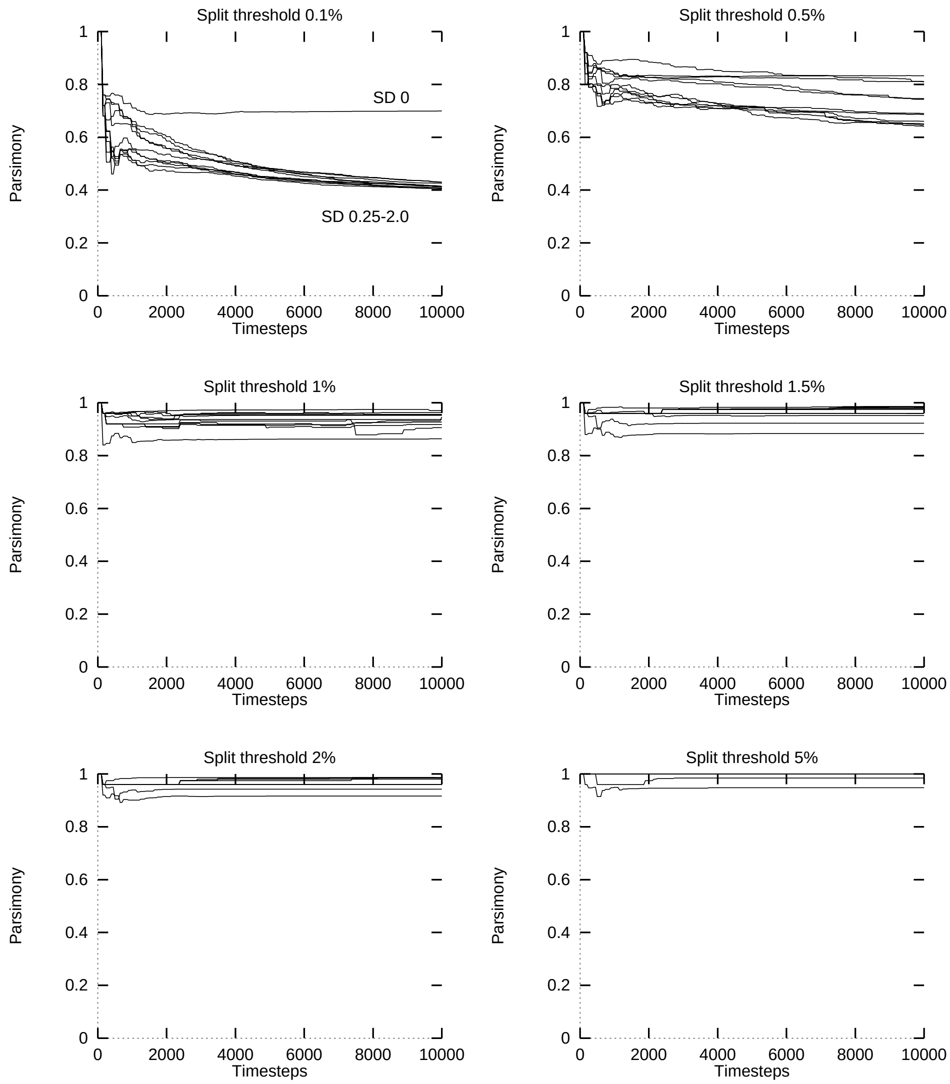


Figure 2.11: Influence of the split threshold on parsimony. For low split thresholds, parsimony is low except in the noiseless case; this is, of course, caused by superfluous splitting. As higher thresholds are chosen, parsimony quickly increases, and remains high (above 0.8) for thresholds of 1% and up for all noise levels.

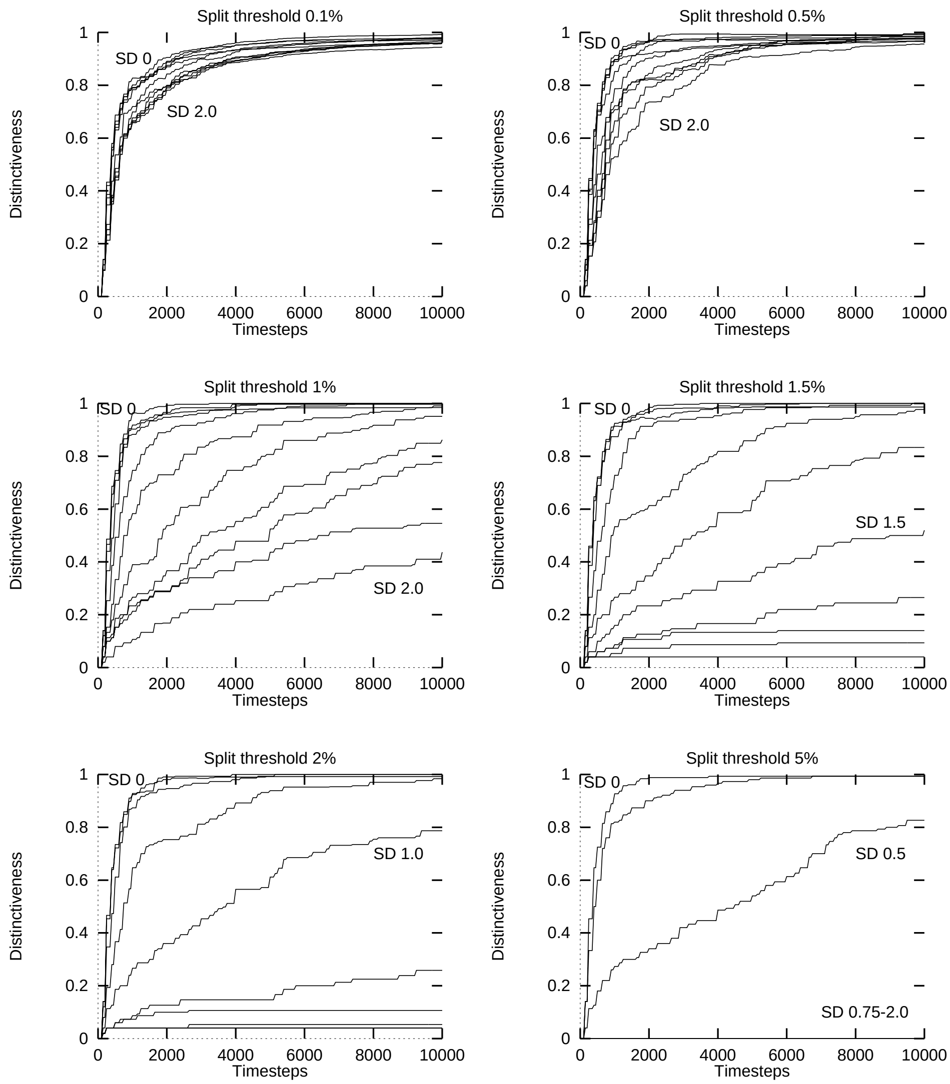


Figure 2.12: The influence of the split threshold on distinctiveness. High distinctiveness is obtained for all noise levels when split thresholds of 0.5% or smaller are chosen. For higher thresholds, the range of noise levels for which high distinctiveness is reached diminishes, starting with the highest noise levels.

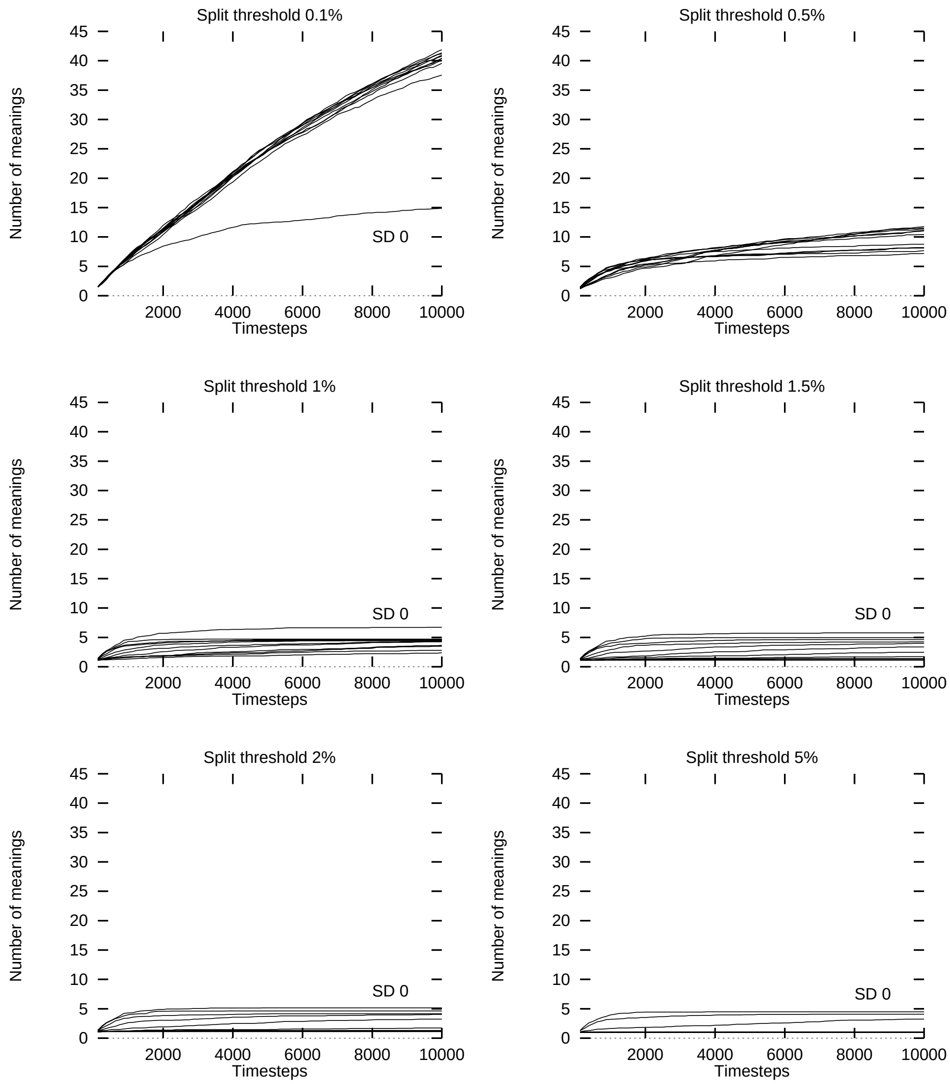


Figure 2.13: The number of meanings generated by the agent as a function of the split threshold. For very low split thresholds (0.1%) large amounts of meanings are generated as soon as a little noise is present. When split thresholds are chosen only slightly higher (starting at 0.5%), this problem is soon solved, and the number of meanings remains limited, up to a point where, when split thresholds are chosen too high in combination with high noise levels (bottom right), the number of meanings can fall below the number of referents.

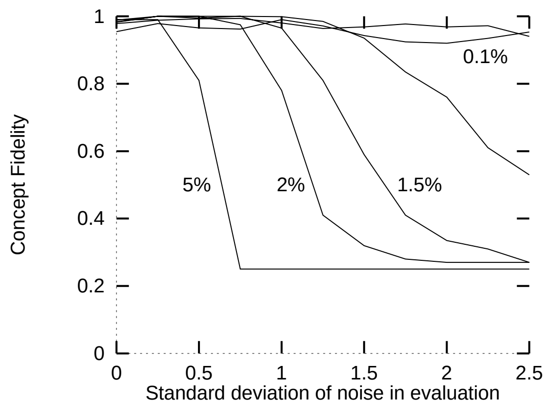


Figure 2.14: Concept fidelity for the six different split thresholds across different levels of noise.

graph. The graphs show the final value of a measure as a function of noise level, and each line depicts a different split threshold, as specified in the graphs.

Figure 2.14 shows the influence of raising the split threshold on concept fidelity. This graph is a condensed form of figure 2.10, showing the final outcomes of the runs. This condensed representation of the experimental results shows more clearly how performance is influenced by noise, since the different noise levels are represented linearly on the horizontal scale instead of as different lines in the former graph.

For a threshold of 0.1% and lower, the concept fidelity measure is high (above 0.9) for all levels of evaluation noise that have been investigated. For higher thresholds the measure is high up to a certain level of noise, after which it quickly drops. When the threshold increases, this drop takes place earlier, i.e. at a lower level of noise. The behavior is very similar to that of distinctiveness. This is understandable, since there is a close relationship between distinctiveness and concept fidelity; if each meaning allows one to identify the referent (high distinctiveness), concept fidelity will also be high, no matter how many meanings there may be (low parsimony). When the distinctiveness is lower, the relationship between distinctiveness and concept fidelity depends on the occurrence frequencies of the referents and the meanings.

Figure 2.15 shows the effect of noise on parsimony. For a very low split threshold (0.1%), the combination with zero noise yields a much higher parsimony than the other combinations. If there is no noise, there are

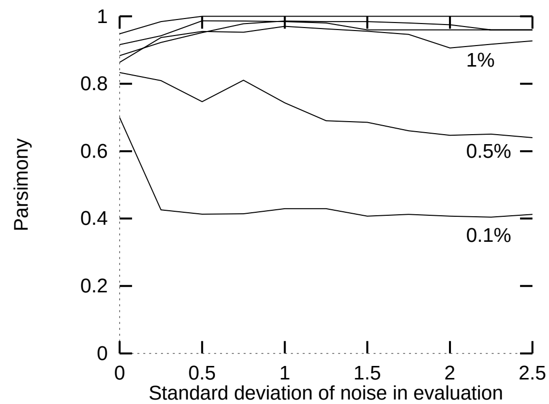


Figure 2.15: Parsimony as a function of noise level for the different split thresholds.

less differences, and has splitting will occur less, and hence the number of meanings remains low. Apart from this effect, the lines in the graph are strikingly horizontal, while differences in the vertical dimension are much more outspoken. From this, we can conclude that the split threshold is a more important factor for parsimony than the level of evaluation noise.

This graph shows the advantage of the concentrated representation. Although all information for discovering the pattern that parsimony depends on the split threshold was present in the graphs of the time series, this relationship is more readily from the current figure.

The similarity between distinctiveness and concept fidelity that was observed in the time series is found back in the concentrated graph of figure 2.16; the noise level at which the measure drops decreases for increasing split thresholds.

Finally, figure 2.17 shows the number of meanings for all combinations of noise level and split thresholds that have been investigated. For very low thresholds, the number of meanings grows quite high, and its final value is constant across the different noise levels (apart from the zero noise case, as discussed above). For higher thresholds, two effects are observed. The first effect is that the number of meanings drops when the threshold increases, as one would expect. The other effect is that the number of meanings after 10,000 time steps drops for increasing levels of noise, especially for higher thresholds. Ostensibly, the variations in evaluations due to noise obfuscate the differences in evaluations. These differences then fail to exceed the higher thresholds, which leads to conservative configurations of concepts.

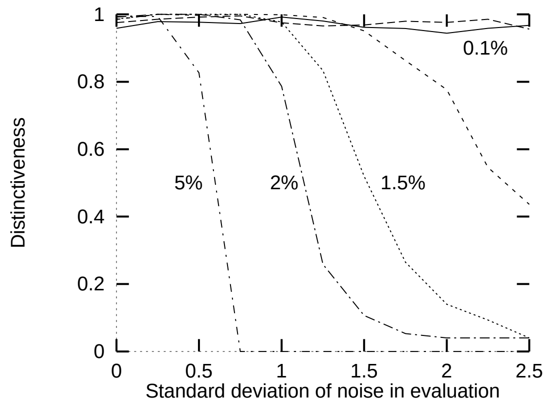


Figure 2.16: Distinctiveness as a function of noise level for the different split thresholds.

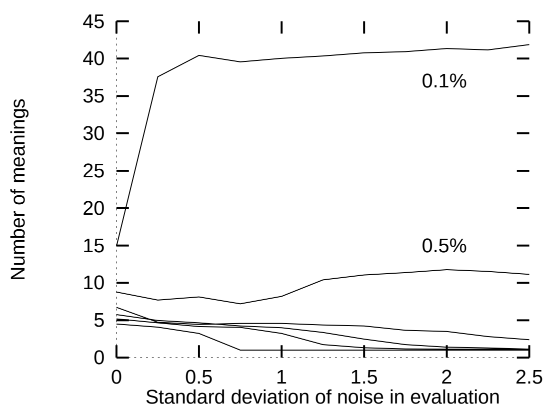


Figure 2.17: Number of meanings as a function of noise level for the different split thresholds.

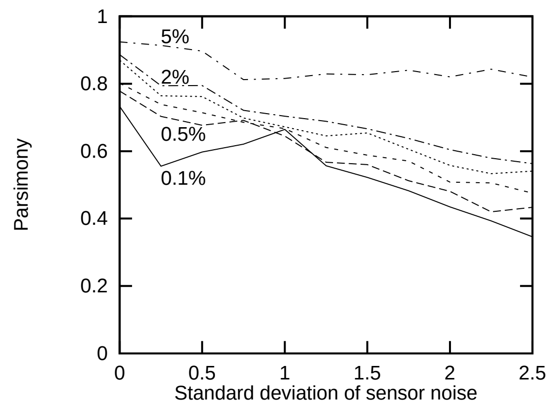


Figure 2.18: Parsimony as a function of sensor noise for the different split thresholds.

Stability under Noisy Sensors

In the previous section, it has been shown how the adaptive subspace method for concept formation reacts to noise in evaluations. Here, the effect of sensor noise will be examined. The experimental procedure is analogous to that of the previous experiments, which allows for a straightforward comparison of the results. In order to save space, only the condensed graphs will be shown. The procedure by which these graphs were produced has been explained in section 2.6.3. The graphs shown in this section correspond to figures 2.15 through 2.17. In the evaluation noise experiments, the results for the concept fidelity measure are almost identical to those of the distinctiveness measure. This is also the case for the sensor noise experiments, and therefore the graph has not been included here.

Figure 2.18 shows the parsimony of an agent's conceptual system for the specified split thresholds and noise levels. As in the evaluation noise experiments, and in line with expectations, low split thresholds yield lower parsimony. A difference however is that for split thresholds of 1% and more, the parsimony is much lower as well.

Figure 2.19 shows how distinctiveness is affected by sensor noise. Comparison with figure 2.16 makes clear that the effect of sensor noise is different in nature from that of noise in the evaluations. Whereas evaluation noise was handled best by low split thresholds, here the opposite is the case; distinctiveness *decreases* for decreasing split thresholds. Apparently, sensor noise leads to splits that are not necessary without noise. Another difference between the two types noise is that the distinctiveness remains

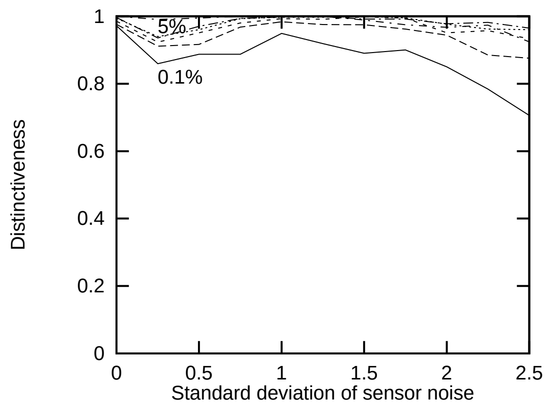


Figure 2.19: Distinctiveness as a function of sensor noise for the different split thresholds.

much higher here than under evaluation noise, where it dropped to zero for high split thresholds.

Figure 2.20 shows the number of meanings constructed by the agent for the different levels of sensor noise. Compared to the evaluation noise case, the number of meanings is much higher. This fits with the results concerning the distinctiveness measure. Whereas the number of meanings for higher split thresholds dropped to zero in the case of evaluation noise, here a substantial number of meanings are produced for even the highest split thresholds. Evidently, the meanings that are created include useful meanings, as they result in high levels of distinctiveness. The fluctuations due to evaluation and sensor noise have the same ranges (standard deviations between zero and 2.5), whereas the ranges of sensor values are larger than those of the evaluations. It might therefore be expected that concept formation is less disturbed by the sensor fluctuations than by the variation in the evaluations. However, although the complete range of the situation sensor is larger than that of the evaluations, the different values that need to be distinguished are only a single unit apart, hence this does not explain the effect. Therefore, a closer analysis will now be performed.

Nature of the Different Noise Sources

Although the different effects of evaluation noise and sensor noise have been charted, the source of these differences has not yet been explained. To this aim, a simulation experiment has been performed where a split in the range of a fictitious sensor is considered. The sensor has a range of $[0, 10>$, and

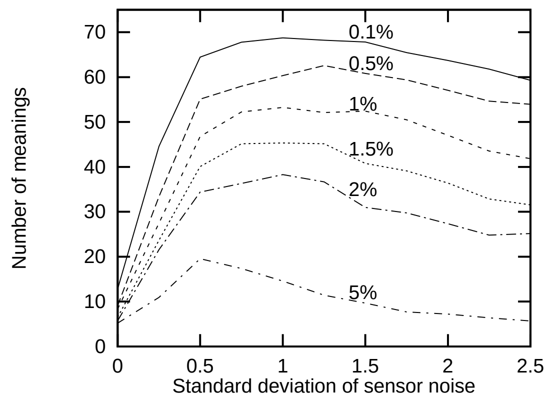


Figure 2.20: Number of meanings as a function of sensor noise for the different split thresholds.

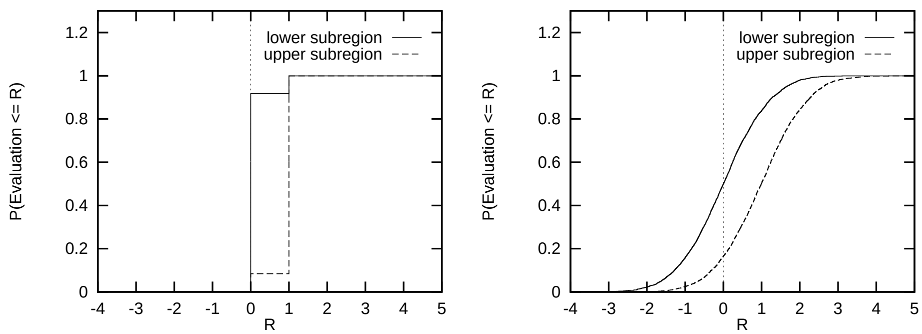


Figure 2.21: Effect of noise in the sensor values (left) and the evaluations (right).

the possible split is the hyperplane $x = 5$. Rewards are in principle zero in the lower half ($[0, 5>$) and one in the upper half ($[5, 10>$) of the sensor range, but either the sensor values or the rewards contain noise with a standard deviation of 1.0. Figure 2.21 shows the effects of these respective noise sources.

The shapes of the distributions in figure 2.21 are explained as follows. If there would be no noise at all, the distribution of the evaluations in the lower subregion ($x < 5$) would rise from zero to one at $R = 0$, and that of those in the upper subregion at $R = 1$. Thus, the graph would be a rectangle with corners (0,0) to (1,1), which has the maximum relative area of 1 and will therefore always lead to a split.

When noise is added to the sensors (left graph in fig. 2.21), some of the evaluations in the lower subregion will be one, and some in the upper subregion will be zero. Graphically, this respectively corresponds to a lowering upper edge and a rising lower edge of the rectangle. The smaller the distance between these horizontal edges, the smaller the area, up to a point where the area does not exceed the split threshold anymore.

When noise is added to the evaluations, the result is rather different. Instead of a vertical shift of the edge of the rectangle, the edges are deformed. The reason for this is that the evaluations are not limited to a few fixed values anymore but are spread around their prototypical values, where the amount of noise determines the extent of the spread. The edges take the shape of the cumulative normal distribution of the noise. As the amount of noise increases, measured in terms of its standard deviation, the slope of the curves decreases and hence their horizontal extent increases.

The graphs suggest an explanation for the higher number of meanings in the case of sensor noise. Whereas the area between the two distributions in on the left side is almost equal to the maximum area (the rectangle between (0, 0) and (1, 1)), this is not at all the case for the evaluation noise. The maximum area is determine by the smallest and largest evaluations that have been observed by the agent during its life, i.e. the leftmost point where the distribution is greater than zero (around R = -3.5) and the point where it reaches one (around R = 3.5). Thus, the maximum area in this case is about 7. Since the areas between the distributions is comparable to that of the sensor noise case, the scaled area will be around 7 times smaller. This explains the different effects of noise that have been observed.

It may be noted that the highest split thresholds yield the best results here. This suggests that even higher thresholds might lead to still better results. Additional experiments have been carried out for a threshold of 10%. For low noise levels, it was indeed observed that both parsimony and distinctiveness increased. For higher noise levels however, the differences between the values in different regions of the sensor space where not detected due to the high threshold, leading to a smaller number of concepts than required. Thus, the effect that has been observed for several of the high thresholds in the evaluation noise case is also present here, but only begins to play a role for higher split thresholds or, presumably, higher levels of noise.

Finally, it is important to realize that the introduction of sensor noise blurs the clear distinctions between the different referents. A referent that is normally represented by a sensor value of x is perceived as $x + \delta$ in the presence of sensor noise. This changes the task with which the agent is faced, including the optimal division of the sensor action space into regions; if the optimal regions in the basic problem consists of intervals $[x-2, x-1>$, $[x-1, x>$, $[x, x+1>$ etc., then in the presence of noise, the higher values within the interval $[x-1, x>$ are more likely to correspond to referent x than to referent $x-1$, and hence an extra split should be made, the optimal location of the split depending on the noise distribution.

The measures that are used to monitor the concept formation process are based on the standard problem, where the referents correspond to fixed sensor values. They are computed directly from the static structure of the conceptual system, based on what referents are located within a certain region in sensor action space. Thus, what the measures measure is to what extent the formation of a conceptual system that is optimal for the described problem is disturbed by noise. However, it is important to realize that the optimal conceptual system from each individual agent's point of view changes when sensor noise is added. Thus, another possible experiment is to investigate whether agents can construct appropriate concepts for the problems that results when sensor noise is added. A difficulty in that case is that the referents are not always separable anymore. The maximum value of the measures thus drops below one, and comes to depend on the noise level. Furthermore, a method for calculating the measures based on sampling would be necessary. Although this experiment would also be interesting, it has not been carried out, as the primary interest in this chapter was to determine whether the agents can develop situation concepts about the problem that will be used in the communication experiments, described in chapter 3.

2.7 Conclusions

This chapter started out from the idea that concepts can be formed individually by agents based on interaction with the environment. The importance of evaluative feedback has been argued for. In combination with the idea that concepts influence behavior, this leads to a framework for concept formation based on evaluative feedback.

Certain generalization methods from reinforcement learning can be used to from a particular kind of concepts representing situations in the environment. One particular method, called the adaptive subspace method, has been described in detail, and tested in experiments. The method distinguishes itself from related methods by the fact that it performs both state and action generalization, and by its suitability for continuous problems. By keeping sensor and action distinctions separated, internal nodes of the tree that the method builds come to represent situations in the environment. This renders the method especially suited as a basis for the development of communication.

Three measures for evaluating the quality of a conceptual system have been introduced. Distinctiveness expresses to what extent the referents of an environment can be identified given knowledge of the meanings that the agent associates with the recent interaction history. If these meanings are appropriate in the agent's environment, this will be indicated by high distinctiveness. A second measure of a conceptual system is its parsimony. Although an agent with many specific meanings may be able to distinguish between all relevant situations in which it may find itself, a set of meanings that is much larger than necessary implies a lack of generalization and slows down learning processes, since more experiences will be necessary to learn the same regularities. Parsimony is important in that it aids the development of communication; if the number of meaning corresponding to a particular referent is higher than necessary, all these meanings need to be associated with the same word independently. Finally, concept fidelity requires distinctiveness, and expresses the probability that coding a referent into a meaning and subsequently decoding it back into a referent yields the original referent.

The experiments have shown that the method for concept formation yields useful situation concepts based on interaction with the environment. For the particular experiment that has been carried out, high quality conceptual systems resulted, in terms of distinctiveness, parsimony, and concept fidelity. A further investigation was concerned with the stability of concept formation under the influence of noise. The conclusions from these experiments are that a substantial amount of noise on sensor readings or on evaluative feedback is tolerated. Furthermore, a qualitative description of the influence of noise has been given. The experiments in this chapter are evidence in favor of the idea that evaluative feedback can be used to create concepts that improve the behavior of the agent. In the next chapter, the role of the concepts as a basis for the development of communication will be investigated.

Chapter 3

Associating Concepts with Words

The previous chapter concerned the autonomous formation of concepts about the environment of an agent. In the current chapter, it will be investigated how agents can develop a language based on individually formed concepts with which information can be transferred. A consequence of the view that agents construct concepts is that information in the head of one agent is incommensurable with another agent's knowledge. At first sight then, it would appear impossible to determine whether information is transferred through communication. However, when information is viewed as an entity that decreases uncertainty, a methodology for investigating this very question becomes available.

It should be stressed that the question addressed here is not in the realm of information theory, see e.g. (Shannon, 1948). In information theory, a fixed set of messages is assumed whose meaning is known. Here, the set of messages is not fixed but open, and moreover, the meaning of the messages is not known in advance but is to be established by the agents by means of interaction. Successful development of communication means that relations between messages and meanings emerge, even though it is not assumed that the agents possess common meanings.

The structure of this chapter is as follows. First, in section 3.1, an algorithm for the association of concepts with words is described. To monitor the development of communication, several *measures* are defined. In section 3.2, the different components of the algorithm are taken out one at a time, to see whether they contribute to the development of communication or not. In section 3.3, the question is posed whether the communication that is developed is *useful* for the agents, in the sense that it improves their ability to produce behavior that is appropriate in their environment. Section 3.4 focuses on the effects of differing concepts on communication. Experiments are reported where the formation of different concepts by the different agents in the system is encouraged, and the development of communication under these conditions is studied. Finally, in section 3.5, conclusions are drawn.

3.1 A Mechanism for Associating Concepts with Words

3.1.1 Production

How can private concepts formed by agents become associated with public words? This question is the subject of the current chapter.

The results of concept formation, described in the previous chapter, can be represented in two matrices. The probabilities that a referent ρ gives rise to a meaning μ in the agent are represented in the matrix $P(\mu|\rho)$. This determines how an agent views the world. On the other hand, an agent's meaning indicates the presence of one or more referents. This information is represented in the matrix $P(\rho|\mu)$.

In a system of communication, the agents will use signals or words σ to express their meanings. This set of signals, the *lexicon*, is an open set, i.e. it is subjective to change. As their usage decreases, some words may become obsolete. On the other hand, new words may also be introduced. At any point in time however, the lexicon is fixed. Therefore, we can define a new matrix $P_{prod}(\sigma|\mu)$ which defines the communicative behavior of an agent. When the environment activates meaning μ in an agent, the probabilities $P_{prod}(\sigma|\mu)$ specify what words the agent might use to describe its situation. The subscript *prod* indicates that the word is produced; although the matrix $P(\sigma|\mu)$ will mostly be used in the context of signal *production*, it is important not to confuse this quantity with the probability of receiving a signal.

3.1.2 Interpretation

Communication is only useful when there are one or more agents that receive a signal sent by some agent and can interpret it. Interpretation can, like production, be captured in a probability matrix. The matrix $P_{int}(\mu|\sigma)$ specifies the probabilities that a signal σ gives rise to a meaning μ.

$$
\begin{array}{cccc}
 & \mathtt{PA} & \mathtt{MI} & \mathtt{LO} \\
\mathtt{r1} & 1 & 0 & 0 \\
\mathtt{r2} & 0 & \boxed{1} & 0 \\
\mathtt{r3} & 0 & 0 & 1
\end{array}
=
\begin{array}{cccc}
 & \mathtt{m1} & \mathtt{m2} & \mathtt{m3} \\
\mathtt{r1} & 0 & 1 & 0 \\
\mathtt{r2} & 0 & 0 & 1 \\
\mathtt{r3} & 1 & 0 & 0
\end{array}
\times
\begin{array}{cccc}
 & \mathtt{PA} & \mathtt{MI} & \mathtt{LO} \\
\mathtt{m1} & 0 & 0 & 1 \\
\mathtt{m2} & 1 & 0 & 0 \\
\mathtt{m3} & 0 & 1 & 0
\end{array}
$$

Figure 3.1: Example of the multiplication of matrices $P(\mu|\rho)$ and $P_{prod}(\sigma|\mu)$ to obtain $P_{prod}(\sigma|\rho)$.

Together, the matrices $P^{\mathcal{A}}(\mu|\rho)$, $P^{\mathcal{A}}(\rho|\mu)$, $P^{\mathcal{A}}_{prod}(\sigma|\mu)$ and $P^{\mathcal{A}}_{int}(\mu|\sigma)$ specify the communicative behavior of an agent $\mathcal{A}$. If this information is known for all agents, a complete system of communication, including the perception of the environment by the agents, their production of signals, and their interpretation can be captured. The set of matrices provides a way to hide the 'private parts' of a communication system by directly linking referents to words ($P_{int}(\rho|\sigma)$) and vice versa ($P_{prod}(\sigma|\rho)$). This can be done by a simple matrix multiplication:

$$P_{prod}(\sigma|\rho) = P(\mu|\rho) \cdot P_{prod}(\sigma|\mu) \tag{3.1}$$

and

$$P_{int}(\rho|\sigma) = P_{int}(\mu|\sigma) \cdot P(\rho|\mu) \tag{3.2}$$

The validity of this matrix multiplication can easily be seen:

$$
\begin{aligned}
P_{prod}(\sigma|\rho)[i][j] &= \sum_{k=1}^{n_m} P(\mu|\rho)[i][k] \cdot P_{prod}(\sigma|\mu)[k][j] && (3.3)\\
&= \sum_{k=1}^{n_m} P(\mu_k|\rho_i) \cdot P_{prod}(\sigma_j|\mu_k) && (3.4)\\
&= P_{prod}(\sigma_j|\rho_i) && (3.5)
\end{aligned}
$$

Figure 3.1 provides an example.

3.1.3 Pointing

In order to learn to what meaning a word refers, a connection between the meaning of the agent that produces the signal (the *speaker*) and the meaning of the agent that receives the signal (the *hearer*) needs to be established. Referents that are simultaneously recognized by speaker and hearer make this connection possible. Such a referent will be accurately captured by some meaning for both agents during concept formation. A referent is not necessarily represented by a single meaning, but the number of meanings to which it is connected is limited. Thus, when both agents are considering the same referent, one of a limited number of meanings will be activated in the agents, thus providing the necessary connection. Given this connection, it becomes feasible to learn a relationship between a meaning and a word.

The situation concepts that were described in section 2.5, are a good example of concepts for which the above condition on referents holds; if two agents are in the same situation, the referent corresponding to this situation will be recognized by both agents, provided their mechanisms for concept formation are functioning properly and are similar in their workings. For many other types of concepts however, an additional mechanism is necessary to identify the subject of communication. The process that establishes the desired connection is referred to as *pointing.* A prototypical example of pointing is the well known special case where an adult wants to teach the name of some object to a child, or to a speaker of a different language. By pointing, the referent is fixed, and it will be understood that the words that are subsequently produced refer, via the private meaning of each agent, to that very referent. For communication about situations, such a mechanism is not necessary; the fact that agents are in the same situation provides the required correspondence between the subject of communication of the agents.

3.1.4 Description of the Algorithm

The experiments in this chapter investigate whether individual concept formation can serve as the basis for a system of communication. That is, can private concepts become associated to public words, such that a means for communicating about the environment becomes available? To investigate this, mechanisms for associating concepts with words have been investigated. Since one of the commitments is that concepts are internal, a condition for the experiments is that such an association can only be adapted by the agent that has created the concept.

Each experiment starts with the situation where no communication system is present. Agents have no words associated with any of the situations, and will produce a random word when present in a situation that has no associated words yet. The words received from other agents will be associated with the situation. Thus, during the experiment, agents interact by uttering words that correspond to their situation, and the association strength of each word that is heard has its association strength with that particular situation increased. As a result of these interactions and adaptations, coherent systems of communication can emerge through self-organization. A defining property of self-organization is that the organization it results in is not caused by a central organizing force as would be the case if one agent were to dictate the meaning of each word. It is essential that agents speak *and* listen; only by the combination of these activities can situations come to be connected to words.

Internal Structures

The agents in the experiment contain the following internal structures:

- Meanings. These are formed during the experiments by the concept formation method described in chapter 2. For each meaning μ, the following information is maintained:
  - A list of experiences gained when the meaning was present. These experiences are used in the concept formation process.
  - For each action, an estimate of the feedback following the selection of the action given the presence of the meaning.
  - A list of words that are associated with the meaning. For each association, two values are stored: *use* and *success*. Use and success are linearly combined into a single value, called association strength. Furthermore, an estimate of the probability $P_{int}(\mu|\sigma)$ is maintained for each word σ; its updates are based on the signals the agent receives.

These structures are used as follows. The use variable measures how often the word is received when the meaning has been observed, and hence may be viewed as an approximation of $P_{int}(\sigma|\mu)$. The success variable represents whether, given the word that was heard, the meaning corresponds to the actual current situation (referent). Naturally, this cannot be determined directly by the agent, since agents have no knowledge about referents. Therefore, it tries to obtain this knowledge indirectly. If the determination of the situation based on the received word is correct, then the feedback following the action should correspond to the estimated feedback of that action if the estimate is accurate enough. If on the other hand there is a

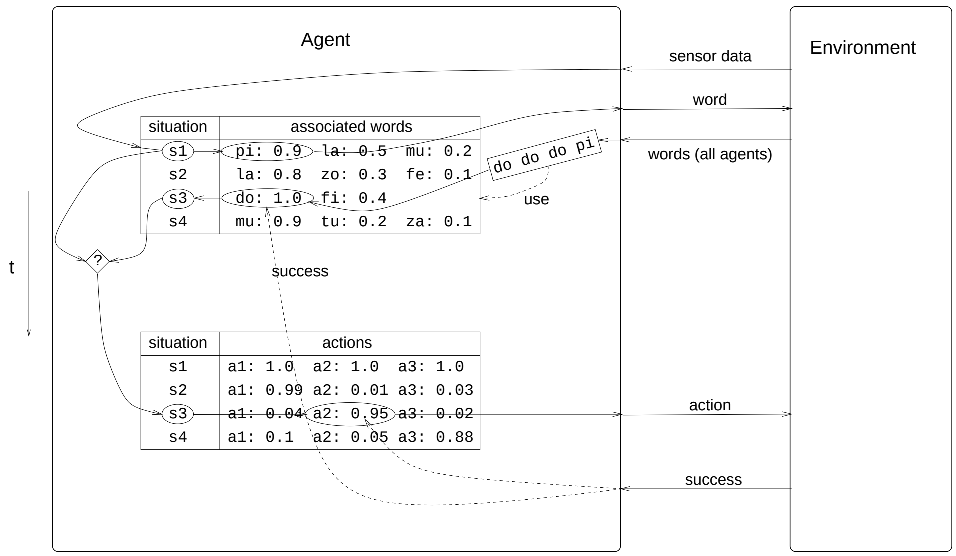


Figure 3.2: Basic cycle of the experiment. The agent receives its sensor data, determines its sensor-situation and produces a word associated with it. It then receives all words produced by the agents within its hearing distance and determines the signal-situation, and selects it if it is probable enough. The action is based on the situation, and feedback from the environment is used to adapt the success of actions and, in case the signal-situation was used, the association between word and situation.

discrepancy between the estimate and the actual feedback, this indicates that the determination of the situation was incorrect. This information is used to update the success value for the used combination of meaning and word.

Basic Cycle of the Experiment

The basic cycle of the experiment is as follows (see figure 3.2). First, the agent receives its sensor data from the environment. Given the set of meanings it has created, the sensor data directly and uniquely determines the present meaning. This meaning will be called the *sensor-situation*, since it follows from the sensor readings. The agent then utters a word. The word is selected from the list of words associated with the sensor-situation. The choice is based on the association strengths of the different words.

All agents produce a word describing their view of the situation. In general, an agent can hear the signals of those agents within hearing distance. Unless otherwise specified, this hearing distance matches the size of the environment, so that all words, including that uttered by the agent itself, will be heard. The agent then determines the most likely situation given the list of words it heard. This situation will be called the *signal-situation.* It is the situation μ for which $P_{int}(\mu|\sigma) \cdot \text{freq}(\sigma)$ is highest for some word σ that was received $\text{freq}(\sigma)$ times.

The estimated probability that the signal-situation is correct, $P_{int}(\mu|\sigma)$, is also used to choose between the sensor-situation and the signal-situation. The latter will be selected with probability $P_{int}(\mu|\sigma)$, and hence the former with probability $1 - P_{int}(\mu|\sigma)$.

The situation thus selected is used to choose an action. In principle, the action with the highest estimated success is selected, but sometimes, other actions are tried as well in order to ensure exploration, as described in chapter 2.

The feedback from the environment is used both to adapt the success of the actions. Furthermore, in case the signal-situation was used for action selection, the success value of the association between word and situation is adapted. If the feedback was sufficiently close to the estimated success of the action, this indicates that the choice of situation was correct, and the success value is increased. If not, it is decreased. Increasing and decreasing are done by adapting the value v towards a goal z of one and zero respectively:

$$v = (1 - \alpha)v + \alpha z \tag{3.6}$$

Updates of this type will be referred to as *recency-weighted average.* The basic cycle is described in pseudo code in figure 3.3.

x := receive-input();
sensor-situation := **determine-situation-from-sensors**(x);
produce-signal(sensor-situation);
signals := **receive-signals**();
signal-situation := **determine-situation-from-signals**(signals);
max-signal := $\arg\max_{\sigma \in \text{signals}} \{P(\text{signal-situation}|\sigma) \cdot \text{freq}(\sigma)\}$;
if $\delta_{avg} \leq$ max-error-for-signals
and P(signal-situation | max-signal) > random(1.0)
    situation := signal-situation;

```
else
    situation := sensor-situation;
    signal-situation := nil;
action := choose-action(situation);
act(action);
store-experience();
R := receive-reward();
previous-value := value;
value := value(action);
if situation = signal-situation; signal based situation determination
    if step > 0 and |R - value| < max-error-for-signals
        increase-association-success(signal-situation, max-signal);
        for s ∈ situations \ signal-situation
            decrease-association-success(s, max-signal);
    else
        decrease-association-success(signal-situation, max-signal);
else
    all-signals := signals(sensor-situation);
    increase-association-use(sensor-situation, signals);
    decrease-association-use(sensor-situation, all-signals\signals);
    update(P_int(μ|σ));
δ := R - previous-value;
update-value(δ);
δ_avg := (1.0 − α)δ_avg + α | δ |;
step := step + 1.
```

□

Figure 3.3: Pseudo-code of the algorithm.

Below, the functions and variables referred to in the pseudo-code description of the algorithm are explained.

- **produce-signal** Select a word for production according to the association strengths between the situation and each word in its word list. The choice is based on the Boltzmann distribution, as will be discussed below. Association strength s is a combination of the *use* and *success* values, and is calculated as follows:

$$s = \alpha success + (1 - \alpha) use \tag{3.7}$$

where α is a parameter called the success-use-ratio. A standard value for this parameter in the experiments is 0.5, which results in an equal weighting of use and success during word production. If no word is associated strongly enough, a random new word is produced. This condition is operationalized by the testing whether the maximum association strength is at least 0.25.

- **determine-situation-from-sensors** The situation is determined by starting at the root of the tree formed as a result of concept formation, and following the branches that are consistent with the sensor information, until an action distinction is encountered.

- **receive-signals** The agent receives the signals produced by all agents within its hearing range.

- **determine-situation-from-signals** Returns

$$\arg\max_{\mu} \sum_{\sigma \in \text{signals}} P_{int}(\mu|\sigma) \cdot \text{freq}(\sigma) \tag{3.8}$$

the situation for which the probability is maximal for some word σ. $\text{freq}(\sigma)$ is the fraction of the signals equal to σ or equivalently its relative frequency, scaled to

- **choose-action** Select an action based on the action value estimates of the situation using the ϵ-greedy method.

- δ_{avg} Running average of the error in the prediction of the value of the action that was last selected.

- **act** Execute a given action.

- **store-experience** Store an experience. Experiences have the following form: $< S_{t-1}, A_{t-1}, R_{t-1}, S_t, A_t >$ where S is a situation, A an action, and R the evaluation received from the environment. This operation is only performed if the previous situation S_{t-1} was determined using sensors as opposed to signals, to prevent incorrect situation determinations from disturbing concept formation. For the experiments here, actions do not influence rewards beyond the immediate next reward. Hence there are no delayed rewards, γ and λ can be zero, and the learning rule becomes equivalent to the recency-weighted average of R. Storage is therefore not dependent on whether the *second* situation determination was based on sensors or signals here, but this condition may be necessary for problems where delayed rewards occur.

- **receive-reward** Receive the reward following the last selected action from the environment.
- **value** Gives an estimate of the value of an action in a situation.
- **increase-association-use** Increase the use value of the association between a situation and a list of words. For each word, the value of the corresponding association v is updated towards the maximum value of 1 using the recency-weighted average update (see eq. 3.6). For words that occur multiple times, the update is executed once for each occurrence. New words that were not associated with a meaning yet are assigned a use and success value of 0.5.
- **decrease-association-use** Identical to increase-association-use, but with a target value of zero.
- **increase-association-success** Analogous to increase-association-use.
- **decrease-association-success** Analogous to decrease-association-use.
- **update-value** Update the estimate of the action values based on the most recent experience. The learning algorithm used is SARSA(λ).
- **max-error-for-signals** This parameter is a maximum threshold for the average approximation error of the reward. It is used in two ways. First, it is used to determine whether the determination of the situation based on signals was correct, by comparing the most recent approximation error to the average one. Furthermore, it is a threshold that determines whether signals may be used to determine the situation; if knowledge about the environment is not stable yet,

the test of whether the determination was correct can not be expected to function.

- $P_{int}(\mu|\sigma)$ This value is approximated by updating estimates for each combination of a meaning μ and a word σ. For each word σ that is received, the estimate for $P_{int}(\mu_i|\sigma)$ of the current meaning is increased, while the estimates $P_{int}(\mu_j|\sigma)$ for other meanings μ_j $(i \neq j)$ are decreased.

- **step** A counter, used to detect whether the current cycle is the first one.

Both use and success takes values in the interval $[0 \ldots 1]$. When an association between word and meaning is first introduced, both use and success are initialized at 0.5.

Word Production

Selection of the word that an agent produces is based on the association strengths with the current situation that the agent determined. Many different schemes for this selection could be envisaged. One would be that agents always selects the association with the highest strength. However, as will be seen in chapter 4, this can lead to a sort of deadlock to due to a lack of exploration.

Another possibility is to view the association strengths as relative probabilities. The effect of this is opposite, in that the word with the highest association strength may be selected in only a fraction of the cases; this happens when many other words are associated with the situation, possibly with much lower association strengths. To give an example, suppose that twenty words have been heard in combination with a certain situation. Since both use and success are initially 0.5, let's suppose their association strengths are 0.5 on average. Then there is one word with a very high association strength, e.g. 0.90. When using equal probabilities, this most strongly associated word will only be used with probability $\frac{0.9}{20 \cdot 0.5 + 0.9} \approx 0.08$.

It would be desirable if selection of words is a little more flexible than simply choosing the hard maximum, while maintaining a strong preference for words with high association strengths. This is possible using a *softmax* method, such as the Boltzmann distribution. The Boltzmann distribution is the probability distribution of the possible states of a gas at high temperature. Using the Boltzmann distribution, the probability of selecting a word σ_i with association strength $v(\sigma_i)$ becomes:

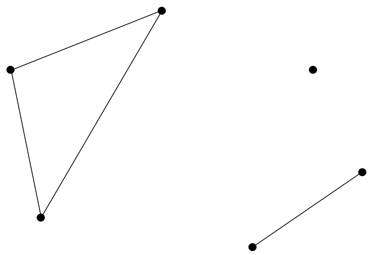
$$P_{prod}(\sigma_i) = \frac{e^{v(\sigma_i)/T}}{\sum_{j=1}^{n_s} e^{v(\sigma_j)/T}} \tag{3.9}$$

the degree to which high association strengths are preferred can be modulated by adjusting the temperature parameter. For high temperatures, more exploration occurs, and thus strong associations are selected *less* often compared to the distribution with probabilities proportional to strengths. In the limit of increasing temperature, the distribution becomes equiprobable, since the fractions $\frac{v(\sigma_i)}{T}$ tend to zero. The opposite effect is achieved by selecting a low temperature, e.g. 0.1. For the same example, the association with the high association strength will now be selected with probability $\frac{e^{\frac{0.9}{0.1}}}{20 \cdot e^{\frac{0.5}{0.1}} + e^{\frac{0.9}{0.1}}} \approx 0.73$.

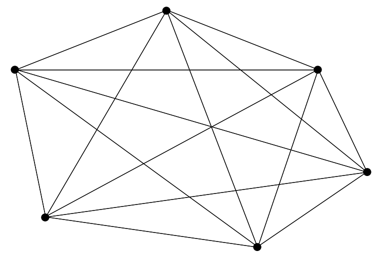
3.1.5 Measures for Association Formation

In chapter 2, two measures have been described that can be used to monitor the concept formation process and to evaluate the conceptual systems that it produces. Here, new measures will be introduced with the same purpose, but concerning the process of forming associations between concepts and words. Both measure concern the *production* of words, not their interpretation. The first measure that will be introduced is *specificity.* This measure is the analogue of distinctiveness. Whereas distinctiveness expresses whether agents have specific *meanings* for referents, specificity captures to what extent agents have specific *words* for referents. A second measure that will be discussed is *consistency.* It expresses to what extent an agent consistently uses the same word for a particular concept, and is to be regarded as the association formation counterpart of the parsimony measure.

Together, distinctiveness and parsimony are sufficient to evaluate the quality of a conceptual system; if both measures are equal to one, every referent is represented by a meaning, and no superfluous meanings are in use. Likewise, specificity and consistency are sufficient to evaluate the system of associations between concepts and words of an individual agent, and thus indicates whether the agent starts to speak a language. However, although these measures are useful for the purpose of analyzing a single agent's word production, they don't capture the most important aspect of association formation, viz. whether the agents start to speak the *same* language. This question can only be investigated with respect to a *population* of agents,



A concept graph: each node represents a concept
A vertex indicates two concepts have the same word



The corresponding fully connected graph:
All concepts are associated with the same word

and is expressed in the *coherence* measure. Coherence is defined as the proportion of the population that uses the same word for a particular concept. Finally, it is interesting to know whether the words produced in a communication system are understood. To this end, the *fidelity* measure expresses the probability that a referent, when encoded by one agent via a meaning into a word, will yield the same referent when decoded by another agent.

Specificity Based on Preferred Words

First a measure of specificity based on preferred words will be described. Preferred words are words which have the highest association value for some referent.

Specificity can be computed by constructing a graph in which the nodes represent the referents. In this graph, two referents are connected if and only if the words primarily associated with these referents are identical. Such referents cannot be distinguished from each other when the preferred word is used. The optimal graph then is a graph without any edges at all. The worst case, in which a single word is used for every situation, results in the fully connected graph. An exception occurs when no word is associated with a certain referent. That referent is effectively connected to all other referents, since the agent has no word to distinguish it from those. This exception appropriately decreases the specificity when one or more referents lack words.

The specificity `spec` is inversely proportional to the connectivity of the graph, i.e. the fraction of edges that are present in the graph:

$$\mathsf{spec} \equiv 1 - \frac{v}{\frac{n_s^2 - n_s}{2}} \tag{3.10}$$

where v is the number of edges and n is the number of nodes in the graph. For reasons of implementation, it is useful to note that the same figure can be obtained by adding the agent-specific frequencies of the words and computing the difference with the maximum sum relative to the maximum value of this difference:

$$\mathsf{spec} = \frac{n_s^2 - \sum_{k=1}^{n_s} f_k}{n_s^2 - n_s} \tag{3.11}$$

The equivalence of the two calculation methods can be seen as follows. The frequency f_k of a concept's word decreased by 1 equals the number of other concepts associated with this word, and hence the number of outgoing edges from this concept's node. Accumulating these numbers and dividing by 2 yields the total number of edges v in the graph:

$$v = \frac{\sum_{k=1}^{n_s}(f_k - 1)}{2} = \frac{-n_s + \sum_{k=1}^{n_s} f_k}{2} \tag{3.12}$$

Substituting 3.12 for v in 3.10 shows the equivalence of 3.10 and 3.11:

$$1 - \frac{2}{n_s^2 - n_s} \cdot \frac{-n_s + \sum_{k=1}^{n_s} f_k}{2} = \frac{n_s^2 - \sum_{k=1}^{n_s} f_k}{n_s^2 - n_s} \tag{3.13}$$

$$n_s^2 - n_s + n_s - \sum_{k=1}^{n_s} f_k = n_s^2 - \sum_{k=1}^{n_s} f_k \tag{3.14}$$

Specificity Based on Entropy

Specificity as it was described in the previous subsection is intuitively clear, since it can easily be depicted graphically. However, the fact that it is only based on preferred words is an idealization that may give a distorted picture if in actual production behavior a meaning is not always expressed by the same word. By taking the same approach as for the distinctiveness and

	ρ_1	ρ_2	ρ_3	ρ_4	$\sum f$	*Specificity*
a_1	1	1	1	1	4	1.0
a_2	1	3	3	3	10	0.5
a_3	1	1	1	1	4	1.0
a_4	1	2	1	2	6	0.83

Table 3.1: An example calculation of specificity based on preferred words. The words are shown in table 3.2. For each agent, frequencies of the words are summed. The specificity of an agent's signals can then be calculated using equation 3.11.

parsimony measures, a principled way for taking into account all associations is obtained. This approach is to calculate the decrease in uncertainty that knowledge of some entity yields. For the concept formation measures, these entities were referents and meanings. Here, they are referents and words. The specificity of a word is thus defined as the relative decrease of uncertainty in determining the referent given a word that was produced.

$$H(\rho|\sigma_i) = \sum_{j=1}^{n_r} -P_{prod}(\rho_j|\sigma_i)\log P_{prod}(\rho_j|\sigma_i) \tag{3.15}$$

$$\mathsf{spec}(\sigma_i) = \frac{H(\rho) - H(\rho|\sigma_i)}{H(\rho)} = 1 - \frac{H(\rho|\sigma_i)}{H(\rho)} \tag{3.16}$$

The calculation of distinctiveness and parsimony was based on the average uncertainty over referents and meanings. Since there are no large differences between the occurrence probabilities of the different referents, this method of calculation is expected to be sufficiently accurate for calculating distinctiveness and parsimony. The occurrence probabilities of words however need not be similar at all, and may also vary much over time as words become obsolete and new words are introduced. Therefore, the calculation of the measures concerning association formation, specificity and consistency, are based on actual measures of the occurrence probabilities:

$$\mathsf{spec}(\mathcal{A}) = \frac{\sum_{i=1}^{n_s} P_{prod}(\sigma_i)\mathsf{spec}(\sigma_i)}{n_s}, \tag{3.17}$$

The specificity of an agent's lexicon is thus defined as the specificities of the words weighted by their occurrence probabilities. The specificity of a communication system, which always contains multiple agents, is defined as the average specificity of the agents of that system:

$$\mathsf{spec} = \frac{\sum_{i=1}^{n_a} \mathsf{spec}(\mathcal{A}_i)}{n_a} \tag{3.18}$$

It is interesting to compare the specificity and consistency measures to the entropy based measure of communication quality introduced in (MacLennan, 1991). There, a matrix linking referents to words is also used defined. A difference is that the entries of this matrix contain co-occurrence probabilities, i.e. the probability that a referent and a word occur together, as opposed to the conditional probabilities used here. MacLennan notes that the objective of ideal communication corresponds to matrices containing a single nonzero values on each row and on each column. However, before translating these criteria into operational measures, he notes that such a matrices are very nonuniform, and then derives three measures based on this criterion. One of these is based on entropy, and computes a scaled and translated entropy over all entries at once. The resulting measure has the property that it is maximal for ideal communication matrices and minimal for maximally dispersed distributions. However, an unmentioned problem with this measure is that it yields the same outcome for a variety of matrices representing communication systems that are far from ideal. If for instance a single row or column contains ones while the rest of the matrix entries are zero, the agents can only understand each other in a fraction of the cases. The measure in this case yields the same value as for ideal communication. This is a result of taking the entropy over all entries of the matrix at once. The shortcoming can easily be solved by computing the entropies per row and per column and relating this to the initial entropy, as has been seen above; the measure can then be based on the extent to which the probability distribution over the elements of a row identifies the column of the event, *vice versa*. These two measures could be combined if it seems important to have a single measure for the quality of communication. In the communication experiments later on in this chapter, it will however be seen that it is often useful to distinguish between the two resulting measures in order to understand what is happening in a simulation.

Consistency

Specificity indicates to what degree the words an agent uses determine the referent that is the subject of communication. Another important property of word use is whether an agent will use the same word when describing a referent, or whether it will use a different word every time. This is expressed in a measure called *consistency*. Perfect consistency means that the agent consistently uses the same word for a particular referent. This measure can be viewed as the analogue of parsimony; a set of meanings is parsimonious if a particular referent always leads to the same meaning, and a lexicon is consistent if a particular referent will always be encoded by the same word.

$$H(\sigma|\rho_i) = \sum_{j=1}^{n_s} -P_{prod}(\sigma_j|\rho_i)\log P_{prod}(\sigma_j|\rho_i) \tag{3.19}$$

$$\mathsf{cons}(\rho_i) = 1 - \frac{H(\sigma|\rho_i)}{H(\sigma)} \tag{3.20}$$

$$\mathsf{cons}(\mathcal{A}) = \frac{\sum_{i=1}^{n_r} P_{prod}(\rho_i)\mathsf{cons}(\rho_i)}{n_r}, \tag{3.21}$$

Coherence

The coherence measures to what extent agents use the same word for a particular referent. Thus, it is a *population* measure, which can only be computed for two or more agents. There is a relationship between consistency. If there is no consistency, i.e. if agents use different words for the same referent at different occasions, then it is very improbable that these agents all happen to use the same word on each occasion. Hence, consistency can be viewed as a necessary condition for coherence.

For each referent, the coherence is measured by computing the fraction of agents that has a word as preferred word for that concept, and then taking the maximum of these fractions. If no word is associated with a concept, the frequency of that concept's word is zero. The coherence of a system of communication is the average of the coherence of its referents. An example of the calculation is shown in table 3.2.

	ρ_1	ρ_2	ρ_3	ρ_4
a_1	PI	PA	PO	PU
a_2	PI	PA	PA	PA
a_3	LU	PA	PO	PU
a_4	PI	PA	PO	PA
Max. freq.	3	4	3	2
Coherence	0.75	1.0	0.75	0.5

Table 3.2: An example calculation of coherence for four agents ($a1 \ldots a4$) and four referents ($\rho 1 \ldots \rho 4$). For each referent, the highest frequency of a word is determined, shown in the bottom row. The average of these figures is the coherence.

Fidelity

Although coherence measures whether agents use the same word for the same referent, a high value of this measure it does not guarantee that agents will understand each other; if specificity is low, the agents might be using the same word for every referent, in which case both consistency and coherence would be perfect, but communication would convey no information whatsoever. A straightforward measure indicating to what extent the agents understand each other is obtained by combining production and interpretation behavior to calculate the probability that a referent encoded by one agent will yield that same referent when decoded by another agent. The measure is calculated as follows. For each combination of two agents $\mathcal{A}_1$ and $\mathcal{A}_2$, the matrices $P^{\mathcal{A}_1}_{prod}(\sigma|\rho)$ and $P^{\mathcal{A}_2}_{int}(\rho|\sigma)$ are multiplied to yield the matrix $P^{\mathcal{A}_1,\mathcal{A}_2}(\rho|\rho)$. This matrix specifies for each referent what the probability is that agent $\mathcal{A}_2$ will understand the word used by $\mathcal{A}_1$ to refer to this referent. In other words, it expresses to what degree agent $\mathcal{A}_2$ understands agent $\mathcal{A}_1$. By taking the average of this matrix over all combinations of two agents $\mathcal{A}_1$ and $\mathcal{A}_2$, and averaging the diagonal entries of this matrix, the average probability that two agents will understand each other, averaged over the referents, is obtained. This value is the fidelity measure.

Measures are Not Influenced by Partial Perception

The purpose of the measures that have been described so far (distinctiveness, parsimony, specificity, consistency, and coherence) is to allow researchers in communication to analyze concept formation and communication systems. To this effect, the measures express whether meanings and words consistently occur in combination with particular referents, not with multiple referents, and whether referents consistently give rise to particular meanings and words. As has been described previously, the particular experiments in this thesis include some amount of partial perception.

When implementing the calculation of the aforementioned measures, it has to be decided whether the incorrect information caused by partial perception should affect measure values or not. In this thesis, the latter has been chosen; partial perception does not influence measures. The reason for this choice is that the purpose of the measure is to determine the quality of a conceptual or communication system relative to the best system the agent can build. For example, specificity expresses whether an agent uses different words for different referents. If partial perception were allowed to influence its measurement, one would investigate whether the agent uses the right word for an object its sensors cannot detect. A similar issue arises in the calculation of fidelity, the probability that an agent will understand another agent when that agent produces a signal to describe a referent. If partial perception were included in the approximation of the conditional probability matrices, this quantity would be obfuscated by the effects of partial perception. A further argument in favor of this choice is that if one would want to know the probability that an agent produces a word in the context of a referent including the effects of partial perception, this quantity can simply be obtained by calculation, since the occurrence frequency of partial perception is known. On the other hand, if the calculation of the measures would be influenced by partial perception, the effects of this uncertainty and any inaccuracies in the communication system are mixed, and the latter cannot be analyzed separately.

A possible argument for the alternative option would be that it is also interesting to know what the probability of correctly determining the situation based on the signals is for an agent in the experiment. Since this information is indeed a crucial aspect of the experiment however, it is expressed in a separate measure called *correctness*, which will be described in the next section. For these reasons, the option where partial perception does not influence measures has been selected.

The above discussed the effects of partial perception on measures. This should not be confused with the issue of noise. An obvious variation on the experiment is to investigate whether useful communication still arises under the influence of noise. This question is investigated in section 3.4, and the calculation method used there ensures that measures reflect the actual behavior of agents in the environment, including noise.

Correctness

If a useful system of communication has been built up, agents can reliably communicate about their environment. The conditions for a useful system of communication are specificity and coherence. If a communication system possesses these properties, each agent uses different words for different referents, and all agents use the same word each particular referent. These properties imply that several other properties are present too; coherence implies consistency, and specificity implies distinctiveness.

What could the use of such a system of communication be? One particularly useful property is the ability to convey information. If agents have imperfect knowledge about their environment, then communication may reduce this uncertainty. It is this form of benefit that will be investigated in the current chapter.

For a description of the experiment, see section 2.6.1. In the basic setup, partial perception plays no role. As discussed above however, some form of uncertainty must be present in order for communication to be useful. This is achieved by masking the situation sensor in 10% of the cases. When masked, the situation sensor returns the same value as when the agent is in the standard situation. However, it will receive the high feedback if and only if it takes the action that corresponds to the *actual* situation of the environment. Clearly, the sensors do not provide the necessary information to select the right action, and there is an aim that can be achieved by communication alone.

An important quality measure of a system of communication then, is to what extent the information provided by it is correct. This quantity is expressed by the *correctness* measure. This measure is calculated as follows: based on the signals it receives from the other agents, an agent determines which situation is most likely. No sensor or other information is used in this procedure. The result is a meaning and hence internal to the agent, but the experimenter can inspect the internals of each agent to see whether this meaning uniquely determines the current referent, i.e. the actual situation. If this is the case, the correctness of the determination is one; otherwise, it is zero. The correctness measure is computed over an interval of time as the average of this value.

Unsuccessful determinations can have several causes. A first prerequisite for a successful determination is that the conceptual system of the agent is accurate; if certain situations are not distinguished between, but represented by a single meaning, as is the case at the beginning of the experiment for example, then successful situation determination may not be possible. As an example, suppose that the agent distinguishes between meanings 1, 2, and $[3 \ldots 4]$. If the referent is 3 and the agent comes up with the closest meaning, i.e. $[3,4]$, the situation is only brought down to the set $\{3,4\}$, and hence not uniquely determined. Given a healthy conceptual system, unsuccessful determination indicates that the system of communication is not functioning properly. In any case, a high correctness measure, which can only have been caused by successful determinations, implies that communication is functioning well.

It is important to note a possible problem with the correctness measure as it has been described so far. Suppose that a new situation arrives, but that the agent's signal based situation determination is incorrect. In this case, the reward following the action will probably not correspond to the agent's estimate of it, in which case the agent decreases the association-success of one of the signals it received (see algorithm, fig. 3.3). At the next time step, the probability that the agent will select the same situation again based on the signals is a little bit less likely, because of the drop in association strength. Consequently, the probabilities for the remaining situations, including the correct one, have risen.

Although this plasticity of the agent's communication system is essential, it obfuscates measurement of the quality of communication. To avoid this possible source of bias, measurements are only made at time steps where a new situation has just arrived.

3.2 Experiments

The mechanism for adaptation of associations is the result of investigating many different mechanisms. Some principles and components that have been considered during this investigation turned out to be crucial, others ineffective. The algorithm that has been described is the result of this process. In the following, experiments will be reported where different aspects of the mechanism have been left out. The effects of these omissions will be discussed. These experiments are intended to give insight into the complete mechanism and the function of its components. Following that, the results of experiments with the complete mechanisms are presented. The section titles are augmented with a code (`sit`,`succ`, etc.) that identifies the experiment in the graphs of this section.

3.2.1 Signal-Based Situation Determination - `sit`

As the algorithm shows, agents sometimes determine in which situation they are on the basis of sensor information, while at other times this decision is determined by the signals the agent received. In the `sit` experiment, it is investigated whether the latter method, which will be termed *signal-based situation determination*, is necessary for the development of communication.

3.2.2 Success Component in Association Strength - `succ`

Associations consist of a success component and a use component. The purpose of this experiment is to test whether the *success* component is essential, or whether the use component would be sufficient by itself. The resulting algorithm is very similar to that of the previous experiment, but there is a subtle difference in the scaling of the association strengths; whereas in the previous experiment the success values were not adapted, and hence retained their initial values, in the current experiment the range for association strength values is devoted to the use component.

3.2.3 Use Component in Association Strength - `use`

Analogously to the previous experiment, it may be questioned whether it is necessary to include a use component in the association strength. The answer to this question is not only technical. Whereas it might well be possible to develop a coherent system of communication based solely on success, i.e. testing whether signal-based situation determination leads to the expected effects of actions, this does not appear a very efficient model for the development of communication. It is intuitively clear that a wealth of information is discarded if the use of words is not taken into account, especially since use information comes 'for free' whereas success information can be costly in cases where situation determination failed. It is interesting to see whether these considerations can be confirmed experimentally.

3.2.4 max-error-for-signals - `maxerr`

This experiment investigates the need for the condition that action values must be accurate before signal-based situation determination may be used. Signal-based situation determination only takes place when the average error in the action value estimate has dropped below this threshold, given by the **max-error-for-signals** parameter. The idea behind this principle is that as long as the estimate of action values are not accurate, the deviation of the actual action evaluation from the expected value should not be used for determining the success of signal-based situation determination. The need for the threshold is tested by removing this condition. The experiment only concerns the described use of the **max-error-for-signals** (line 7 of the algorithm), and not its use in comparing the received reward to its estimate (line 20).

3.2.5 Probability of Signal-based Situation Determination - `Psit`

The *max-error-for-signals* parameter regulates when the signals *may* be used for situation-determination. But even when the average action value estimation error does not exceed this threshold, situation determination is not always based on signals. Rather, as the algorithm shows, this option is chosen with a probability equal to the conditional probability of the situation given the signals, weighted by the frequency of each signal. In this experiment, it is investigated whether it is necessary to use this probability, or whether always using signal-based situation determination would also do the job. Since the use component of the associations is not updated for signal-based situation determination, the effect on the algorithm is similar to that of the `use` variation, with the difference that the use *is* updated when the average (line 7) or current (line 20) action value estimation error exceeds the **max-error-for-signals** parameter.

3.2.6 Decrease Association Success - `decsucc`

As the algorithm shows, successful signal-based situation determination not only causes the success value of the signal most influential in the choice to be increased (line 21 of the algorithm), but also results in a decrease of the success values of associations of the signal with other situations (line 23). In this experiment, the need for this mechanism is tested. This test is performed by removing the decrease statement from the algorithm.

3.2.7 Results

This section presents the results of the experiments described above. The experimental system is a complex dynamical system. In such systems, it is seldom possible to identify a single cause for phenomena that are observed. This is because the large number of interactions between the

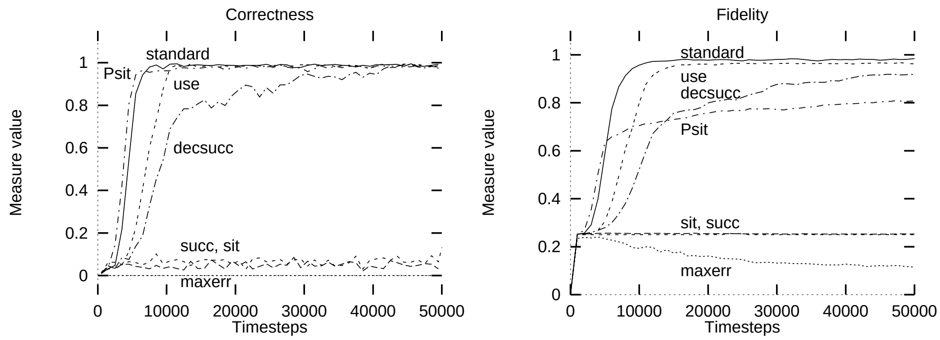


Figure 3.4: Correctness (left) and fidelity (right) measures for the standard experiment and its variations.

parts of such a system allow subtle changes to propagate in ways that are difficult to envision, possibly causing large changes in the overall behavior of the system. For the same reason, it can be difficult to predict or explain the effects of changes of a particular aspect of the system. In this section, the results of the experiments will be shown, and the differences caused by the changes to the basic algorithm will be explained where possible.

Figure 3.4 (left) shows the correctness measure for the standard version of the algorithm and its modifications. Since correctness expresses whether communication leads to correct determinations of the situation, it is an important measure of the quality of the communication system. Several of the modifications cause substantial harm to the workings of the system. Only the `use` and `Psit` modifications yield results that are comparable to the basic algorithm.

If correctness is high, this implies that agents can correctly determine their own situation most of the time. For this to happen, it is in principle sufficient to have one agent that always announces the correct situation, provided that all agents understand this agent's language and, moreover, are able to select that agent's signal for situation determination. A stricter measure of quality therefore is the extent to which *all* agents can understand each other, i.e. will decode the same referent that was encoded by another agent. The fidelity measure expresses exactly this, for all pairs of agents.

Figure 3.4 (right) show the fidelity for the standard algorithm and its modifications. This stricter evaluation of the different methods brings to light differences between the variations that could not be detected from the correctness measure. Specifically, the graph shows that the standard algorithm brings about better systems of communication than all of its variations. This indicates that all components of the algorithm that have been tested contribute to the development of communication. Two cautions are in place here. First, it is necessary to determine which level of statistical significance the differences have. Second, even if significant, the results only hold for this particular experiment, consisting of the problem environment and the parameter settings. It is to be expected, and has been observed in actual experiments, that the variations have different effects for different parameter settings. The conclusions that a component is useful or necessary can only be made if its omission is compared to the standard algorithm for every possible parameter settings. Since some of the parameters are real valued and since there is a substantial number of parameters, this form of exhaustive test is impossible to carry out both practically and theoretically. Thus, the only type of statement that can be made is that a particular component of the algorithm *can* be useful. It is these questions that we will attempt to answer here.

3.2.8 Statistical Significance of the Results

In the above, it was mentioned that the statistical significance of the results needs to be demonstrated before any strong claim concerning the need for particular components can be made. The experiments all had the following structure: a component of the algorithm is removed, and the performance is compared to that of the basic system to determine whether it decreases. As has been shown above, this was the case for all components. However, the differences might be the result of random variations in the course of the experiments. For that reason, it is important to test the statistical significance of the results.

As the graph in figure 3.4 shows, fidelity is not a constant measure, but rather develops over time. One possibility for testing statistical significance of differences in the fidelity measure would be to consider its value at the end of the experiment. However, the value at a particular point is not what one wants to know. The desirable property of a system of communication is not that it's fidelity is high at some arbitrarily chosen point in time, but that communication develops quickly and reliably. A better measure therefore is the average fidelity of the system over time. Since this measure takes into account the behavior of the system over the complete period of the experiment, it is a more stable indication of how the development of communication proceeded.

In figure 3.5, a plot is shown where at each time step the average fidelity of the system up to the point of measurement is given. As could already

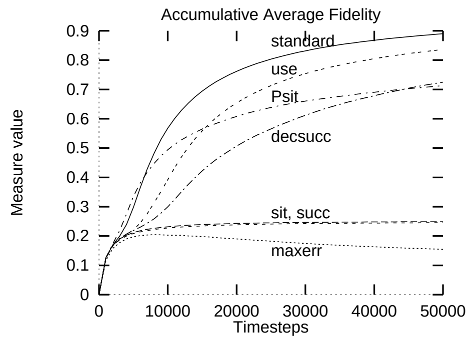


Figure 3.5: Fidelity for the standard experiment and its variations, averaged up to the point in time indicated by the horizontal axis.

be deduced from the time series graph of fidelity (figure 3.4), the standard algorithm has the highest average fidelity. The question to be answered now for each component of the algorithm is whether the final value in this graph, which is the average fidelity over the complete experiment, is significantly higher with the component than without it.

Each experiment has been repeated ten times. Each of the ten measurements is determined by the behavior of a dynamical system under different initial conditions. There is no good reason to assume that such a set of measurements is normally distributed. As an the extreme example of this, one could imagine a system that, depending on its initial conditions, either moves towards an attractor where the measure of interest is high, or to an attractor where it is low. The results for such a system are binomially distributed, not normally distributed. Although the system at hand is probably not of this type, the example goes to show that the assumption of normality cannot be made. Indeed, in the case of high measure values (near 1), the limited range of the measures resulted in highly skewed distributions, and thus tests based on this assumption can not be used here.

The Wilcoxon Rank Sum Test

Since the statistical distribution of the measurements is not known, it is necessary to resort to a distribution-free statistical, or nonparametric tests. In general, such tests are less powerful than tests based on some distribution, which is only logical since the latter poses a condition on the data, although in rank-sum tests the loss is often very small. A test that is suitable for our purposes is the Wilcoxon rank-sum test, see e.g. (Lehmann &

D'Abrera, 1975), also called *Mann-Whitney U Test.* This test determines whether a treatment is effective. To determine the necessity of a component of the algorithm, the test can be applied by determining whether removing the component is detrimental, i.e. the treatment has a negative effect. Conversely, the addition of a component can be viewed as the treatment, in which case it is tested whether the treatment has a positive effect. The test is symmetrical after a normalizing by subtracting the minimum rank of $\frac{1}{2}n(n+1)$, and therefore it is not necessary to distinguish between these two views here.

The workings of the Wilcoxon rank-sum test, applied to the current situation, are as follows. The measurements of two versions of the algorithm are ranked, the lowest result receiving a rank of 1, and each subsequent result having a rank that is 1 higher than its predecessor. Thus, high ranks correspond to high performance. The ranks of one of the versions are then summed. The result of this calculation is the rank-sum, to which the test thanks its name. If the rank-sum of the control is low enough, the null-hypothesis, i.e. that the distributions of the measurements of the two versions are equal, can be rejected, and the superiority of the algorithm with the highest rank sum is confirmed (Lehmann & D'Abrera, 1975). Text books on nonparametric statistics contain tables of the significance level with which the null-hypothesis can be rejected for particular combinations of the numbers of examples in both sets and the rank-sum. For the current investigation, the tables from (Lehmann & D'Abrera, 1975) have been used.

As the actual measurement values are discarded in the calculation of the rank sum, the interpretation of the test can not be stated in terms of a difference in the mean, unless the additional restriction is imposed that the distributions are identical apart from a shift in the mean as is sometimes done. Instead, the outcome of the Wilcoxon rank-sum test can be interpreted in terms of pairwise comparisons between measures of the two samples. When corrected with the above term, the rank sum of a measurement equals the number of comparisons with the measurements of the other sample that it would win. Under the null hypothesis of equal distributions, these pairwise comparisons are equally likely to turn to either side. However, if one of the algorithms performs better than the other, this algorithm is more likely to outperform a run of the other algorithm than not, and hence the majority of the pairwise comparisons are expected to be won by it. It is this aspect of the algorithms that is tested by the Wilcoxon rank-sum test.

Figure 3.6 shows the error bars (minimum and maximum values) for the

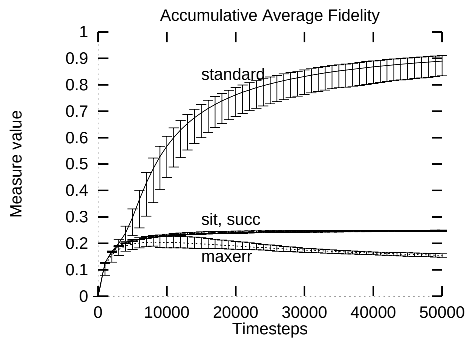


Figure 3.6: Error bars of average fidelity for `sit`, `succ`, and `maxerr` variations (left) and the `Psit` variation.

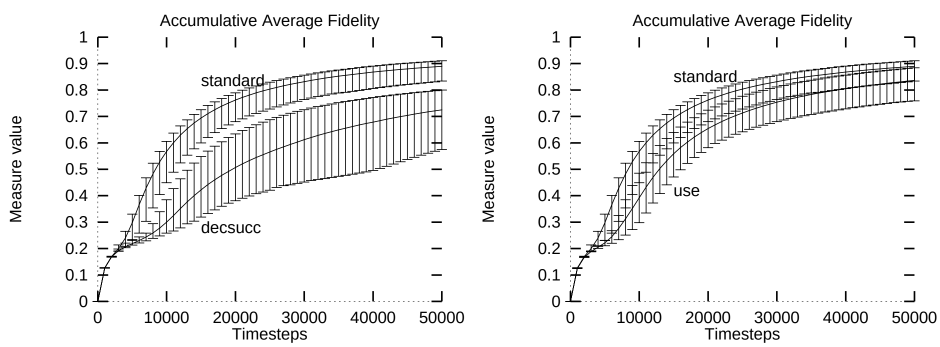


Figure 3.7: Error bars of average fidelity for the `decsucc` variation (left) and `use` variation (right).

`sit`, `succ`, and `maxerr` variations. An interesting phenomenon is the fact that the height of the error bars tends to decrease over time. Apparently, the initial differences between the different runs stabilize in the course of the experiment.

A more important observation is that the rightmost error bar of `sit`, `succ`, and `maxerr`, each showing the minimum and maximum value over ten runs of the average fidelity over the experiment, is lower than the corresponding error bar for the standard algorithm. This implies that the lowest value for `standard` was higher than the highest value of `sit`, `succ`, and `maxerr`. Therefore, the ranks of the `standard` runs are $[11, 20]$, and those of the `sit`, `succ`, and `maxerr` runs are $[1, 10]$, and so the rank-sums of the variations are all $1+2+\ldots+10 = 55$. The table in (Lehmann & D'Abrera, 1975) yields a significance level of $\alpha = 0.0000$ for this case, which means that the difference between `standard` and each of these three variations is statistically significant at this level of significance.

Figure 3.6 and 3.7 show the error bars of the `Psit` and `decsucc` experiments. For both of these experiments, the complete final error bar is below that of the standard case. Even though the difference is smaller here, this property yields the same conclusion as for the first three variations, i.e. the difference between `standard` and `Psit` and `decsucc` is significant at a level of $\alpha = 0.0000$.

For the `use` variation, where use information plays no role in the association strengths, the situation is different; here the error bars overlap (see figure 3.7). Therefore, the ranks have to be computed by sorting the values of the 10+10 runs of `standard` and `use`. Table 3.3 shows the sorted values and their ranks. The rank-sum of the `use` variation is 67. The significance level of the statement that there is a difference between the two distributions is 0.0014, and so it may be safely said that the `use` component is useful.

3.2.9 Utility of the Use Component

In the above experiment it was already demonstrated that the use component of associations can be useful, even though its advantage in terms of performance was small. However, the performance issue is not the only reason for including it in the algorithm. An important role of use information could be that it reduces the necessity for obtaining success information. Since the signals are initially unreliable, signal-based situation determination involves an increased risk of selecting the wrong situation and, consequently, an action that is based on the wrong situation. This possibility

	Rank	With use comp.	Rank	Without use comp.
	20	0.910807	13	0.883869
	19	0.91	12	0.875613
	18	0.908599	10	0.873765
	17	0.905289	9	0.869088
	16	0.90134	7	0.851777
	15	0.901281	6	0.847442
	14	0.889206	4	0.819828
	11	0.874593	3	0.791192
	8	0.864297	2	0.790466
	5	0.833662	1	0.759108
Rank-sum	143		67	

Table 3.3: Average fidelity for ten runs of the algorithm with and without the use component. The runs have been ranked based on this measure, and the sum of the ranks is used as a statistic for the Wilcoxon rank-sum test (see text).

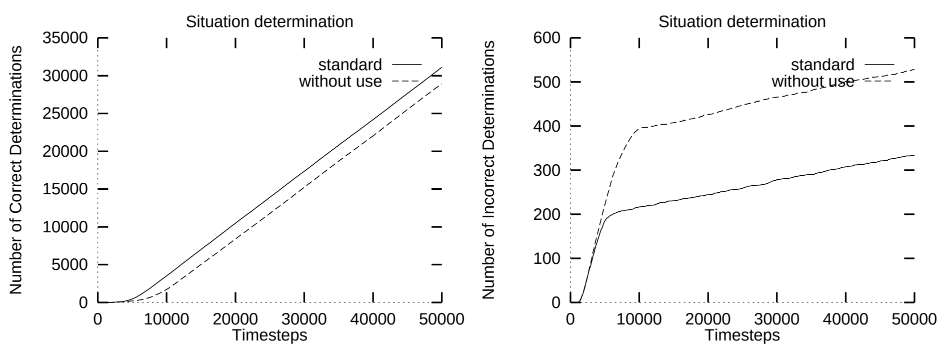


Figure 3.8: Number of correct (left) and incorrect (right) signal-based situation identifications for the `standard` and `use` experiments. The graphs show that the use component increases the number correct determinations *and* reduces the number of incorrect determinations.

has been investigated by determining the number of occasions on which the agent successfully determined the situation based on signals and occasions on which the signal-based situation determination was incorrect. By comparing these numbers for the standard algorithm and the `use` variation, it can be determined whether use information reduces the proportion of costly wrong signal-based situation determinations.

Figure 3.8 shows the number of correct and incorrect signal-based situation determinations for the `standard` and `use` experiments. The graphs show that the use component not only increases the number of correct identifications, but also decreases the number of incorrect determinations. These numbers are not merely complementary; the total number of signal-based situation determinations is not fixed, but is determined by the agent's confidence in the signal, since signal-based situation determination is selected with a probability equal to the highest situation probability given a signal.

The statistical significance of this result has been tested. The previously described Wilcoxon rank-sum test yielded a significance level of $\alpha = 0.0177$ for the higher number of correct determinations and for the lower number of incorrect determinations even $\alpha = 0.0000$, and hence the component's role of reducing the number of harmful situation determinations and increasing the number of correct ones has been demonstrated.

A further argument for the use component is that the specificity, consistency and coherence all develop earlier and to higher final values with the component than without it, see figures 3.11 and 3.12. The hypothesized reason for this, is that although the information conveyed by use is of less value for the development of communication than success information, its frequency is much higher, since it is available after every time step. Together with the above considerations, this provides a strong enough case for the inclusion of use information in the model.

3.2.10 Analysis of the Success Component

One of the arguments for the use component was that it reduces the need for expensive success information. This leads to the question of *why* the success component of the algorithm was found to be indispensable. A possible explanation for the prominence of success information is the following. Let us consider the bias caused by agents with respect to the occurrence frequency of words. Each agent estimates the occurrence frequency of words for all situations, and produces words with probabilities proportional to these frequencies. Due to the soft-max choice (see section 3.1.4), this proportional

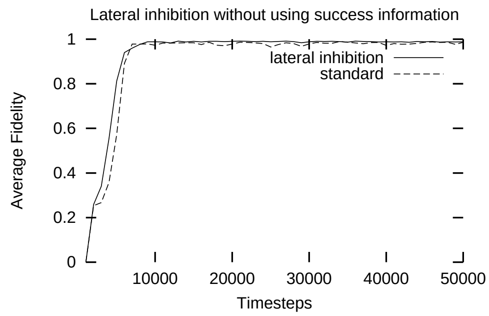


Figure 3.9: Average fidelity for the standard experiment compared to the variation where the lateral inhibition of the success component is maintained, but increases are not dependent on success anymore. In combination with earlier results, the graph shows that the lateral inhibition aspect, not the success information, is instrumental in achieving specificity.

relationship is nonlinear, with a preference for frequent words. This results in a positive feedback loop; words that are used frequently will tend to be used even more often, and thus one may expect convergence towards using a *consistent* and *coherent* system of communication. However, *specificity*, the property that different concepts are referred to with different words, is not guaranteed by this setup. Indeed, the `sit` and `succ` experiments point out that using only use information in the associations, and hence not taking into account success values, results in communication systems with moderate or high coherence, but very low specificity measures.

Whereas use information is biased towards words that are a little more frequent than other words, the effect of the success component in association strengths is different. This value is increased when a word gave rise to a correct situation identification, and decreases for incorrect determinations. This introduces a different kind of bias. Specifically, it reinforces words that are strongly associated with only a single situation, since situation determinations based on such words have a high chance of being correct. The algorithm ties such words to the situation even stronger. On the other hand, words that are associated with several situations are less specific and will therefore not be selected as often. In addition, *if* they are selected, they stand a high chance of yielding an incorrect situation determinations, since only one of the situations can be correct. The success component in association strengths thus punishes such words. The combined effect of these processes is that the system of communication will tend towards a state where words are only associated with a single situation and, hence, systems with a high specificity. Since specificity is a necessary condition for fidelity, this effect explains why the success component is necessary in order to bring about communication.

A close look at the algorithm shows that the success-component of the associations fulfills two functions. First, it represents information about the ability of a received word to distinguish between different situations. This is achieved by determining whether the reward received after the action was selected corresponds to its estimate. Second, it implements *lateral inhibition* between one association and other associations of the same word with other meanings. In the following, the former will be referred to as the *success information*, the latter as *lateral inhibition*.

There is a possibility that it is not the success information *itself* that is required to achieve specificity, but rather the lateral inhibition. As was already remarked, if one situation is linked more strongly to a word than other situations, the effect of the success component is to enlarge this difference. This effect might even persists when no distinction between correct and incorrect situation determinations is made. This can be seen as follows. The word leading to a situation determination must already be somewhat more specific than other words uttered; the very fact that it was used in the situation determination procedure indicates that this word has a high conditional probability $P(\mu|\sigma)$. The effect of lateral inhibition is that the success component of this association increases and that of the associations with other meanings decreases. The effect of this on production behavior is that the specificity of the word in identifying the situation increases. Given time, this change then also affects reception behavior due to the changes in reception probabilities. Hence, the increase in specificity that is necessary for successful communication could in principle be provided by the lateral inhibition, even if this procedure does not employ success information.

To determine whether this is the case, two experiments have been performed: one where success is used without lateral inhibition, and one where lateral inhibition is used but not success. The first of these experiments is achieved by removing line 23 from the algorithm, which corresponds precisely to the **decsucc** variation. In the **decsucc** experiment, performance was lower than normal, but successful communication still developed.

The other experiment is to use lateral inhibition, but not success. One way to obtain such a system is to remove the second term of the condition on line 20 of the algorithm; the effect is that the success component of associations is not increased when the reward corresponds to its estimate and decreased when it doesn't, but that after *each* signal-based situation determination this component is increased for the selected word and meaning. The accumulative average fidelity of this experiment is shown in figure 3.9. As the graph shows, the performance in this experiment is at least as high as in the standard case, even a fraction higher.

We now move on to the interpretation of these results. What the second experiment proves is that success information is not required under these experimental conditions; by removing the specified condition, no information about the accuracy of the action value estimations is used in the association adaptation mechanism, while communication still develops at least as successfully. The implications of the first experiment are less crisp; whereas the most obvious source of lateral inhibition between meanings has been removed, increasing the success component of the associations involved in the signal-based situation determination indirectly also has the effect of enlarging existing differences between the associations, thus increasing specificity. This can be seen as follows. The higher association values cause the corresponding words to be uttered more frequently in the corresponding situation. This decreases the relative conditional probability of being in *another* situation given that word is received, and thus the success components of these associations of other meanings with the word will be increased less often. Therefore, this can be viewed as an indirect form of lateral inhibition. The conclusion is that lateral inhibition, not success, is responsible for the necessity of the success component of the algorithm.

To determine whether success information can be of use when differences between the conceptual systems of agents are stimulated by means of noise, additional experiments have been carried out. These experiments did not contradict the above conclusion. Furthermore, experiments have been performed where the hearing range of the agents was decreased from a radius of 5 to 4, 3, and 2, causing the utterances to be received by only a subset of the agents rather than all agents as in the standard experiment. Again, the results did not require a revision of the conclusions, as the changes had hardly noticeable effects on the success of communication in both the standard case and the case where no success information is used.

It is interesting to compare these results to the work of Oliphant (1997). As noted, Oliphant does not distinguish between meanings and referents, but assumes that all agents have access to a public set of concepts. One of his central findings was also that a bias is required to induce specificity. In Oliphant's Bayesian learner, this is achieved by observing the communica-

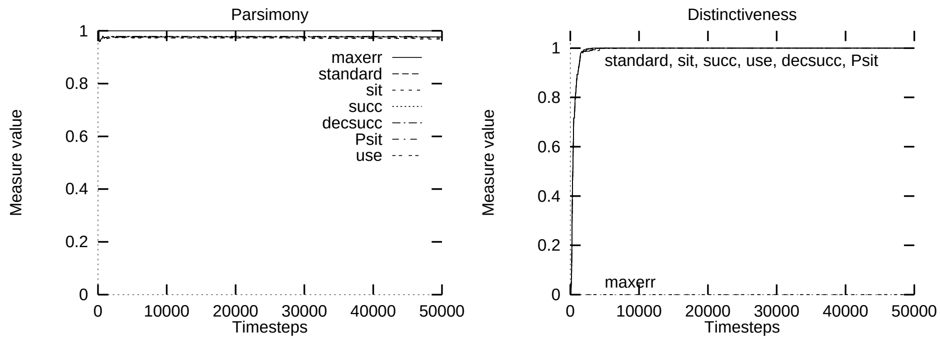


Figure 3.10: Parsimony (left) and distinctiveness (right) measures for the standard experiment and its variations.

tive behavior of the population and using this to set the communicative behavior of a selected individual to a matrix which for each meaning sends the word best understood by the population, and interprets each signal as the meaning for which the word is produced most often. All other entries of the matrices are set to zero. In the current context, this mechanism would exclude stable systems of communication with agents that have multiple meanings associated with a single word. In view of the idea that conceptual systems may differ without disabling the potential for communication, this procedure can not be used in the current context. Although the two approaches share the objective of increasing the specificity of words, the gradual updates of the algorithm presented here offer a potential for shared communication based on differing conceptual systems. This idea will be explored in detail at the end of this chapter.

3.2.11 Discussion of the Remaining Variations

Figure 3.10 shows measures describing concept formation. Parsimony is continuously high for the standard algorithm and all of its variations. Distinctiveness quickly rises to its optimum value of one for all but one of the experiments; for `maxerr`, distinctiveness remains zero. This implies that the knowledge of which meaning is present yields no information whatsoever about the actual situation, and hence that no useful concepts have been developed. The disastrous effect of not using the **max-error** threshold can be explained by inspecting the proportion between signal-based and sensor-based situation determinations; right from the beginning, the agent almost solely depends on its signals in determining the situation, see figure

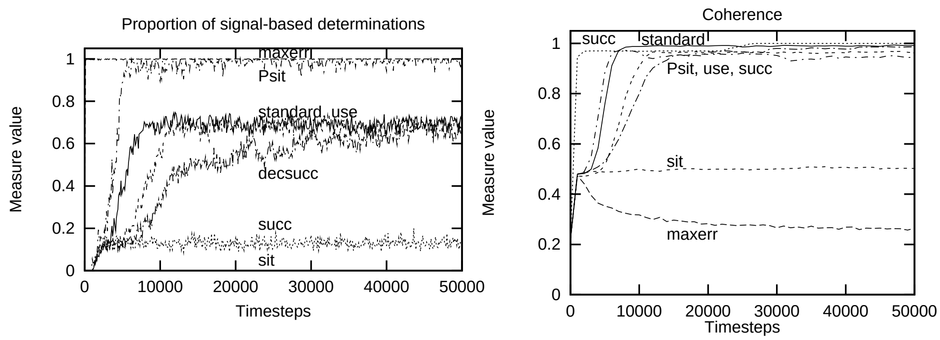


Figure 3.11: Left: fraction of signal-based situation determinations. Right: coherence.

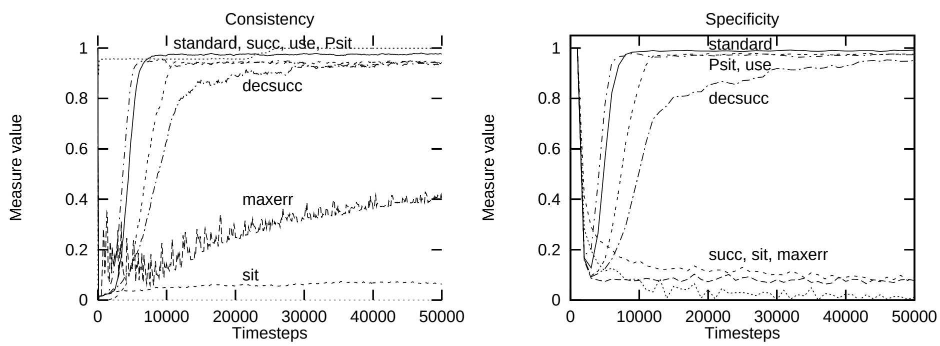


Figure 3.12: Consistency (left) and specificity (right) measures for the standard experiment and its variations.

3.11. This prevents it from gaining experience with its sensor data, which causes the breakdown of the concept formation process. Apparently, in this experimental setup, it is important that the agents do not rely on communication before they have formed concepts for interpreting communication. This result may be viewed as an example where the coupling between concept formation and association formation, as depicted in the introduction (figure 1.2), is dysfunctional. The problem is not with concept formation nor with association formation itself, but in the interaction between these two processes.

Figure 3.11 (right) shows the coherence of the experiments. The standard algorithm and most of its variations achieve a high coherence. One variation however (`succ`) has a coherence that ends up slightly higher than the standard algorithm. As has been explained in chapter 2 however, a high coherence on its own is no guarantee for good communication. The specificity graph (figure 3.12 right) shows that for the same variation, specificity is very low. Thus, the communication systems that developed during the runs of the `succ` variation is not a very useful one; it consists of agents that agree on what words to use, but different referents are referred to with the same word, and so their language is not successful at conveying information. The concept formation measures shown above (figure 3.10) reveal that this lack of distinctiveness is not caused by the concept formation process. Apparently, the success component is crucial in obtaining good communication. Furthermore, the prevalence of the standard algorithm that was reflected in the fidelity measure is found back in the two communication specific measures shown here.

An interesting additional observation is that the specificity graph bears similarities to the fidelity graph of figure 3.4. In several cases (`standard`, `use`, and `decsucc`), the shapes and relative positions of the curves are very similar. Furthermore, together with the `Psit` variation, these experiments all have a high specificity whereas the other three experiments (`sit`, `succ`, and `maxerr`) all result in low specificity. The fidelity graph also shows this pattern. This suggests that specificity is a good indicator of fidelity. For the `Psit` modification however, specificity is on the positive side when viewed as a fidelity indicator. The earlier increase of the `Psit` modification as compared to the standard experiment is explained by its higher fraction of signal-based situation determinations (the left graph in figure 3.11), which yields more success information.

An intriguing phenomenon concerning `Psit` is that although its communication measures (specificity, consistency and coherence) are all approximately as high as the standard experiment, its fidelity was notably lower (fig. 3.4). The large fraction of signal-based situation determinations, and consequently low fraction of sensor-based ones, also explains this. Since use information is only update after sensor-based situation determinations, this information is not updated very frequently in the `Psit` variation. The low fidelity is a result of these inaccuracies. An example is provided by agent no. 1, which had the following interpretation probabilities $P_{int(\mu|\sigma)}$ of the word `fu` for its four meanings: 0.2, 0.55, 0.33, and 0.1. Its own production probabilities of the word `fu` for the same meanings were 0.00005, 0.998, 0.00011, and 0.00012. Clearly, its estimates of word use had not converged yet, and indeed the fidelity graph is still slowly rising. Nonetheless, it communicative behavior concerning the word is accurate; its highest interpretation probability is associated with the meaning for which the word is most often produced, and hence it will always select that meaning for the interpretation of the word.

For the three experiments with low specificity (`sit`, `succ`, and `maxerr`), the measure is lower than fidelity. This can be explained from the different nature of the measures. Whereas specificity measures a reduction in uncertainty, fidelity measure the probability of making the right choice. Since there is only a limited number of choices, the fidelity of a communication system that allows almost no transmission of information will still have an expected fidelity of at least 1 over the relative frequency of the least frequently occurring referent. This explanation is only valid when the agents have the same set of words associated with their meanings; this explains the exception of the `maxerr` variation in which, as will be seen below, the associated set of words of the agents continually changed.

The consistency graph, see figure 3.12, shows several interesting phenomena. First, the high value of this measure for the `succ` experiment is reminiscent of the coherence graph. In itself this should not be surprising, since consistency is a necessary condition for coherence. An inspection of signal production revealed that in all of the ten repetitions, all agents used a particular word for all of their meanings with probability 1 at the end of the experiment. This finding is consistent with the results concerning specificity, where the experiment ranked lowest and appears to converge towards zero. As the distinctiveness measures showed, this can not be the result of a lack of appropriate concepts. The characteristic property of this experiment is that no success information is used. Apparently then, the success component is crucial for the development of specificity. This is understandable, since success values are one source of differentiation between the word associations of different meanings.

Another striking fact concerning consistency is that for `maxerr` this measure rises only very gradually; apart from the level to which it seems to converge being rather moderate compared to most other variations, the level continues to rise during the whole experiment, whereas other variations rise to 90% of their final level in less than a tenth of this time.

An inspection of the data showed that many words are associated with the single meaning of an agent in the `maxerr` experiment, all with a low association value, in agreement with the low consistency. Combined with the information that choices in this experiment are almost always based on signals, this leads to the hypothesis that unsuccessful signal-based situation determinations, a necessary consequence of the lack of distinctiveness, bring down the success values of the associations, causing the agents to introduce new words when the highest association strength falls below a threshold (0.25, see algorithm). Since the association values of the existing words are low, the new word will have the highest association value and be selected for production, until its association value has decreased by the same principle, and the process repeats itself. Inspection of the association strengths supports this hypothesis. Success values components of the associations varied around 0.1 with only minor variations, and use components around 0.45, which implies that whatever the success-use ratio (0.75 in these experiments), the association strength of a newly introduced word will supersede all existing words and hence be selected for production. The probabilities $P(\sigma|\mu)$ varied around 0.02, which indicates that the number of words associated with a meaning was around 50. This is clearly anomalous in comparison with the standard experiment, where typically four or five different words are used. These findings explain why the signal production of the agents in this experiment is not very consistent. Given this information, the question why the consistency measure for `maxerr` keeps rising in an unnatural manner still demands an answer. However, this is merely an artefact. For practical reasons, the number of words for which the probability matrices are computed is limited to 20. Since the actual number of words exceeds this number in this anomalous case (none of the other variations exhibited this behavior), the uncertainty corresponding to the words not present in the matrix falls outside the calculation, yielding a value higher than the actual consistency. This effect starts to play a role as soon as the number of words exceeds 20 which, judging from the graph, appears to be around 8000 time steps. It is interesting to see that the variations in the value over time appear to have been superimposed onto an independent smooth curve; this observation fits well with the above explanation.

3.3 The Benefit of Communication

It has been demonstrated that communication can be used to convey information, and so it might be said that communication is *useful* for the agents that use it. However, communication only has an overall benefit if the correct situation determinations are not overshadowed by incorrect situation determinations that could have been avoided if no communication was used. In order to investigate whether communication as it is developed in these experiments presents an *overall* benefit to the agents, an additional variation is introduced here which excludes communication from the

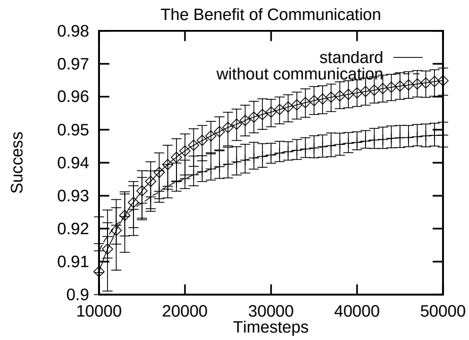


Figure 3.13: The influence of communication on success. At every time step, the graph shows the success obtained up to that point in time averaged over agents and runs.

experiment. The measure that will be used to investigate the systems is the success received by the agents. Figure 3.13 shows the success, averaged up to the point in time on the horizontal axis, for the standard experiment and its new variation. Two aspects are of interest. First and foremost, success is higher when communication is used. In other words, the benefit conferred by the increase in available information is not overshadowed by errors made as a result of the signal based situation determinations that are performed as part of the communication development process. Second, the success of the system with communication is initially lower than that of the non-communicative system, but supersedes it after some time. The initial lower success is explained by the costly incorrect signal-based situation determinations that are necessary to build up communication in this experiment. Once a good system of communication exists, its fruits can be reaped in the form of a sustained higher success value.

The statistical significance of the results has been assessed as follows. The basis of the test is the average success per time step obtained by all agents up to the point in time indicated on the horizontal axis. The error bars in the graph show the minimum and maximum of this value over the ten runs at 50 equidistant time steps. The lines show the average of the value over all runs. The rightmost two error bars do not have any overlap. As noted before, for the number of runs used, this yields a significance level of $\alpha = 0.0000$ in the Wilcoxon rank test. Thus, agents that developed successful communication in the experiments produce significantly better *behavior*, as expressed by the evaluations they receive, than agents that do not.

3.4 Private Concepts, Public Language

A central question in this thesis is that of how agents can come to use the same words for the same things, even though their experiences are different. The previous chapter already suggested a partial explanation: since each agent generalizes over its own experiences, the concepts that one agent forms may be very similar to that of another agent. The current chapter suggests a second part of the explanation. This complementary explanation is the idea that even if agents have formed different meanings based on their individual experiences in the environment, they may associate these private concepts with words in such a way that they use the same words in the same situation. Thus, the differences in their conceptual systems are internal, and, although they may determine the behavior of the agent to a large extent, the existence of such differences is not manifested in the communicative behavior of the agents.

3.4.1 Experimental Setup

To investigate whether such a scenario is feasible in practice, the communication experiment as it has been described thus far will be slightly modified. In the basic version of the experiment, the distinctiveness and parsimony measures are often both equal to one, implying that the conceptual systems of agents are ideal and (therefore) equal. To properly address the question of whether private concepts can be the basis for public language, it is required that the agents develop different conceptual systems. Although the sensory experiences of the agents are already different in the basic experiment, these differences, the generalizing capabilities of the agents evidently are sufficient to overcome these differences. In the following experiments, extra differences are therefore introduced, by means of adding noise to sensory inputs and evaluations.

The specific setup of the experiment is as follows. As in the basic experiment, five agents are present in the environment. The sensory inputs and evaluations of these agents are differentiated by noise as follows: the first agent has noise added to its first sensor dimension, the second to its second sensor dimension, and the third to its third sensor dimension. The fourth agent has no sensor noise, but its evaluations are augmented with noise. The fifth agent finally has no additional sensor or evaluation noise. However, like the other four agents, its information about the environment is not always correct, because of the partial perception of the situation sensor (10%). Furthermore, the minimum number of experiences that is required in the two halves of a region for a split to be considered is decreased from 50 to 25, in order to further increase the likelihood of extra distinctions.

3.4.2 Calculation of Matrices Based on Sampling

In section 2.6.3, it was explained that the concept formation measures as they have been described so far measure to what extent the ideal conceptual system for the basic experiment is formed. For determining the effect of noise on the experiment, this is interesting information. In this section however, a modification of the basic experiment is investigated. The question thus becomes to what extent the conceptual and communication systems formed by the agents are appropriate for this modified task. To determine this, it is not sufficient anymore to map the conceptual systems onto the ideal conceptual system and thus determine the matrices $P(\mu|\rho)$ and $P(\rho|\mu)$. The reason for this is that the original referents, defined as regions in sensor space, no longer exactly correspond to the regions where particular behaviors are appropriate. Rather, the boundaries of these regions become blurred as a result of sensor noise, and the distinctions between different regions become less clear as a result of evaluation noise. Because of these effects, the ideal referents can no longer be easily identified by corresponding sensor values, and hence the direct calculation of the parsimony and distinctiveness measures reflects to what extent the conceptual system of the original problem is formed rather than to what extent ideal concepts for the new problem are formed.

Fortunately however, the direct approach is not the only way to calculate the concept formation measures. Although the referents no longer have a one to one correspondence to sensor regions, they are known at each point in time, because the internal state of the simulation, and hence the environment, is accessible to the investigator. Moreover, for the same reason, the current meaning of each agent is available at each time step. Together, these bits of information are sufficient to dynamically estimate the co-occurrence probability of every combination of meaning and referent. Thus, provided that the system is observed for a sufficiently long interval of time, accurate estimates of the probability matrices $P(\mu|\rho)$ and $P(\rho|\mu)$ can be made. This procedure will be referred to as the *sampling* method, and it is important to distinguish the resulting measures from their direct counterparts.

The above distinction between the direct and sampling methods only referred to concept formation matrices, viz. $P(\mu|\rho)$ and $P(\rho|\mu)$. However, since communication measures depend on concept formation, these are also affected by a change in the environment. Specifically, consistency is based on the entropy in the $P_{prod}(\sigma|\rho)$ matrix, which was calculated multiplying $P(\mu|\rho)$ with $P|(prod)(\sigma|\mu)$, and hence depends on $P(\mu|\rho)$. Likewise, in the direct calculation scheme, specificity is based on $P_{prod}(\rho|\sigma)$ and hence depends on $P(\rho|\mu)$.

Since the contents of both of these matrices depend on whether the direct calculation or the sampling method is used, it is necessary to distinguish between direct and sampling versions of consistency and specificity as well. Accurate estimates for these measures can be obtained by directly sampling $P_{prod}(\sigma|\rho)$ and $P_{prod}(\rho|\sigma)$, without using $P(\rho|\mu)$ or $P(\mu|\rho)$. These matrices are still necessary for the concept formation measures though, and for fidelity; fidelity is calculated by multiplying the $P_{prod}(\sigma|\rho)$ matrices with $P_{int}(\rho|\sigma)$ matrices for all combinations of two agents. The latter of these matrices is calculated by multiplication of $P_{int}(\mu|\sigma)$ and $P(\rho|\mu)$, since there is no principled way to directly determine as what referent an agent interprets a given word.

A final note concerns the correctness measure. Since this indicator measures whether the current referent is determined by the agent, its exact determination requires a one-to-one correspondence between meanings and referents. Since this correspondence is lost by the introduction of noise, this measure will not be used in the following experiments. Although an approximation of correctness is possible when the estimates of $P(\rho|\mu)$ are used, there is a more direct way to determine whether the agent has developed the capacity to produce the right behavior in each circumstance. This is simply to see whether it selects the right action. Particularly, we are interested in whether communication allows the agents to overcome their imperfect sensor information. By counting the occasions on which sensor information was misleading but the right action was selected nonetheless, the degree to which communication reduces uncertainty can be measured. To avoid possible bias, as was discussed in relation with the correctness measure (see section 3.1.5), this information must only be registered at the first time step after the advent of a new situation.

3.4.3 Results of the Private Concepts Experiment

In this section the result of the private concepts experiment are presented. The central question here is whether a public language can emerge when agent have different conceptual systems. The experiment as it was described at the beginning of this section has been repeated tenfold. For each run, the conceptual system of every agent has been tracked over time. A requirement for the experiment is that agents develop different conceptual systems. Investigation of the data made clear that this is indeed the case. Often, several agents in an experiment developed equivalent conceptual systems, but in no experiment did all agents develop the same conceptual system. Figure 3.14 show some examples of conceptual systems that were developed by the agents.

The top right system, with six distinctions in the first sensor dimension, and several variants of it, was often developed by the first agent. This can be explained from the distribution of noise over the agents as follows. Without any noise, the ideal conceptual system is the upper left graph of figure 3.14. In this system, the intervals $[-1, 0>$, $[0, 1>$, $[1, 2>$ and $[2, 3>$ are distinguished. The first agent however receives noise on its first sensor dimensions. For example, an initial value of 2 will be received by the agent as $2 \pm \delta$, where δ is drawn from a $\mathcal{N}(0, \sigma)$ distribution. For $0 \leq \delta < 1$, the sensor value falls within the interval $[2, 3>$, and no problems arise. If $\delta \geq 1$, as happened in many of the concept formation experiments of chapter 2, the sensors do not provide enough information to determine the referent. Since $\sigma = 0.15$ in these experiments however, the probability of this event is negligible. However, for $\delta < 0$, which happens in an expected 50% of all cases, a similar problem arises: if the agent would have developed the 'ideal' conceptual system, its sensors now point to the meaning $[1, 2>$. The net effect of this is that most experiences in the lower halves of the integer intervals ($[-1, -0.5>$, $[0, 0.5>$, etc.) correspond to the referent of the low end (-1, 0, etc.), whereas experiences in the upper halves occur almost solely in the presence of the upper referent (0, 1, etc.). Apparently, the concept formation mechanism is able to detect this correlation, which explains the fact that in all experiments, the first agent introduced 6 distinctions in the first dimension as opposed to the basic three.

Figure 3.15 shows the concept formation measures over time. All results reported concern the sampling version of the measures. The average distinctiveness at the end of the experiment is 0.999, which implies that all agents developed virtually perfectly distinctive conceptual systems in all of the experiments. Parsimony is suboptimal. This is no surprise however; it is a property of the experiment that agents develop different conceptual systems, and since there is only a single ideal conceptual system, suboptimal parsimony is a necessary condition for the experiment.

In figure 3.16, the communication measures are depicted. Specificity is very high. The fact that distinctiveness was high already fulfills a first condition for high specificity. Given high distinctiveness, the requirement for

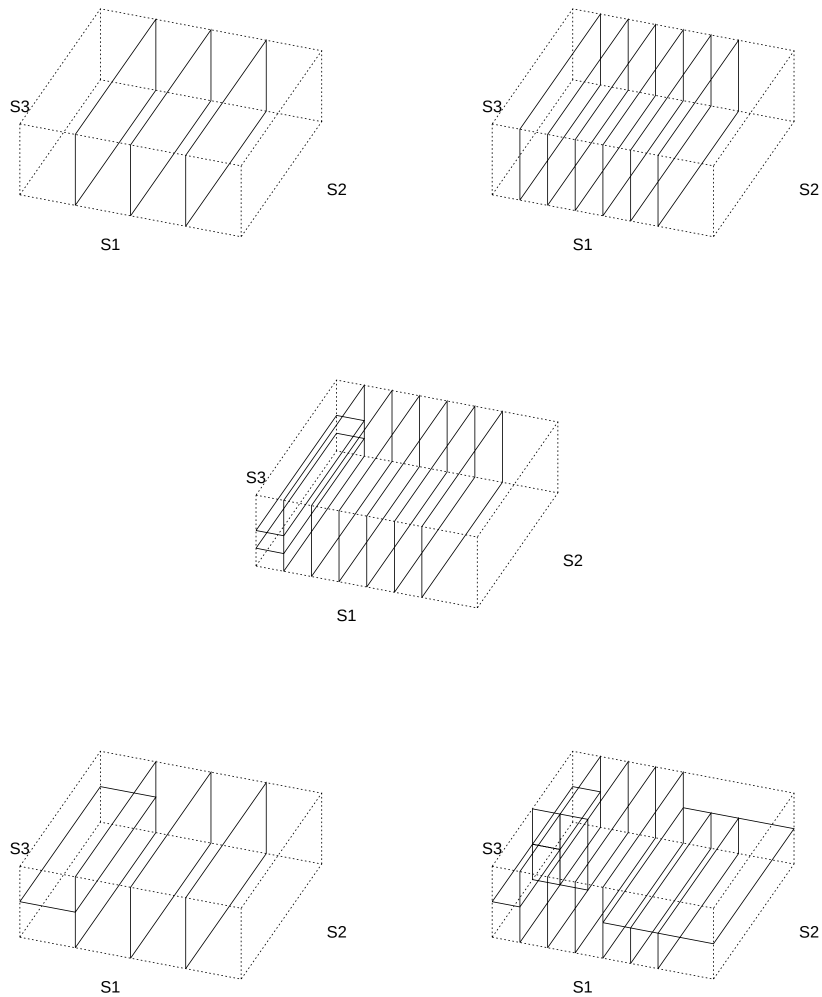


Figure 3.14: Some examples of conceptual systems developed by agents in the experiment. The labels (S1, S2, S3) mark the three sensor dimensions.

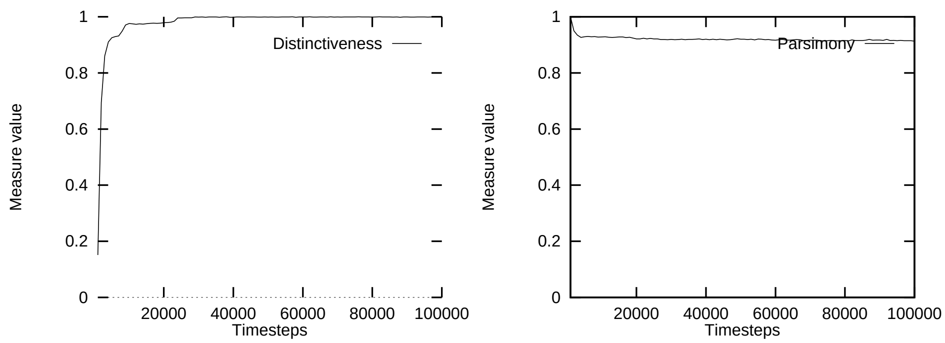


Figure 3.15: Distinctiveness (left) and parsimony (right) in the private concepts experiment.

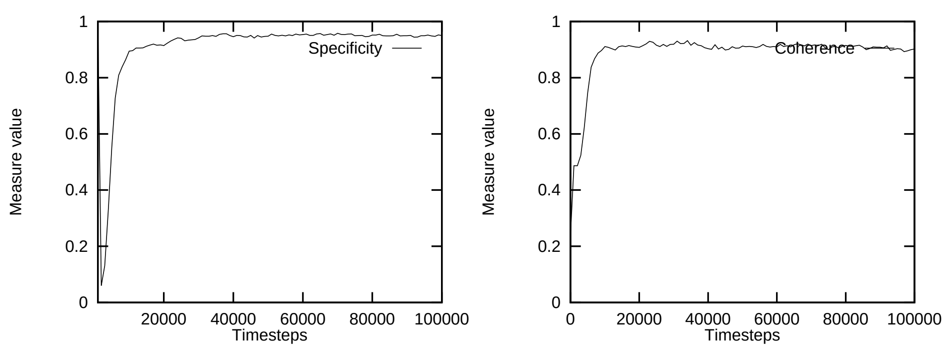
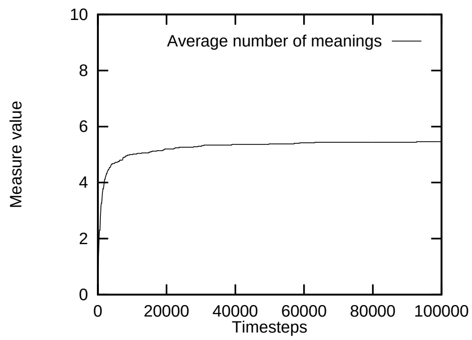


Figure 3.16: Specificity (left) and coherence (right) in the private concepts experiment.



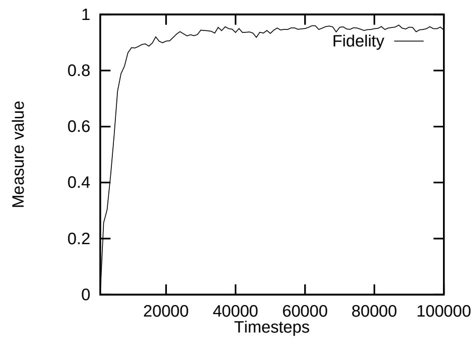


Figure 3.17: Average number of meanings (left) and Fidelity (right) in the private concepts experiment.

high specificity is that the meaning(s) for different referents are associated with different words. The value of the specificity measure shows that this is the case. Coherence is reasonably high, although not as high as possible. This suggests that agents sometimes use several different words for one or more referents. Nevertheless, the fidelity (figure 3.17) is high. Together, this information indicates that although agents do not always use the same word for the same referent, they do understand the words that are produced in connection with each referent. Figure 3.15 shows the average number of meanings that was developed by the agents. This number appears to converge to a value around 5.5, a bit higher than the number of meanings that is required when no noise is present (4).

The measures that have been reported demonstrate that indeed a good system of communication has been developed, even though concept formation is private and often resulted in different conceptual systems. In order to gain some insight into *how* this is possible, a closer look will be taken at the internals of the agents, specifically at the associations between meanings and words that the agents have formed.

The first question to be asked is what conceptual systems are formed. Table 3.4 represents the concept formation matrix $P(\mu|\rho)$. The corresponding graphical representation is shown in the middle graph of figure 3.14. The distinctions in the first sensor dimension can also be read from the headings in the table. As should be expected for the first agent (see explanation above), the conceptual system contains six distinctions in the first sensor dimension. The first interval, $[-1, -0.5>$, represents the first referent. Subsequent referents are represented by two regions, the lower ones ($[-0.5, 0>$, $[0.5, 1>$, and $[1.5, 2>$) corresponding to experiences of which noise the sensor value representing the corresponding referent (0, 1, and 2) was decreased due to the addition of noise, the higher ones corresponding to the cases where sensor noise was positive. The last interval is not split in half; this is because there is no referent 4 which would cause ambiguity when the noise δ would be negative. As would be expected, given that the standard deviation of the noise is 0.15 and will therefore exceed the value of 0.5 with very low probability, the probabilities $P(\mu|\rho)$ are near 0.5 (rightmost 6 columns)[1].

The lowest interval, $[-1, -0.5>$, occurs in the three topmost rows of the $P_{prod}(\sigma|\mu)$ table. These rows represent three different meanings which all have the same range for the first sensor dimension. They are distinguished in the third dimension, and are visible in the figure as the three strips to the left. As would be expected given the fact that there are three possible values of the third sensor dimension, each bearing a one to one correspondence with a meaning, the probabilities in the matrix vary around $\frac{1}{3}$.

Table 3.5 shows the word production probability matrix $P_{prod}(\sigma|\mu)$ for agent number one. A striking observation is that for the first three meanings, all corresponding to the same referent, different words are produced (`ro`, `ra` and `sa` respectively). The association values are all around 0.95, implying that the agent almost always uses the particular word that is associated, and not one of the other two words that correspond to the same referent.

For referent 0 (meanings $[-0.5, 0>$ and $[0, 0.5>$), signal production is not as ambiguous as in the previous case; for both meanings, the word with the highest association (0.709 and 0.999) is `ga`. Referent 1 (meanings $[0.5, 1>$ and $[1, 1.5>$) is represented as `co` during production, and referent 2 finally is represented by meanings $[1.5, 2>$ and $[2, 3>$. The agent refers to these meanings with `za` and `pi`.

The measures reported earlier already revealed that agents start to speak the same language. However, in order to gain more insight into *what* language has been developed, it is necessary to inspect the word interpretation of one or more other agents that were present during the same run of the experiment. In particular, it is interesting to select an agent that has a different conceptual system. Table 3.6 show this information

[1]The lowest interval, $[-1, -0.5>$, also contains experiences with values below its lower bound. Likewise, the relevant coordinate of experiences in the highest intervals may exceed the upper bound of those intervals. Thus, a more accurate rendering of the intervals would be to replace the bound by an arrow, representing the indefinite extension of the interval. This notation is not adopted since the bound conveys useful information; it determines at which point a possible split will be located.

	-1..-0.5	-1..-0.5	-1..-0.5	-0.5..0	0..0.5	0.5..1	1..1.5	1.5..2	2..3
ρ_1	**0.338**	**0.342**	**0.319**	0	0	0	0	0	0
ρ_2	0	0	0	**0.492**	**0.508**	0	0	0	0
ρ_3	0	0	0	0	0	**0.502**	**0.494**	0.004	0
ρ_4	0	0	0	0	0	0	0	**0.487**	**0.513**

Table 3.4: Table listing the probabilities $P(\mu|\rho)$. Each column corresponds to a meaning, the label denotes the interval in the first sensor dimension; as splits in dimensions other than the first are also made, this interval may be identical for different meanings (e.g. columns 1-3).

	`co`	`ga`	`pi`	`ra`	`ro`	`sa`	`za`
$[-1, -0.5>$	0.001	0.001	0.001	0.038	**0.952**	0.007	0.001
$[-1, -0.5>$	0.002	0.002	0.002	**0.949**	0.009	0.035	0.002
$[-1, -0.5>$	0.001	0.001	0.001	0.006	0.011	**0.981**	0.001
$[-0.5, 0>$	0.032	**0.709**	0.032	0.045	0.106	0.043	0.032
$[0, 0.5>$	0	**0.999**	0	0	0	0	0
$[0.5, 1>$	**0.999**	0	0	0	0	0	0
$[1, 1.5>$	**0.582**	0.030	0.032	0.059	0.213	0.052	0.033
$[1.5, 2>$	0.002	0.002	0.067	0.003	0.007	0.003	**0.915**
$[2, 3>$	0	0	**0.929**	0	0	0	0.070

Table 3.5: Word production ($P_{prod}(\sigma|\mu)$) for agent 1 at the end of the first run of the experiment.

	-1.00..0.00	-1.00..0.00	0.00..1.00	1.00..2.00	2.00..3.00
`co`	0	0	0	**1**	0
`ga`	0	0	**1**	0	0
`pi`	0	0	0	0	**1**
`ra`	0.137	**0.863**	0	0	0
`ro`	0	**1.00**	0	0	0
`sa`	**0.915**	0.086	0	0	0
`za`	0	0	0	0	**1**

Table 3.6: Interpretation ($P_{int}(\mu|\sigma)$) for agent 2 at the end of the first run of the experiment.

	ρ_1	ρ_2	ρ_3	ρ_4
$[-1.00, 0.00>$	**1**	0	0	0
$[-1.00, 0.00>$	**1**	0	0	0
$[0.00, 1.00>$	0	**1**	0	0
$[1.00, 2.00>$	0	0	**1**	0
$[2.00, 3.00>$	0	0	0	**1**

Table 3.7: Table listing the probabilities $P(\rho|\mu)$.

for the second agent in the same run. As the table shows, this agent has only three distinctions in the first sensor dimension, compared to six for the first agent. The difference between the two agents is that whereas the first agent received noise on its fist sensor dimension, noise for the second agent affected its second sensor dimension, thus obviating the need for extra splits in first dimension. Apart from the splits in this first dimension, the second agent also has a split in the third dimension, displayed vertically in the three dimensional graphs of figure 3.14. Specifically, this horizontal plane is located at the same height as the upper horizontal plane of the first agent's conceptual system, but necessarily extends over the complete region to the left of the first vertical plane, yielding the system displayed in the lower left of figure 3.14.

By analyzing what words agent one uses for each referent and how agent two interprets those words, the language of the agents can be investigated. As an example, let's consider referent ρ_1. The meanings that agent one has developed for different instantiations of this referent can easily be identified on the basis of their intervals, but can also be read from the matrix 3.4.

	co	ga	pi	ra	ro	sa	za
ρ_1	0.004	0	0	**0.346**	**0.319**	**0.327**	0.004
ρ_2	0.012	**0.892**	0.008	0.008	0.046	0.012	0.023
ρ_3	**0.794**	0.012	0.025	0.021	0.107	0.012	0.029
ρ_4	0	0	**0.521**	0.004	0	0	**0.474**

Table 3.8: Production probabilities $P_{prod}(\sigma|\rho)$.

In the word production table, these meanings correspond to the first three rows and are, as noted above, referred to with the words `ra`, `ro`, and `sa`. Table 3.6 shows that the latter of these words is interpreted by agent 2 as its first (leftmost) meaning, while the other two words lead this agent to consider its second meaning as a possible situation. No other meanings are associated with these words and no other words have associations with these meanings. The final step in the analysis of the use of this words when produced by agent one and interpreted by agent two, is to consider to what referents the meanings of agent two correspond. This information is provided by table 3.7, and shows that both meanings unambiguously correspond to the referent ρ_1, i.e. the same referent that was encoded by agent one.

Instead of going through all productions and interpretations of both agents in this manner, a more comfortable way to investigate the question is to simply compute the probabilities that the referent found by the receiver is the one described by the sender of a word. When this probability is averaged for all referents, and furthermore over all combinations of two agents, a sender and a receiver, the result is the fidelity measure.

The first step in this procedure is, as in the manual version of the procedure just described, to determine the probability that a certain word is produced for a certain referent. All combinations of words and referents are needed, and this matrix, $P_{prod}(\sigma|\rho)$ is obtained by multiplying the matrices, $P(\mu|\rho)$ and $P_{prod}(\sigma|\mu)$, shown in tables 3.4 and 3.5. This operation hides the private meanings of the agent, and yields a matrix that shows its word production behavior.

The second step of the procedure is to determine how words are interpreted by the receiving agent. Word interpretation is captured in the $P_{int}(\rho|\sigma)$ matrix. To obtain this matrix, all that has to be done is another single matrix calculation: multiplying $P_{int(\mu|\sigma)}$ with $P(\rho|\mu)$ yields the desired information, shown in table 3.9.

	ρ_1	ρ_2	ρ_3	ρ_4
co	0	0	**1**	0
ga	0	**1**	0	0
pi	0	0	0	**1**
ra	**1**	0	0	0
ro	**1**	0	0	0
sa	**1**	0	0	0
za	0	0	0	**1**

Table 3.9: Interpretation probabilities $P_{int}(\rho|\sigma)$.

Now that both the production information of agent one and the interpretation information of agent two are available, all that rests is to combine these two sources to obtain the desired result. When these two matrices of figure 3.8 are multiplied, the words are hidden from the picture, just as the private meanings were eliminated in the previous calculations:

$$P_{\mathcal{A}1 \to \mathcal{A}2}(\rho|\rho) = P^{\mathcal{A}1}_{prod}(\sigma|\rho) \cdot P^{\mathcal{A}2}_{int}(\rho|\sigma) \tag{3.22}$$

where $P^{\mathcal{A}1}_{prod}(\sigma|\rho)$ is the production matrix of agent $\mathcal{A}1$ and $P^{\mathcal{A}2}_{int}(\rho|\sigma)$ the interpretation matrix of agent $\mathcal{A}2$. In the resulting matrix $P_{\mathcal{A}1 \to \mathcal{A}2}(\rho|\rho)$ each row represents a referent. The values in the matrix specify the probability that this referent, when communicated from agent one to agent two, will be interpreted as the referent corresponding to the column. The resulting matrix for our example is given in table 3.10.

The above showed how the matrix $P_{\mathcal{A}1 \to \mathcal{A}2}(\rho|\rho)$, determining the interpretation by agent $\mathcal{A}2$ of words produced by agent $\mathcal{A}1$, can be obtained. In this manner, the communication between any two agents can be analyzed. From this information, it is straightforward to determine the fidelity of the communication system as a whole, by computing the average matrix for all possible combinations of two agents:

$$P(\rho|\rho) = \sum_{\mathcal{A}1, \mathcal{A}2 \in \mathcal{A}} \frac{P_{\mathcal{A}1 \to \mathcal{A}2}(\rho|\rho)}{n_a^2} \tag{3.23}$$

If a referent that is encoded by some agent is always interpreted as that same referent for all pairs of agent in some system of communication, the $P(\rho|\rho)$ matrix will contain ones on its diagonal and zeroes at all other

	ρ_1	ρ_2	ρ_3	ρ_4
ρ_1	**0.992**	0	0.004	0.004
ρ_2	0.065	**0.892**	0.012	0.031
ρ_3	0.14	0.012	**0.794**	0.054
ρ_4	0.004	0	0	**0.996**

Table 3.10: Matrix containing the average probabilities that a word will be decoded to the referent it encoded. The average over the diagonal entries yields the fidelity measure.

entries. In this case, the *fidelity* of the communication system is complete, i.e. one.

In the private concept experiment reported here, the fidelity measure is 0.92, as can be computed from the matrix in table 3.10. This implies that in 92% of the cases, communication between two agents would be successful; the agents have indeed learned to speak a common language based on private conceptual system. In the model for communication that is used here however, agents do not receive information from a single agent, but from all agents within hearing distance. Fidelity can then be considered as a lower bound for communication success, assuming that agents do not selectively attend to signals from agents with which they have a lower than average $P_{\mathcal{A}1 \rightarrow \mathcal{A}2}(\rho|\rho)$. In fact a positive bias due to selection is more probable given the algorithm; each agent selects the signal which it estimates to most probably yield the actual situation (see the algorithm, section 3.3).

3.4.4 Synonymy and Ambiguity

Finally, it has been investigated whether communication based on private concepts can be used to overcome uncertainty. As explained during the introduction of the sampling versions of the measures, the correctness measure is no longer suited to measure this aspect, but a replacement is available in counting the number of right actions that were selected upon the advent of a new situation that remained invisible to the agent. Figure 3.18 shows this fraction for both sensor-based and signal-based situation determinations. The sensor-based situation determinations should not significantly exceed $\frac{1}{3}$, since the sensors do not provide information about the situation, and the agent can only select one of the three actions at random. A count of this number yielded a fraction of 0.335803, which is indeed close to $\frac{1}{3}$. The signal-based actions were more successful. These were correct

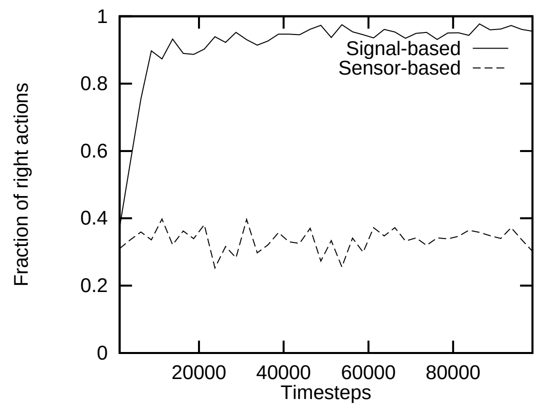


Figure 3.18: Fraction of right actions for signal based and sensor based choices in the private concepts experiment.

in more than 95% of the cases, which shows that the communication system that has been developed on top of the private concepts of the agents enables accurate transmission of information about the situation of the environment. It furthermore confirms the above expectation that the actual accuracy of communication is higher than the average fidelity between two agents, due to a selection effect; apparently, the agents pay more attention to words they can understand.

3.5 Conclusions

This chapter has considered the question of how agents may form associations between privately formed concepts and public words such that a shared system of communication results. Following the commitments outlined in chapter 1, there is no central control mechanism to guide the process of association adaptation, and hence an algorithm had to be found that can be executed by individual agents that has the effect that a global of communication emerges. Such an algorithm has been described in detail, and it has been examined whether all components of the algorithm are required.

The result of the comparison is that all components of the algorithm as it has been presented are useful in that they contribute to the development of communication. There is no component that can be removed without a statistically significant decrease in the quality of the communication system.

Some components are crucial to the functioning of the system, whereas others merely result in a small drop in communication quality when removed. The `sit`, `succ`, and `maxerr` experiments revealed that the corresponding components (signal-based situation determination, the success component in association strengths, and the **max-error-for-signals** threshold respectively) are of the first type; without these, communication does not build up. The experiment where the **max-error-for-signals** threshold was not used showed how the coupling between the concept formation and association formation processes may fail; in this experiment, the communication process was attempting to make use of concepts too early, and by doing so prevented the concepts from developing.

The other two experiments were both concerned with signal-based situation determination. The function of this component is to obtain success information. A striking result of these experiments is that this component is necessary for the development of communication. A further analysis of the success component of the algorithm revealed however that it is not the success information *itself* that is required to achieve this specificity, but rather the lateral inhibition between different meanings it implements. Thus, lateral inhibition can perform the necessary function of providing a bias towards specificity. It does this without imposing a one-to-one relation between the words and the meanings of agents. This is an important property given the commitments of the research, and has been demonstrated in the private concepts experiment. In this experiment, agents developed different meanings but related these to words in such a way that the word used by one agent to describe the particular meaning it observed was interpreted by the other agent as one of its meanings corresponding to the original referent. The matrix describing the communication system demonstrated that this was not an exception, but that reliable communication had developed in spite of substantial differences in the conceptual system. Thus, if the meanings of agents are different, this does not have to prevent them from successfully communicating about their environment.

In the conceptual systems of the agents in this experiment, a single word is often stably associated with multiple meanings. This is necessary for successful communication, and would not be possible using a mechanism for specificity that imposes a one-to-one relation, such as the normalization procedure in (Oliphant, 1997). The gradual changes made by the success component provide specificity without imposing this one-to-one relation.

Also, measures for the quality of communication have been developed. The specificity measure expresses to what extent a word identifies a referent, while consistency indicates whether an agent consistently uses the same word for the same referent. To measure whether the word used for a referent is the same for all agents, the coherence has been defined. Finally, fidelity measures the average probability that a word produced by one agents will be interpreted by a random other agent to yield the original referent. The specificity and consistency measures are principled in that they are based on the decrease of uncertainty, measured as entropy in a conditional probability matrix, and overcome a fundamental limitation of a related measure of communication quality introduced in (MacLennan, 1991).

Chapter 4

Communication as a Dynamical System

When language is viewed as a dynamical system, a natural question is whether the tools that are available for analyzing such systems can offer new insights into it. This is the question that will be investigated in the current chapter. A central notion in the domain of dynamical systems is that of an *attractor*. It will first be investigated whether the system of communication that has been investigated possesses these. Another tool related to the dynamical perspective is that of phase plots. Phase plots characterize the possible states of a system. The most straightforward use of phase plots is to show the trajectories that the system visits. This method will be used to view the processes of association and dissociation of words and meanings in a different way. Finally, a parameter called *temperature* has been found to have a large influence on the development of communication. To better understand the effect of this control parameter, its effect on the development of communication is investigated.

The structure of this chapter is as follows. First, in section 4.1, some basics of dynamical systems theory that are necessary to understand the following investigation are described. The variables of the dynamical system that models our communication system are determined in section 4.2. The presence of attractors in a deterministic version of the system is investigated in section 4.3. Both experimental evidence and a theoretical proof of the presence of attractors are provided. Although the deterministic version functions reasonably well, experiments in section 4.4 suggest that stochasticity is a useful element in language evolution by demonstrating its positive effect on the development of communication. The behavior of the stochastic system is further explored in section 4.5. Finally, section 4.6 concludes.

4.1 Basic Dynamical Systems Theory

A dynamical system is a system that changes over time. Although the basis for dynamical systems theory was already laid by Newton, important parts of the field, such as chaos theory, only developed during the second half of this century, when the advent of computers enabled researchers to quickly perform complex simulations.

A dynamical system is characterized by a number of *variables*. The systems specifies how these variables change. If the specification of the system involves *time*, the system is called *non-autonomous*. In *autonomous* systems, the changes of the variables are defined relative to one another. Changes can either be stepwise or discrete, or continuous.

The *phase space* or *state space* of the system is a space where each dimension represents a variable of the system. At each point in time, the state of the system is characterized by a single point in this state space, determined by the values of the different variables. The *trajectories* of a system are the paths along which the system moves through phase space over time. The derivative at each point in phase space specifies the direction at that point of trajectories that pass through the point. The magnitude of the derivative specifies the speed at which the system travels through the point.

A central notion in dynamical systems theory is that of an *attractor*. Informally, an attractor is a part of the state that appears to attract the system, in the sense that the state of the system will move toward the attractor when it is near. In a physical interpretation, a stable state to which a non-conservative system tends is an attractor. Several forms of attractors exist, such as *fixed points*, *limit cycles*, and *strange attractors*. Fixed points are points where the derivative of the state equals zero. Thus, when the system is in the state that corresponds to such a fixed point, it does not leave this state anymore, since the speed with which it moves is zero. Three variants of such points exist: fixed point attractors, fixed point repellers, and saddle points. A point attractor is a point where all neighboring trajectories are directed towards the fixed point. Such a point 'attracts' the system to itself from some neighborhood. This neighborhood is called the *attractor basin*. For a repeller on the other hand, the opposite is the case; in all points near the fixed point, trajectories head away from the fixed point, and thus the system will be repelled from the immediate neighborhood of such points. A saddle point is a combination of a fixed point attractor and a fixed point repeller. From some directions the system is attracted towards the fixed point, while it is repelled in other parts of the neighborhood.

Although the informal definitions of attractors may seem quite clear, there is no consensus about a formal definition for attractor. Here, the definition from (Strogatz, 1994) will be used. There, an attractor is defined as a closed set $\mathcal{Z}$ for which the following conditions hold:

1. $\mathcal{Z}$ is an *invariant set*: any trajectory $\mathbf{x}(t)$ that starts in $\mathcal{Z}$ stays in $\mathcal{Z}$ for all time.

2. *$\mathcal{Z}$ attracts an open set of initial conditions*: there is an open set $\mathcal{U}$ containing $\mathcal{Z}$ such that if $\mathbf{x}(0) \in \mathcal{U}$, then the distance from $\mathbf{x}(t)$ to $\mathcal{Z}$ tends to zero as $t \to \infty$. This means that $\mathcal{Z}$ attracts all trajectories that start sufficiently close to it. The largest such $\mathcal{U}$ is called the *basin of attraction* of $\mathcal{Z}$.

3. $\mathcal{Z}$ is *minimal*: there is no proper subset of $\mathcal{Z}$ that satisfies conditions 1 and 2.

Another pattern that may occur in the phase space of dynamical systems is the *limit cycle*. A limit cycle is a closed curve where the derivative of the system is such that, once on the curve, the system will stay on it forever. A well known example of a limit cycle is the oscillator used by the Dutch scientist Balthasar van der Pol to describe the operation of an electronic valve oscillator. Another, more interesting type of attractors is the *chaotic attractors*. Such attractors display sensitivity to initial conditions, meaning that two points starting near each other will move away from each other exponentially. Many chaotic attractors have fractal shapes, and are therefore sometimes called *strange attractors*.

4.2 Determining the Phase Space

A dynamical system is characterized by a number of variables that change over time. The space of possible configurations of the system contains all possible combinations of admitted values for these variables, and is called the phase space. The first thing that needs to be asked about our system of communication in order to investigate its dynamics, is what these variables are. The complete state includes the state of the environment, and that of all agents in it. However, not all aspects of this state are of interest to this investigation. Particularly, the state relating to concept formation and action selection will not be regarded. The experiments in the previous chapters learned that the concept formation processes are concentrated in the initial phase of the experiment, and that the conceptual system of the agents does not change after the first thousand time steps under normal conditions. On the time scale of the development of communication, the effects of concept formation play no role of any importance. For action selection, the situation is similar; after 10,000 time steps, the average error in the agents' expectation of their rewards has dropped below 10^{-7}, which implies that the agents can virtually always choose the right action when situation determination is successful. For this reason, the state relating to action selection, which consists of the estimates of the values of the action, is not relevant to the investigation of the dynamics of communication.

The key factor determining the communicative behavior of an agent is its set of associations between meanings and words. Like the conditional probabilities characterizing a conceptual or communication system, these association strengths can be written in a matrix of which the row and column determine a meaning and a word. The **produce-signal** routine, see the algorithm (figure 3.3), specifies how these associations are used by the agents to select a signal for production. However, meanings are private entities, and therefore analyzing the phase space of association values is more convenient when based on the public referents. Such a matrix is obtained by multiplying the association matrix with the $P(\mu|\rho)$ matrix. When association values are used directly, as opposed to their corresponding production probabilities, this operation requires perfect parsimony and distinctiveness; production probabilities depend on the association values of a meaning with other signals, which implies that the association values of different meanings corresponding to a referent cannot be directly compared. The conditions concerning parsimony and distinctiveness are satisfied in the experiments[1], and therefore the signal-associations corresponding to the referents will initially be considered as the variables of the system.

Apart from word production, the interpretation of words can also be viewed as being part of the communicative behavior of an agent. As the al-

[1] The reader may recall that in the standard experiment, these conditions were almost, but not completely satisfied. The difference is a side effect of the removal of stochasticity from the system, as documented in section 4.3.1.

gorithm showed, interpretation is based directly on occurrence probabilities of words as observed by the agents. These estimates indirectly determine signal production since they are the basis for signal based situation determination. However, if a good system of communication develops, word interpretation corresponds closely to word production, and thus its added value is limited. On the other hand, the costs of including interpretation in the investigation of the state space very large; for every single state in production space, a complete space with a size comparable to that of the production space is added (in other words, the volume of the phase space is squared). Given these considerations, the course of action that will be taken is to first investigate only word production. If the behavior of the system as viewed in this space is sufficiently stable, the interpretation state does not have to be included in the investigation.

4.3 Presence of Attractors

When viewing language as a dynamical system, see e.g. (Steels, 1997a), an interesting hypothesis is that in a population of language learners, good systems of communication are attractors. This could explain how a language can emerge without central control. In the following, it will be determined whether this is the case for the system that has been described.

4.3.1 Stochasticity and Attractors

A first observation is that the system as it has been described cannot be said to have attractors in the strict sense. The reason is that it is stochastic. The fact that chance is involved in signal production implies that there is a possibility that agents do not use their preferred words, even if these are very strongly associated, but one of the weakly associated words that are used for other meanings. Even if this possibility decreases to a level where it almost never occurs during a simulation of standard length, the nonzero probability implies that the event can occur in principle. When it does, the association strengths of the agents that receive such a signal will be adapted accordingly, i.e. they rise. This rise increases the distance to the candidate attractor. Since a nonzero probability implies that this event can not only occur once, but may occur (with extremely low probability) on any number of subsequent occasions, the distance to the candidate attractor may increase indefinitely. Thus, the stochasticity of the system violates the second condition of the definition of attractors. This is not so surprising, as the theory behind the attractor definition assumes a deterministic system.

How serious is this problem? For practical purposes, there may be points in the phase space that can be considered as attractors if the system behaves as if they were attractors on virtually every occasion. Mathematically thought, those points are not attractors. To keep this distinction clear, such points will be called pseudo-attractors here. It is instructive to see whether a deterministic system that is completely analogous to the original system has attractors in the formal sense. This question will be investigated in the following.

4.3.2 A Deterministic Version of the System

There are several sources of stochasticity in the system as it has been described. These are:

1. Action selection
   As noted in section 2.4.1, learning the values of actions requires exploration. Most forms of exploration, including the one adopted here, involve stochasticity.

2. Signal production
   In the basic system, the association strengths between words and meanings are used as parameters for a Boltzmann distribution that determines word production.

3. Situation Determination
   In the basic system, signal based situation determination occurs with a probability equal to the estimated probability of the situation.

To remove these forms of stochasticity, the following changes have been made:

1. To remove the stochasticity in action selection, exploration is stopped after a fixed number of time steps (10,000), when the values of the different actions are sufficiently accurate.

2. Instead of using the association strengths to determine production probabilities, agents always select the signal that has the highest association strength.

3. The probability of a situation that is estimated by the agent during situation determination is not used as a probability of selecting that situation, but is compared against a threshold; the choice is based on signals if and only if this value is larger than $\frac{1}{2}$.

The system that these modifications yield is deterministic. The reader may not that the approach to removing stochasticity from the action selection mechanism differs from the other two cases. This is because exploration is required to learn action values, but once these are learned, no further adaptation of these values is necessary. Using this scheme, action selection has no influence whatsoever on the development of communication once it has been frozen. The other two cases are part of the communication system itself and hence form part of the investigation. They can therefore not be treated in this way. The next question that should be asked is whether, apart from stochasticity, there are other properties of the system that exclude the presence of attractors; if so, it would be of no use to investigate this presence. Three such factors have been identified:

1. Convergence to an attractor requires that both the success and use components of all associations except one preferred signal for each meaning tend to zero. However, in the standard algorithm, the success component of the association between a signal σ and a situation μ only decreases when a signal-based determination of μ based on preferred signal σ is estimated to be *incorrect* (see algorithm), or when a signal-based determination of a situation other than μ based on preferred signal σ is estimated to be *correct*. This implies that once a reasonable system of communication has developed, the association success of words that are not the preferred word of some situation will not be updated, and cannot tend to zero as required to move the distance to the attractor towards zero.

2. Regarding the use components of associations, the case is similar. When a good system of communication arises, the fraction of signal-based situation determinations can approach one, especially with the modification concerning signal-based situation determination that had to be made above to ensure determinism. Since the use component of associations is not updated after a signal-based situation determination (see algorithm), the decrease of the use values of non-preferred signals can stall.

3. Partial perception can in principle occur for all agents simultaneously. In this case, the agents have no way of determining their situation. The incorrect situation determinations that are the effect of this cause the associations of the preferred words of the situations that were determined by the agents to decrease, thus increasing the distance to the nearest candidate attractor with good communication. This

possibility is not merely theoretical; when five agents are present and partial perception occurs once in every 10 occasions, as is the case in the experiments, the probability of this happening simultaneously for all agents equals $\frac{1}{10}^5 = 10^{-5}$, which means that during the 10 runs of 200,000 time steps, it is expected to occur 20 times.

These factors have been addressed by the following changes:

1. The success components of the associations between a situation μ and a signal σ_i that were not the preferred signal for that situation are decreased after a successful signal-based situation determination.

2. The use component of associations is also adapted when situation determination is based on signals.

3. The perception of the agents is complete, i.e. the agents do not receive misleading information as is the case in the basic experiment.

4.3.3 Determining the Location of Attractors

The purpose of the experiments that follow is to determine whether the system has attractors that correspond to good communication. The second condition in the definition of an attractor specified that when the system is in a neighborhood of an attractor, the distance to this attractor must tend to zero. Thus, the procedure for testing whether the system has attractors requires that the location of these attractors be known, and that a neighborhood around them be chosen within which the distance to the attractor can be monitored to see whether it tends to zero. In the following, the locations of possible attractors and their neighborhoods will be determined.

The locations of which we want to know whether they are attractors of the system, are the locations where the multi-agent system accommodates a perfect system of communication. These locations are characterized by the necessary and sufficient condition that for each referent, each agent has only one word strongly associated, and this is the same word for every agent, but different for each referent. All other associations between referents and words must be low.

As a result of the modifications that were applied to make the system deterministic, signal production always selects the signal with the highest association, and the words with lower, nonzero associations are never produced. This implies that for the deterministic version of the system,

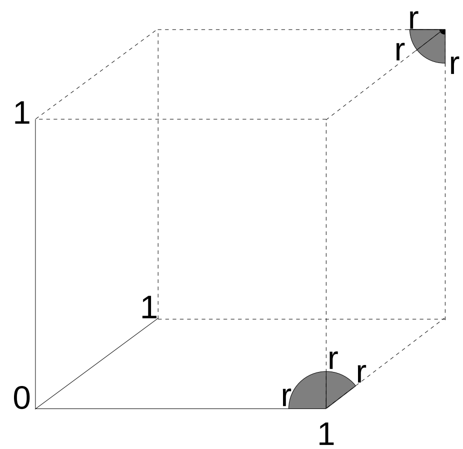


Figure 4.1: Schematic representation of attractors and their neighborhoods (see text).

there are *regions* in phase space, not merely dimensionless points, where communication is perfect. Within these regions, the associations continue to change indefinitely. The candidate attractors of the system now are the points towards which these movements are directed. In these points, the high associations have their maximum value of 1, and the low associations the minimum value of zero. It may be noted that due to the recency-weighted average update rule, these locations can never be reached exactly. The question that has to be answered by the experiments is whether the *distance* to them tends to zero.

4.3.4 Determining Neighborhoods for Attractors

The condition that the systems should tend towards attractors has to hold within a neighborhood around these attractors. Thus, it is necessary to define neighborhoods around the attractors within which the movement of the system is to be monitored. Here, these neighborhoods will be determined as a region centered around each attractor where the distance to the attractor is below a threshold. This definition yields an n-dimensional hypersphere with a radius equal to this threshold. The figure illustrates this for the three dimensional case. In reality, the minimal dimensionality of the phase space is at least 80, since there are 4 referents, 5 agents, and at least 4 words to distinguish between the referents.

Figure 4.1 shows a schematic representation of attractors and their neighborhoods. Each axis represents the association strength for some word, referent, and agent. In reality the number of dimensions is much larger (80 or more), and the radius of the sphere much smaller (0.001). The phase space is a hypercube with edge length one that starts in the origin and extends along the positive axes. The coordinates of an attractor representing ideal communication are either zero or one, placing it on one of the corners of this hypercube. Only a fraction of these corners ($\frac{n_r!}{2^d} \approx \frac{24}{1.2 \cdot 10^{24}}$ for $d = 80$) are such attractors, since most corners correspond to communication systems where specificity, consistency, or coherence are low.

4.3.5 Experimental Investigation of the Presence of Attractors

In this section, experiments are reported that investigate whether states that correspond to perfect communication are attractors of the system. The conditions of the attractor definition will be addressed in reverse order in the following three subsections: presence of proper subsets, open set of initial conditions, and invariance.

Presence of Proper Subsets

The last condition states that the set of points that constitute the attractor may have no proper subset that satisfies the first two conditions. This condition is always satisfied in the case of point attractors, since the attractor consists of a single point and hence the set has no non-empty proper subsets.

Open Set of Initial Conditions

The second condition for attractorhood states that an attractor must have a neighborhood such that whenever the initial condition of the system falls within this neighborhood, the distance to the attractor must tend to zero (see figure 4.2). This subsection starts with a description of the experimental procedure to investigate this, which is based on perturbations, and presents the results of the experiments.

The Perturbation Procedure The simulation is run until a fixed number of time steps, chosen at 50,000, to allow concept formation to settle down and to see whether the system moves towards a state corresponding

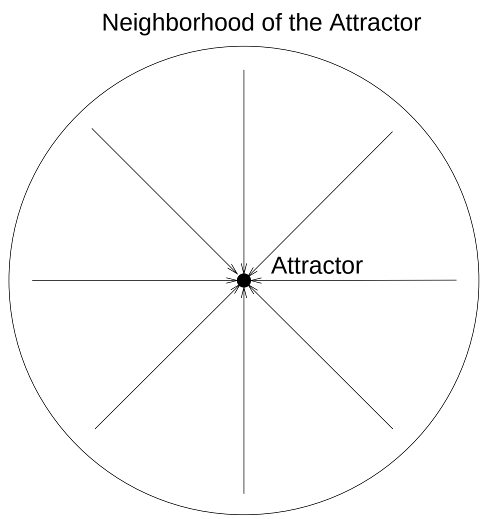


Figure 4.2: Schematic rendering of an attractor in the deterministic system. When the initial state of the system is within a neighborhood of the attractor, it moves towards the attractor. The arrows show only one possible configuration of vectors; many vector fields, e.g. fields with spiraling trajectories, satisfy this condition.

to good communication. The system is then perturbed. A perturbation here signifies that the system is taken out of its current state and moved to a random point within a distance r of the nearest corner of the hypercube that contains the phase space (see fig. 4.1). That is, its new state is a random point within the hypersphere with radius r and the corner as its center. Most of the corners do not represent successful systems of communication. However, if the corner corresponds to good communication and is a candidate attractor (see section 4.3.3), the distance of the system's state to this candidate attractor is monitored over time. The question that is then investigated is whether this distance tends to zero.

Perturbations are achieved as follows. First the closest corner of the hypercube needs to be determined. This position is found by taking the current position of the system and moving each coordinate, i.e. each association between some referent and word of some agent, to either zero or one, depending on which is closest. Next, each of the variables is moved a random distance away from its current value. Let r be the radius of the hypersphere within which the new location of the system has to lie. The coordinates with value zero are increased, those with value one are decreased, both over a distance that is randomly chosen from the interval $[0, r]$. This moves the system to a point within a hypercube with edge length r whose corner furthest from the origin coincides with the selected corner. If the point is inside the hypercube, but outside the segment within phase space of the hypersphere with radius r centered around the same point [2], the state is moved in the direction of the center until it reaches a distance from the center randomly selected from $[0, r]$. This procedure ensures that the system is moved towards a point within the hypersphere with radius r, and that all points within this hypersphere have a nonzero chance of becoming the new location of the system in phase space.

The next question is how the desired association values thus selected are achieved. As described in chapter 3, an association value is a linear combination of a use value and a success value. The combinations of use and success values are subject to the constraint that both use and success should be within the interval $[0, 1]$. Since the association values are weighted combinations of use and success, the intervals are further restricted depending on the desired association value. For example, if success has a weight of 0.8, and use therefore a weight of 0.2, and the desired association value is 0.6, then the minimum success value is $\frac{0.6-0.2}{0.8} = \frac{1}{2}$. The use and success value are chosen randomly under these constraints, so that all possible combinations that yield the desired association value have a chance of being selected.

Another issue concerning perturbations is that of the interpretation information. As remarked above, this information is not represented in the phase space. However, it does affect the communicative behavior of the agents, since signal-based situation determination is based on it. This raises the question of how the interpretation information, which does not form part of the state vector, should be initialized. To address this issue, all interpretation information stored by the agents is reset during a perturbation. This renders the state related to interpretation neutral, and has the result that the subsequent behavior is only determined by the production information, which does form part of the state.

Results of the Perturbation Experiments In the experiments, a perturbation is performed every 10,000 time steps, starting at time step 50,000.

[2] This happens more often than one might think; for three dimensions, the probability is one minus the ratio between the volume of a sphere and that of the smallest cube that surrounds it, i.e. $1 - \frac{\frac{4}{3}\pi r^3}{(2r)^3} \approx 0.48$; for higher dimensional spaces, this probability rapidly increases.

In the following, an experimental investigation of the question whether the system will tend to move towards a hypothesized attractor when placed in a neighborhood of it is presented. Such hypothesized attractors will be called *candidate attractors*, since the hypothesis is that they are attractors, but this can only be confirmed after an investigating whether the conditions of the attractor definition hold. The first test to be performed is whether the deterministic system moves towards a state of good communication. If this is the case, the question this subsection is concerned with can be examined: does the system, when placed in a neighborhood of a candidate attractor, move towards it? If it does, the second condition of the definition for attractors is satisfied.

Locating Candidate Attractors Figure 4.3 shows the distance of the state of the communication system to the nearest corner of the aforementioned hypercube. The top left graph shows the complete time series. Apart from the distance, two other lines are plotted: *fidelity* and *indicator*. Fidelity is plotted to show the quality of communication at different points in time. *Indicator* is an indicator function that tells whether the system is in the neighborhood of a candidate attractor. It is binary, and takes the value of one when the following conditions are met:

- The system is within a hypersphere with radius 0.001 of a corner of the hypercube.
- This corner represents an ideal system of communication, i.e. the fidelity measure equals one when the system is in that corner.

In figure 4.3, it is seen that during an initial period, until around time step 5000, there are only small changes in the distance. Note that the corner to which the distance is calculated may change during this period, as the location of the system in phase space keeps changing. Then the system starts to move towards a particular corner. As long as the graph stays below 0.5, which it continues to do indefinitely, one can be sure that the corner towards which the system is moving does not change anymore; if it did, at least one of the coordinates would have to change from zero to one or vice versa, and would pass the value of 0.5 in doing so, at which point the total distance would be greater than 0.5.

The decrease of the distance to the corner is steep, and corresponds to a rise in fidelity. After 11,000 time steps, the fidelity measure equals one, which implies that the system is moving towards a corner with ideal

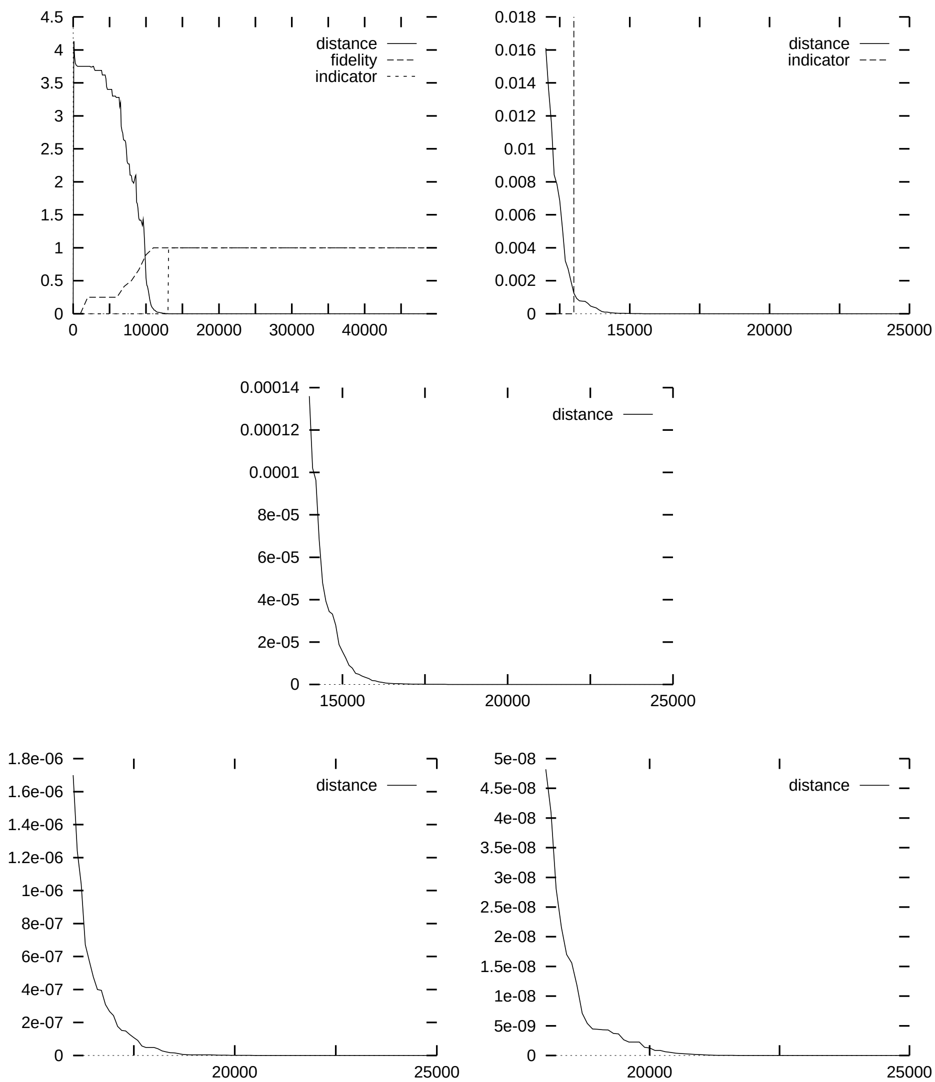


Figure 4.3: Evolution of the distance to the attractor over time. The graphs show the same data, but on different scales. Whereas the in the top left graph the distance appears to have converged to zero, closer inspection reveals that the distance keeps on diminishing at an ever slower pace.

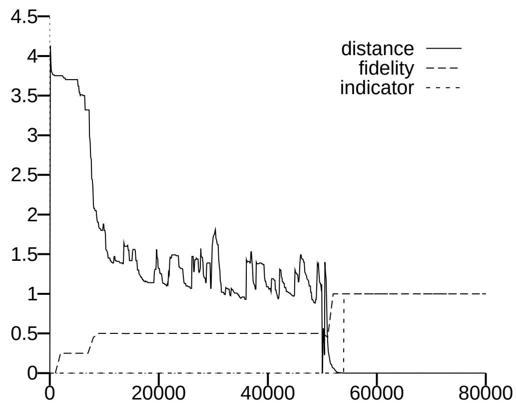


Figure 4.4: Distance to the attractor over time. After the perturbation (time step 50,000), the system is moved towards a corner that does not correspond to good communication. The result is a sharp peak, indicating a move away from the corner, followed by convergence to a corner that does correspond to good communication.

communication. The neighborhood of this corner is not entered until just after time step 13,000, when the distance has dropped below 0.001.

An optical inspection of the top left graph would appear to show that the distance quickly drops to zero. However, as can be seen from the recency-weighted average equation 3.6, the distance will never actually *reach* zero. Therefore, to get a better feeling for the shape of this curve, the graph is shown at increasingly small scales, each subsequent graph starting 2000 time steps later. In the top left graph, the distance to the attractor appears to have reached zero by time step 12,000. However, on the smaller scale of the top right graph, which starts at time step 12,000, it can be seen that it has not quite reached zero yet. The middle graph, which starts at time step 14,000, shows that even at time step 16,000, the distance was not zero. The distance of the graph to zero is decreasing approximately exponentially, and at each point will reach a tenth of its current value within a few thousand time steps. Similarly, whereas the distance appears to have converged to zero in the middle graph around t=17,000, the next graph shows that here it is still decreasing, at a very similar relative rate. The bottom right graph shows one more magnification step, and subsequent magnifications will continue to look very similar to those shown. The origin of this shape is the weighted average update rule, which makes steps of exponentially decreasing size when moving towards a constant point.

The above results were presented to illustrate *how* the distance to one of the corners tends to zero. The particular corner to which the system converges in this experiment is a location where the communication is ideal, in the sense that both specificity, consistency, and coherence are all equal to one; in other words, every referent corresponds to a single word that is used by every agent, and this word is different for different referents. In five of the ten runs, the deterministic system converged towards such a state. In the other half of the runs, the system did not spontaneously develop accurate communication.

Figure 4.4 shows a typical example of a run where accurate communication did not develop spontaneously. The distance to a corner also drops substantially in the beginning, but not as far as in the previous case. Then, instead of converging exponentially to the corner, it keeps varying at an intermediate level. As the graph shows, the fidelity of communication during this period is only 0.5. Then, at time step 50,000, the first perturbation takes place. The system is moved to a random location within 0.001 distance of the closest corner, visible as a quick drop in the distance value. However, this closest corner does not correspond to good communication. This can be seen from the fact that the fidelity measure does not immediately rise to one. An examination of word production showed that a single word was consistently used for three different meanings by all agents. Only the remaining meaning had its own, unambiguous word. This ambiguity is the result of the deterministic selection mechanism that governs word production; if one word happens to be slightly stronger associated to several meanings at one point by many agents, the agents continue to use this word for these meanings.

It is interesting to see that the system itself moves away from this corner, and the distance to the corner rises again with a sharp peak just after time step 50,000. Then, the system manages to find a region in phase space with good communication; the distance to the corner gradually decreases, and the fidelity measure reaches one. When the movement towards this corner has decreased the distance below the radius of 0.001, the neighborhood of a candidate attractor is reached, as is seen from the value of the indicator function.

Course of the Distance to Candidate Attractors The experiments above illustrated two ways for the system to reach corners of the hypercube where communication is ideal: spontaneously, or as the result of a perturbation. In all of the ten runs the system arrived at such a corner

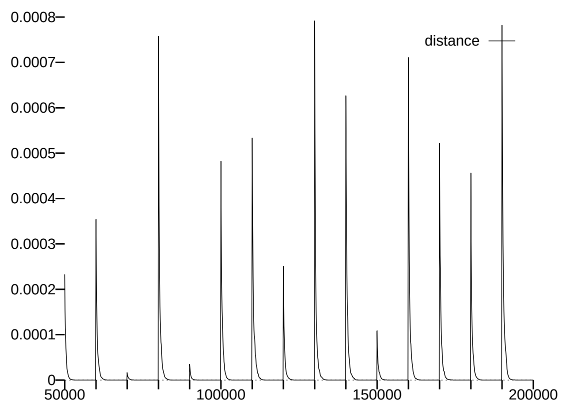


Figure 4.5: Results of one of the perturbation experiments. After every single perturbation, visible as a sharp peak, the system reacts with a steep descent towards the candidate attractor.

after one or two perturbations or, in five of the cases, without requiring any perturbation at all. This situation enables us to investigate the actual question that needs to be answered: once placed within the neighborhood of a candidate attractor, does the distance to the candidate attractor tend to zero?

Since the location of these corners is known by the experimenter, it would be possible to place the system in such a location directly. The present procedure was preferred both for implementation reasons and because it demonstrates that attractors are found by the system from random initial conditions, which indicates the attractor basin is not so small that reaching an attractor is merely a theoretical possibility.

In the following experiment, the candidate attractors found using the procedure described above are used as a starting point. At fixed intervals, a perturbation is performed, which consists of moving the location of the system in phase space towards a point within the neighborhood of the nearest candidate attractor. The evolution of the distance to this candidate attractor is then monitored over time. For the candidate attractor to be a real attractor, this distance must tend to zero. In this experimental investigation, this criterion is examined in two ways.

First, the evolution of the distance to the candidate attractor after the perturbations is inspected visually. Figure 4.5 shows this evolution for the 15 perturbations of the first run of the experiment. Every 10,000 time steps, the system is perturbed. These events are seen as steep peaks in the graph, with values that can vary between zero and 0.001, the boundary of the hypersphere that has been selected as a neighborhood within which convergence is experimentally investigated. After every single perturbation, in all of the runs, the system reacted with a steep descent towards the candidate attractor.

Second, it has been investigated whether the descents towards the attractors are monotonic by calculating the differences between every pair of subsequent distances. A descent is monotonic if at every time step the distance to the attractor either decreases or remains at the same level; if it increases on one or more of the time steps, the monotony property is not satisfied. If all descents are monotonic, this is an *indication* that the distance to the attractor tends to zero. This examination showed that after every single perturbation, in all of the runs, the distance to the attractor decreased monotonically.

In summary, all of the candidate attractors that were encountered had the property that the system, when having its initial condition in a neighborhood of such an attractor, tends to move towards it. This concludes the experimental investigation of the middle condition of the attractor definition.

Invariance

The final part of the experimental evidence required to investigate the presence of attractors is the condition that attractors must be invariant sets; any trajectory that starts in $\mathcal{Z}$ stays in $\mathcal{Z}$ for all time. This question has been examined for all of the attractors that were found in the perturbation experiment. The procedure was as follows. At the end of each of the perturbation experiments described in the previous subsection, at time step 200,000, the system is always at close distance of a candidate attractor. Since the weighted average update rule never really converges to equal the goal value, this distance will never be zero, save for the very unlikely case where the initial state of the system happens to be an attractor. At time step 200,000 then, the system is moved to the location of the candidate attractor. Subsequently, the experiment is continued as usual, without any perturbations. If the location of the system in phase space remains exactly equal for a substantial number of steps, this is experimental evidence that the candidate attractor is an invariant set.

Figure 4.6 shows the distance to the attractor over time, in units of 10^{-14}. Apart from the evolution of this distance before time step 200,000, the only event that is visible is the perturbation onto the attractor. From

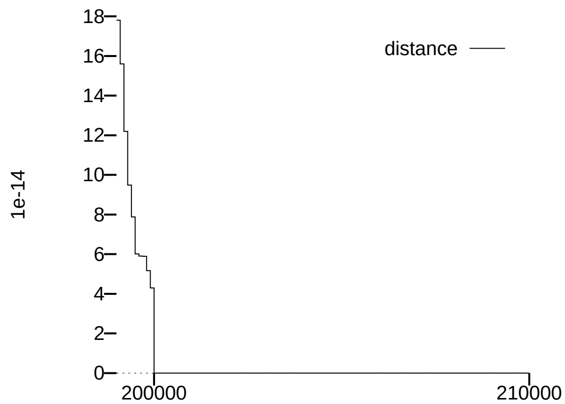


Figure 4.6: Distance to the attractor over time. At time step 200,000, the system is moved to the nearest attractor. From then on, the system remains in the attractor state, as required in the definition of attractors.

time step 200,000 on, the system has a zero distance to the candidate attractor. Inspection of the data showed that the distances do indeed equal zero, i.e. there are no fluctuations at a scale that escapes visual detection. This result was identical for all of the runs.

4.3.6 Theoretical Proof of the Presence of Attractors

After this experimental demonstration of the presence of attractors in the deterministic version of the system, a theoretical proof will now be given. The following assumptions about the initial conditions are made:

- For each agent, there is a one-to-one correspondence between referents and meanings. In the deterministic version of the system, this was always the case.

- For each agent, there is a one-to-one correspondence between meanings and their strongest associated words. Furthermore, for each referent, the word that is most strongly associated with each agent's corresponding meaning is the same. This assumption defines the attractor neighborhood within which the system is placed. As the experiments showed, this condition is not always reached spontaneously by the deterministic version of the system, but in all of the runs, the condition was satisfied after one or two perturbations. The condition is equivalent to the condition that the fidelity measure must equal one. In the deterministic version, there are large regions where this condition is satisfied, due to the different word production mechanism. In the stochastic version of the system however, this condition is only satisfied in a finite number of points towards which the system can converge in the limit, but which can never be actually reached.

- The estimates of the action values have converged to a point where the difference between the estimate and the actual value does not exceed the threshold **max-error-for-signals**. The first time step at which convergence to the attractor is investigated is chosen such that this condition is satisfied. This guarantees that the evaluation of the correctness of signal based situation determinations by the agents is correct.

For convenient notation, it will be assumed, without loss of generality, that the referents, meanings, and words that correspond to each other have the same index. The number of meanings equals the number of referents. The number of words however is in principle unlimited. Thus, σ_j corresponds to a referent if $j \leq n_r$.

First, the tendency towards zero of the distance to the attractor is considered. Let the word produced by the agents in connection with a referent ρ_i be referred to with σ_i. The above conditions guarantee that such words exist and are unique, i.e. $\sigma_i \neq \sigma_j$ for $i \neq j$. The procedure parallels that of the experimental investigation, i.e. the system is placed at a random position within the neighborhood of an attractor, and the agents' reception information, consisting of estimates of $P_{int}(\mu|\sigma)$, is removed.

Word Production

Let i be chosen such that ρ_i is the current referent. During signal production, each agent will select the corresponding word σ_i.

Association Updates

Since each agent produced the word σ_i, all agents will receive this same word n_{ag} times. At the first time step after a perturbation, no reception information is present, and hence the agents will select sensor based situation determination. Since there is a one-to-one correspondence between referents and meanings, the agents will all arrive at their own corresponding meaning $\mu_i^{\mathcal{A}}$.

Let $a^{\mathcal{A}}(\rho_i, \sigma_j)$ be the association strength between an agent's meaning for referent ρ_i and the word σ_j. The use and success components of an association $a^{\mathcal{A}}(\rho_i, \sigma_j)$ are referred to with $a_u^{\mathcal{A}}(\mu_i, \sigma_j)$ and $a_s^{\mathcal{A}}(\mu_i, \sigma_j)$, respectively.

After a sensor-based situation determination, the following association updates take place. For each agent $\mathcal{A}$ the use component of the association strength between the received word σ_i and the situation $\mu_i^{\mathcal{A}}$ is increased. The recency-weighted average update rule causes a weight update for each occurrence of a word in the input of:

$$a_{n+1} = \alpha a_n + (1 - \alpha) z \tag{4.1}$$

where z is the goal value. It can easily be seen that the result of repeated application of this update rule is given by:

$$a_{n+r} = z - \alpha^r (z - a_n) \tag{4.2}$$

Thus, the association strength updates after receiving the word σ_i n_a times (once for every agent) are governed by the following rule, which will be referred to as *update rule 1*:

$$a_{u,t+1}(\mu_i^{\mathcal{A}}, \sigma_j) = \begin{cases} 1 - \alpha^{n_a}(1 - a_{u,t}(\mu_i^{\mathcal{A}}, \sigma_j)) & \text{for} \quad j = i \\ 0 - \alpha(0 - a_{u,t}(\mu_i^{\mathcal{A}}, \sigma_j)) & \text{for} \quad j \neq i \end{cases} \tag{4.3}$$

The success components of the associations are not updated after a sensor based situation determination. At subsequent time steps, i.e. two or more time steps after the perturbation, there are two possibilities: if the referent has not been encountered yet, as was the case above, its corresponding word has not been produced yet, and hence the interpretation information estimates that the probability of being in the corresponding situation given the received word is zero. On these occasions, which after a perturbation occur once for every referent, a sensor based situation determination is performed by each agent, with the association updates given in equation 4.3.

After a signal based situation determination, the association success components are updated. The assumptions specified that the agents all produce the same word σ_i after a referent ρ_i. The fact that signal based situation determination was selected, implies that for each agent $\mathcal{A}$ this situation $\mu_i^{\mathcal{A}}$ has occurred before, and therefore the only word for which the interpretation information of this meaning $P(\mu^{\mathcal{A}_i}|\sigma_j)$ contains nonzero estimates is σ_j. This ensures that the signal based situation determination can only yield $\mu_i^{\mathcal{A}}$, if the estimate exceeds 0.5, or revert to sensor based determination, in which case the updates of eq. 4.3 are applied. For the signal based determination, the absence of partial perception and the assumed convergence of the action value estimates ensure that the determination will be correct. Hence, *update rule 2* specifies the following success component updates:

$$a_{s,t+1}(\mu_j^{\mathcal{A}}, \sigma_i) = \begin{cases} 1 - \alpha(1 - a_{s,t}(\mu_j^{\mathcal{A}}, \sigma_i)) & \text{for} \quad j = i \\ 0 - \alpha(0 - a_{s,t}(\mu_j^{\mathcal{A}}, \sigma_i)) & \text{for} \quad j \neq i \end{cases} \tag{4.4}$$

Furthermore, since this analysis concerns the deterministic version of the system, the success components of associations between the current meaning and signals other than the preferred one are decreased by *update rule 3*:

$$a_{s,t+1}(\mu_i^{\mathcal{A}}, \sigma_j) = 0 - \alpha(0 - a_{s,t}(\mu_i^{\mathcal{A}}, \sigma_j)) \quad \text{for} \quad j \neq i \tag{4.5}$$

Finally, also specific to the deterministic system, the association use updates of equation 4.3 are performed.

The Influence of Association Updates on the Distance to the Attractor

Now that the association updates have been specified for all possible cases, it is time to investigate their effect on the distance to the nearest candidate attractor. The reader will remember that in a candidate attractor, all association strengths are either zero and one, there is only one word per referent for which this value is one, and this word is the same for every agent. The assumptions of this analysis specify that there is only one word per referent that is *produced* by the agents, and that this word is the same for every agent. Given the workings of the deterministic word production mechanism, this implies that the preferred word of a meaning has a higher association value than the other words associated with it. Therefore, the closest attractor is that for which these highest associations all have a strength of one, and all other associations have zero strength. This allows us to specify the distance to the nearest attractor for n_a agents, n_r meanings, and n_w words thus:

$$\text{dist} = \sqrt{d_t} \tag{4.6}$$

$$d_t = \sum_{k=1}^{n_a} \sum_{p=1}^{n_r} \sum_{q=1}^{n_w} \begin{cases} (1 - a_t(\mu_p^{\mathcal{A}_k}, \sigma_q))^2 & \text{for} \quad p = q \\ (0 - a_t(\mu_p^{\mathcal{A}_k}, \sigma_q))^2 & \text{for} \quad p \neq q \end{cases} \tag{4.7}$$

where

$$a_t(\mu_p^{\mathcal{A}_k}, \sigma_q) = \beta a_{s,t}(\mu_p^{\mathcal{A}_k}, \sigma_q) + (1 - \beta) a_{u,t}(\mu_p^{\mathcal{A}_k}, \sigma_q) \tag{4.8}$$

$\beta \in <0, 1>$ is the **success-use ratio** parameter. It will be shown that all updates decrease this distance. The first update rule (eq. 4.3), has two parts, the active part depending on whether the update concerns the word that was uttered by the agents (σ_i) or another word. The distance function also contains two parts. For each agent, the first part of the association update rule only affects the first part of the distance term, and only for the index i of the current referent (and meaning and word). The effect of the first update on the distance function is the following:

$$d_{\text{after}} - d_{\text{before}} = \sum_{k=1}^{n_a} (1 - a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 - (1 - a_t(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 \tag{4.9}$$

The first part of the update rule can be written as

$$a_{u,t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i) = 1 + \alpha^{n_a}(a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_i) - 1) \tag{4.10}$$

Since the success components of the association are not affected by the update, and using eq. 4.8, $a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i)$ can be written in terms of $a_t(\mu_i^{\mathcal{A}_k}, \sigma_i)$:

$$\begin{aligned} a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i) &= a_t(\mu_i^{\mathcal{A}_k}, \sigma_i) + \\ & \quad (1 - \beta)(a_{u,t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i) - a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_i)) \\ &= a_t(\mu_i^{\mathcal{A}_k}, \sigma_i) + U_{1A} \end{aligned} \tag{4.11}$$

where

$$U_{1A} = (1 - \beta)(\alpha^{n_a} - 1)(a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_i) - 1) \tag{4.12}$$

It can now be seen that for every agent $\mathcal{A}_k$, the difference in eq. 4.9 is below zero:

$$(1 - a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 - (1 - a_t(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 < 0 \tag{4.13}$$

This is true if

$$(1 - a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 < (1 - a_t(\mu_i^{\mathcal{A}_k}, \sigma_i))^2 \tag{4.14}$$

Since $a \in < 0..1 >$, this holds whenever

$$\begin{array}{rcl} a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_i) & > & a_t(\mu_i^{\mathcal{A}_k}, \sigma_i) \\ a_t(\mu_i^{\mathcal{A}_k}, \sigma_i) + U_{1A} & > & a_t(\mu_i^{\mathcal{A}_k}, \sigma_i) \\ U_{1A} & > & 0 \end{array} \tag{4.15}$$

U_{1A} contains three terms, two of which are negative, and is therefore positive, which proves that the first part of the first update rule (eq. 4.3) causes the distance to the attractor to decrease. The second part of the same rule specifies that for $j \neq i$

$$\begin{array}{rcl} a_{u,t+1}(\mu_i^{\mathcal{A}_k}, \sigma_j) & = & 0 - \alpha(0 - a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_j)) \\ & = & \alpha a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_j) \end{array} \tag{4.16}$$

Thus,

$$a_{u,t+1}(\mu_i^{\mathcal{A}_k}, \sigma_j) = a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_j) + U_{1B} \tag{4.17}$$

where

$$U_{1B} = (1 - \beta)(\alpha - 1) a_{u,t}(\mu_i^{\mathcal{A}_k}, \sigma_j) \tag{4.18}$$

For $j \neq i$, the increase in the distance to the attractor (eq. 4.6) is negative if

$$\begin{array}{rcl} (-a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_j))^2 & < & (-a_t(\mu_i^{\mathcal{A}_k}, \sigma_j))^2 \\ a_{t+1}(\mu_i^{\mathcal{A}_k}, \sigma_j) & < & a_t(\mu_i^{\mathcal{A}_k}, \sigma_j) \\ a_t(\mu_i^{\mathcal{A}_k}, \sigma_j) + U_{1B} & < & a_t(\mu_i^{\mathcal{A}_k}, \sigma_j) \\ U_{1B} & < & 0 \end{array} \tag{4.19}$$

U_{1B} contains two positive and one negative term, and therefore satisfies this condition, which together with the above result shows that the first update rule strictly decreases the distance to the attractor.
The analysis of the second update rule is very similar to that of the first one. The equivalent of U_{1A} in eq. 4.11 is

$$U_{2A} = \beta(\alpha^{n_a} - 1)(a_{s,t}(\mu_i^{\mathcal{A}_k}, \sigma_i) - 1) \tag{4.20}$$

U_{2A} has a positive sign. Thus, substituting it for U_{1A} in eq. 4.15 gives the same result, i.e. the distance to the attractor is decreased as a result of the update. The equivalent of U_{1B} is:

$$U_{2B} = \beta(\alpha - 1)a_{s,t}(\mu_j^{\mathcal{A}_k}, \sigma_i) \tag{4.21}$$

U_{2B}, like U_{1B}, has a negative sign, and therefore does not affect the outcome when substituted in eq. 4.19. Finally, the third update rule also parallels the second part of update rule one, but with the following term substituted for U_{1B}:

$$U_3 = \beta(\alpha - 1)a_{s,t}(\mu_i^{\mathcal{A}_k}, \sigma_j) \tag{4.22}$$

Once more, the change does not affect the sign, and the effect of the third update rule also is to decrease the distance to the attractor. This result concludes the theoretical investigation of the second condition of the attractor definition.

The first condition of this definition requires that when the system is *in* the attractor, it must remain there forever. This is equivalent to the requirement that U_{1A}, U_{1B}, U_{2A}, U_{2B}, and U_3 are all equal to zero. This condition is indeed satisfied, which can be seen as follows. U_{1A} and U_{2A} both contain a term $a(\mu_i, \sigma_i) - 1$, where the subscript in U_{1A} indicates the use component and that in U_{2A} points to the success component. In the attractor, the associations $a(\mu_i, \sigma_i)$ must all equal 1, and since association strength are a linear combination of use and success components, both of which are limited to the interval [0..1], both these components must equal one. Thus, the $a(\mu_i, \sigma_i) - 1$ terms in U_{1A} and U_{2A} equal zero, bringing the result of both expressions to zero.

For U_{1B}, U_{2B}, and U_3, the situation is analogous; these expressions all contain a term $a(\mu_p, \sigma_q)$, where $p \neq q$. Given the precondition that the system is located in the attractor, all such association strengths must equal zero, which implies that the remaining three expressions also yield zero. Together with the previous result, this establishes that none of the association update rules alters the location of the system in phase space when this location is an attractor.

In summary, in this section it has been shown theoretically that all association strength updates cause the distance of the system to the nearest attractor to decrease, except when the system is located *in* the attractor, in which case the updates do not affect the location in phase space. Together with the third condition of the attractor definition, which was shown trivially true, this constitutes a theoretical proof of the statement that the deterministic version of the system of communication that has been described has point attractors that correspond to perfect systems of communication.

4.4 Stochastic Version of the System

The basic concepts from dynamical systems theory assume that the dynamical system is deterministic. For that reason, the above investigation of attractors in the communication system was based on a deterministic variant of the system. However, as was noted, this deterministic version of the system differs from the basic, stochastic system in that communication does not always develop spontaneously; in half of the runs of the experiment that was reported, high fidelity was only reached after a perturbation that put the system near the nearest corner of the phase space. When these perturbations are not performed, the phase of the system keeps wandering indefinitely, without any further development of communication. The results of this modified experiment, which only differs from that in section 4.3.5 in that no perturbations are performed, are shown in figure 4.7. The graph also shows that the sudden drop in the distance to the attractor of figure 4.4 was indeed due to the perturbation at time step 50,000, since the corresponding line in figure 4.7 continues to vary around a distance of just above 1.

A logical question to ask given the problematic development of communication in the deterministic system, is which of the modifications caused the problems. The most fundamental change to the basic system was the removal of stochasticity in word production. Therefore, the hypothesis that will be tested in the following is that the stochasticity in the original system actually played a valuable role in bringing about communication. To

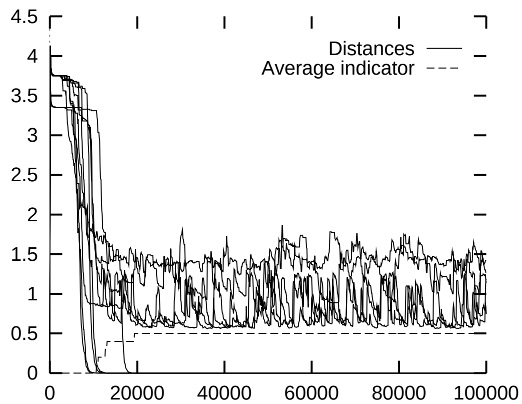


Figure 4.7: Distance of the deterministic systems' state to the nearest hypercube corner over time for ten different runs. Only in half of the cases the system spontaneously develops ideal communication and moves towards an attractor, as verified by the average indicator.

test this hypothesis, a system will be used that is in principle equal to the deterministic system, except that the modifications that were made to ensure determinism are reversed. This concerns the first three modifications of section 4.3.2:

- Action Selection
- Signal Production
- Situation Determination

The other modifications of section 4.3.2 are maintained.

4.4.1 Choice of the State Variables

In the deterministic system, the variables were the association strengths between meanings and words; since communicative behavior does not change as long as the word which has the highest association strength remains the same for each referent, this yields more information about the system than the production probabilities. For example, if the association strengths of the words associated with a referent of a particular agent are 0.01, 0,05, 0.03, and 0.89, the communicative behavior of this agent is much more stable than when they are 0.25, 0.24, 0,25, and 0.26 respectively. In the

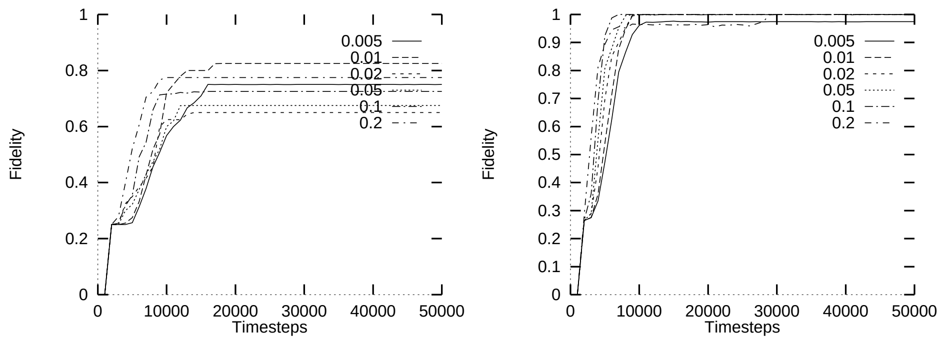


Figure 4.8: Fidelity averaged over ten runs for six different values of the **max-last-error-for-signals** parameter for the deterministic (left) and stochastic (right) system.

deterministic system, this information would not be visible if the production probabilities (0, 0, 0, and 1 respectively) would be used as the state variables.

In the stochastic system however, the case is different; here, changes in the magnitudes of the association strengths do influence communicative behavior when the system is in the neighborhood of a pseudo-attractor. Therefore, the production probabilities will be used as the variables determining the state of the stochastic system.

4.4.2 Comparison Between the Deterministic and the Stochastic System

The only remaining parameter that needs to be tuned differently for the two systems is the **max-last-error-for-signals** parameter. To gain insight into the behavior of both systems, experiments have been conducted for the following parameter values: 0.005, 0.01, 0.02, 0.05, 0.1, 0.2. For each parameter value, data has been gathered for ten runs with both systems. The principal measure indicating communication quality, fidelity, is shown for all experiments in figure 4.8.

The graph shows that for the stochastic system (on the right), the development of communication is consequently observed over the whole range of parameter values that has been examined. Only for the very lowest value of 0.005 is the average fidelity substantially lower than one. For this parameter setting, one of the ten runs reached a fidelity of 0.75, while the fidelity of the other nine runs all exceeded 0.99. For the deterministic system on the other hand (on the left), the average fidelity exceeded 0.8 in only one of the experiments. This was the case for a parameter value of 0.01, the value that was used in the experiments with the deterministic system that have been presented in this chapter. In summary, these experiments show that indeed stochasticity appears to be a useful characteristic of the system that is of positive influence on the development of communication.

This is in line with the findings reported in (Steels & Kaplan, 1998). There, the effect of stochasticity on concept formation, word production, and word interpretation has been investigated. It was found that there is an upper bound to the amount of stochasticity that can be tolerated, and that stochasticity causes and maintains language variation.

It was remarked at the beginning of this section that the basic concepts of dynamical systems theory that have been described assume a deterministic system. However, as the stochastic system evidently is more likely to develop good systems of communication, the question of what the dynamics of this stochastic version are like suggests itself. In figure 4.8, the distance to the nearest hypercube corner is plotted for both the deterministic (left) and stochastic (right) system and over the same set of values for the **max-last-error-for-signals** parameter.

In the left graph, none of the parameter values causes the system to consistently move towards an attractor spontaneously; if this were the case at least one of the lines would converge to zero. The lowest line corresponds to parameter value 0.01, and hence displays the average of the ten distances in figure 4.7, five of which converged to an attractor.

To the right, the corresponding plot for the stochastic system is shown. The experiment with the highest average distance is, as was to be expected, that with parameter value 0.005. In this experiment, the average distance remains below 0.2 once converged, while in the deterministic experiment even the most successful experiment has an average distance of around 0.5.

4.4.3 Pseudo-Attractors in the Stochastic System

Given the improvement in the development of communication induced by the introduction of stochasticity, a natural question is whether the stochastic system also has points towards which the state of the system is drawn when approaching these points. In the deterministic system, these points were called attractors. A condition of the definition of attractors was that the distance from the system to an attractor tends to zero for initial conditions within some neighborhood of the attractor. In the stochastic system, this can not be guaranteed; if the system is not *in* the pseudo-attractor but

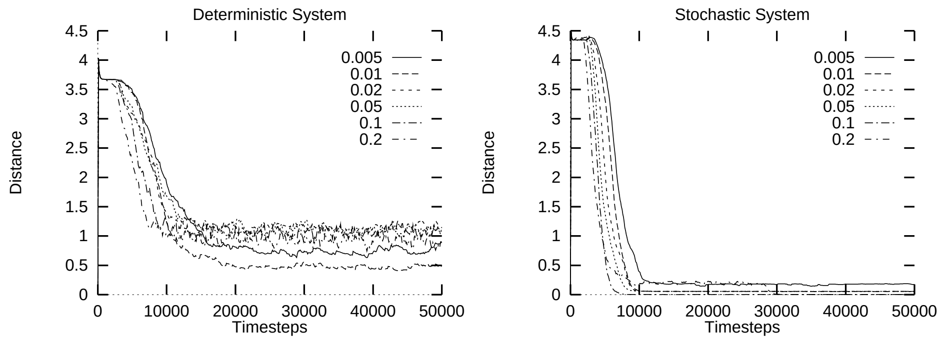


Figure 4.9: Distance to the nearest hypercube corner averaged over ten runs for six different values of the **max-last-error-for-signals** parameter for the deterministic (left) and stochastic (right) system.

very near to it, associations must exist whose probability of being selected for production is almost zero, but not quite. Thus, the corresponding words have a nonzero change of being produced. If the coordinate of the pseudo-attractor corresponding to the word has a value near zero, this implies that the associations strengths of the agents that receive this word will be increased, which increases the distance to the pseudo-attractor. In short, it cannot be guaranteed that the distance to the pseudo-attractor will tend to zero; there may always be an unlucky combination of random values that prevents the system from moving towards the pseudo-attractor.

The theoretical argument that has been given implies that the behavior of the stochastic system around pseudo-attractors will not be mathematically equivalent to that of a deterministic system near attractors. However, the behavior may well be very similar. This is suggested by graph 4.9, which shows that the stochastic system is strongly attracted to corners of the hypercube, even more so than the deterministic system. To test this hypothesis an experimental investigation of the presence of pseudo-attractors will be performed, analogous to that of the deterministic system reported in section 4.3.5.

For stochastic systems, the distance to an attractor continues to fluctuate, and cannot be proved to converge to zero from some neighborhood of initial conditions. However, it may well be the case that such a system is attracted by a point, quickly moves towards it, and stays within a small *neighborhood* of the point. If this operational test is satisfied, this neighborhood may be viewed as a pseudo-attractor. To test whether a neighborhood acts as a pseudo-attractor, tests analogous to that of section

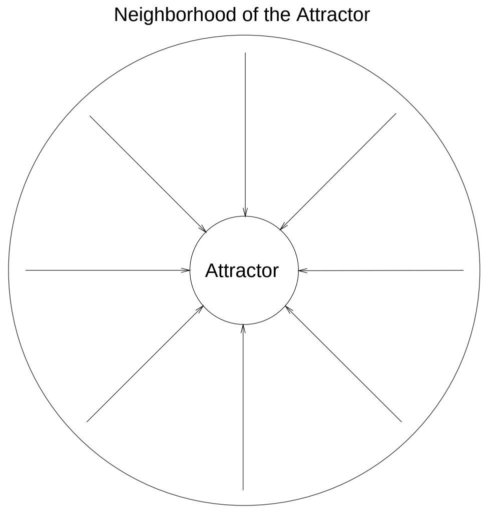


Figure 4.10: Schematic rendering of a pseudo-attractor in the stochastic system. When the initial state of the system is within some neighborhood of the pseudo-attractor (outer sphere) its state moves into a smaller neighborhood of it (inner circle). Once within the pseudo-attractor, the system remains within it.

4.3.5 may be carried out; the system should, when placed within a larger neighborhood of the pseudo-attractor, with high probability move into the smaller neighborhood determining the pseudo-attractor, and it should, once within this neighborhood, with high probability stay within it, see figure 4.10. If these two conditions hold, the small neighborhood may be viewed as a pseudo-attractor.

For the successful parameter values (0.1 and 0.2), the distances of the individual runs stay well below 0.02 once converged. Therefore, this value will be used as the radius defining the pseudo-attractors of the stochastic system. In the following experiments, a value of 0.1 will be selected for the **max-last-error-for-signals** parameter, since the system reaches low distances earliest for this parameter value. A perturbation is performed by putting the system outside the neighborhood determined by the pseudo-attractor, but within a larger neighborhood of the center of the pseudo-attractor. For the radius of this larger hypersphere, a value of 0.1 will be used. Thus, the experiments should test whether the system, when within a radius of 0.1 of the center of the pseudo-attractor, moves into the pseudo-attractor (radius 0.02), and whether the system, once within this radius

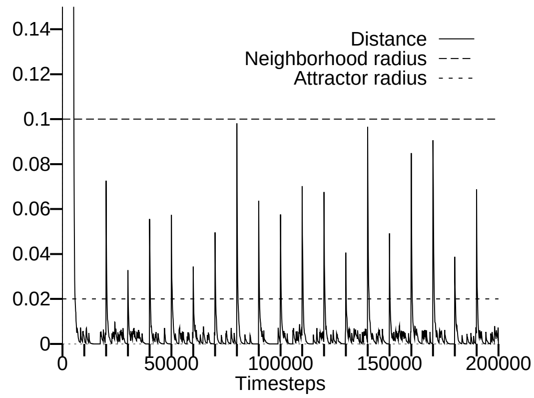


Figure 4.11: Distance of the stochastic system's state to a pseudo-attractor over time for one of the runs. After every perturbation, the system quickly recovers and reenters the pseudo-attractor.

of 0.02, remains there. These two conditions are the counterparts for the stochastic system of conditions one and two of the deterministic attractor definition.

Figure 4.11 shows a run of the first perturbation experiment. At time step 5900, the distance of the system first drops below 0.02. Until time step 20,000, the distance to the center of the pseudo-attractor continues to vary, but does not exceed 0.02, i.e. system remains within the pseudo-attractor. Starting with time step 20,000, and repeated at every multiple of 10,000 time steps, a perturbation is performed. The system is moved to a location outside the pseudo-attractor, but within a neighborhood with radius 0.1. As the graph shows for one of the runs, the system quickly recovers from each perturbation, and reenters the pseudo-attractor within several hundreds of time steps. The graph is typical for the ten runs.

As has been explained, the stochastic nature of the system implies that sometimes the behavior will differ from what is generally observed. This was indeed observed on one occasion during the ten runs, see figure 4.12. In that case, the system did not immediately reenter the pseudo-attractor after a perturbation, but first moved outside the neighborhood of the pseudo-attractor. However, before the next perturbation, the system has already reentered the neighborhood of the pseudo-attractor again. Since all words have nonzero production probabilities, such exceptions necessarily exist for any neighborhood, and hence limiting the neighborhood to a smaller region would not make a structural difference, apart perhaps from prolonging the

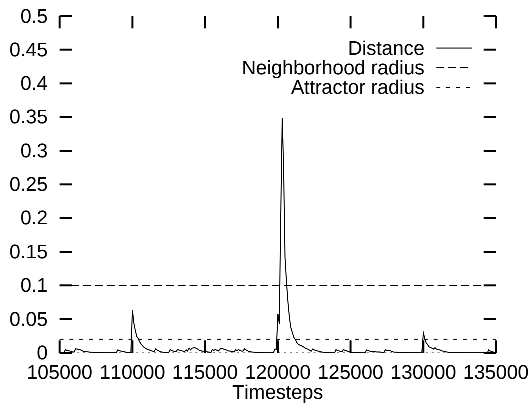


Figure 4.12: Distance of the stochastic system's state to a pseudo-attractor over time for one of the runs. At each perturbation, where the system is taken out of the pseudo-attractor (indicated by the ticks), it quickly recovers and reenters the pseudo-attractor.

time required to discover such exceptions.

Figure 4.13 shows a run of the second perturbation experiment, where the system is moved to a random location *within* the pseudo-attractor. After every perturbation in every run the system remained in the pseudo-attractor, with (again) one exception where the system temporarily escaped from the pseudo-attractor, but after 1000 time steps, long before the next perturbation, it had reentered the pseudo-attractor again.

Together, the two perturbation experiments that were described are an experimental demonstration of the hypothesis that the stochastic system has pseudo-attractors that correspond to good systems of communication.

4.5 Behavior of the Stochastic System

The development of communication has been viewed as a dynamical system. A relevant question then is how the behavior of a system of communication changes under the influence of control parameters. A very influential control parameter is the temperature T, which controls the amount of exploration in signal production (see the description of the algorithm in section 3.3). In the following experiment, its influence on fidelity will be determined. Fidelity characterizes the quality of communication in a single measure; such measures are sometimes called *order parameters* in dynamical systems research.

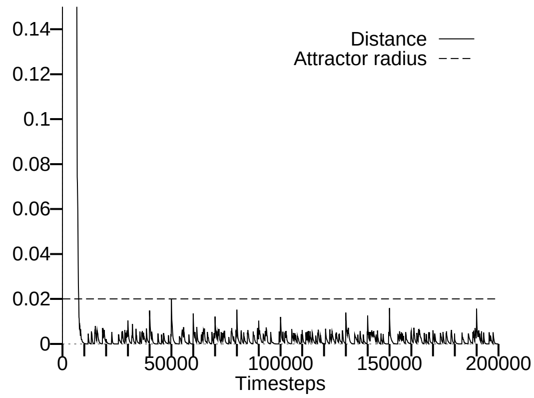


Figure 4.13: Distance of the stochastic system's state to a pseudo-attractor over time for one of the runs. At each perturbation, where the system is moved to a random position within the pseudo-attractor (indicated by the ticks), it remains within the pseudo-attractor.

When the relationship between temperature and the development of communication has been clarified, it will be interesting to see how this relationship manifests itself in phase space. In order to better appreciate the phase space behavior of communication systems, it will be useful to first consider the phase space behavior of the standard stochastic system of communication. Some insight into this behavior has been provided by the perturbation experiments in the previous section. In the next section, a typology of phase plots for communication systems will be introduced, and examples of the different types of phase plots are given to provide insight into the phase space behavior of the system under normal conditions. Finally, in section 4.5.3, the influence of temperature on the phase space behavior of a communication system is examined.

4.5.1 The Effect of Temperature on the Fidelity of Communication

In this section, the effect of temperature on fidelity is investigated. The form of the experiment that allows this is straightforward; for each data point within a chosen range, a number of simulations are performed, each with different random seeds. With the results of these experiments, the average fidelity and the variance, measured using the sample standard deviation, can be determined as a function of temperature. Although the

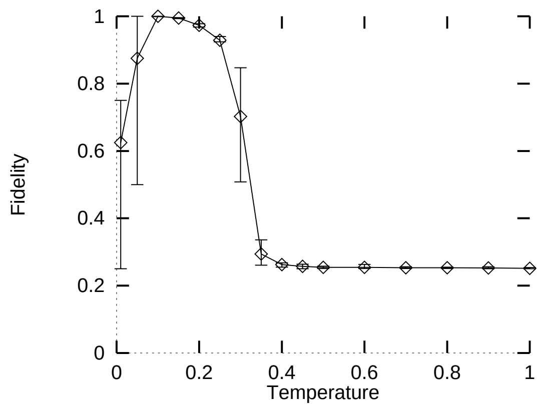


Figure 4.14: The effect of temperature of the fidelity of communication. The error bars show the minimum and maximum values of ten runs; the line shows the average value over these runs.

procedure is simple, the experiment is computationally intensive, since for each temperature, ten complete experimental runs are carried out. The results are shown in figure 4.14.

As the graph shows, the temperature can be too low and too high. For temperatures below 0.1, specificity drops, while consistency remains high. This means that some words are used in combination with multiple meanings. Low temperatures have the effect that words that occur slightly more often than other words have a much higher probability of being selected for production. Since this may happen for some word in combination with several meanings, words may become associated with multiple meanings, which implies low specificity. The fact that words that become slightly stronger associated than other words will be selected much more often and hence become associated even stronger results in consistent selection of one word for each meaning, which explains why consistency is high.

For temperatures higher than 0.25, specificity is also low, but consistency is low as well. This is precisely what would be expected for high temperatures; since word production becomes more explorative, meanings will be expressed by different words at different times, which implies low consistency and low specificity. For temperatures between these values, i.e. within the interval $[0.1, 0.25]$, the fidelity of communication is high, and remarkably stable; as the error bars indicate, there is only a very small difference between the minimum and maximum values measured.

An interesting interpretation of these results arises once it is realized that stochastic signal production with very low temperatures approaches deterministic signal production; the loss of exploration as temperature decreases results in a greedy selection of words that is identical in the limit to the behavior of the deterministic system. Therefore, the shape of the curve depicting fidelity as a function of temperature is not only useful for determining suitable values of this control parameter of the stochastic system. Moreover, it provides an explanation of *why* the deterministic system is less effective in developing communication, namely that it lacks exploration. When some word is the preferred (most strongly associated) word for multiple referents, as was observed in the experiments with the deterministic system, greedy selection results in a deadlock. The success components of this word's associations with each meaning are increased in some interactions, but, due to its association with the other meanings, decreased in others. Indeed, in the deterministic system, success components of such ambiguous words have been observed to remain at intermediate values (between 0.2 and 0.6), while the use components of these same associations had already converged to 1.In such cases, exploration allows new words to be used sporadically, and since these are not restrained by their associations with other meanings, their association values can increase and surpass those of the ambiguous word.

4.5.2 Typology of Phase Plots

Phase plots of the following combinations of associations will be distinguished:

1. The associations of two different words σ_i and σ_j with one meaning μ: s(σ_i, μ) against s(σ_j, μ).

2. The associations of one word σ with two different meanings μ_i and μ_j: s(σ, μ_i) against s(σ, μ_j).

3. The associations between one word σ and one meaning μ of two different agents $\mathcal{A}_i$ and $\mathcal{A}_j$: $s_{\mathcal{A}_i}(\sigma, \mu)$ against $s_{\mathcal{A}_j}(\sigma, \mu)$.

Different combinations are possible (e.g. plotting the association between one word and one meaning of one agent against the association between a different word and a different meaning of another agent), but will not be used.

It will be noted that type 3 plots require the notion of a corresponding meaning for pairs of agents. This notion is only relevant when concept

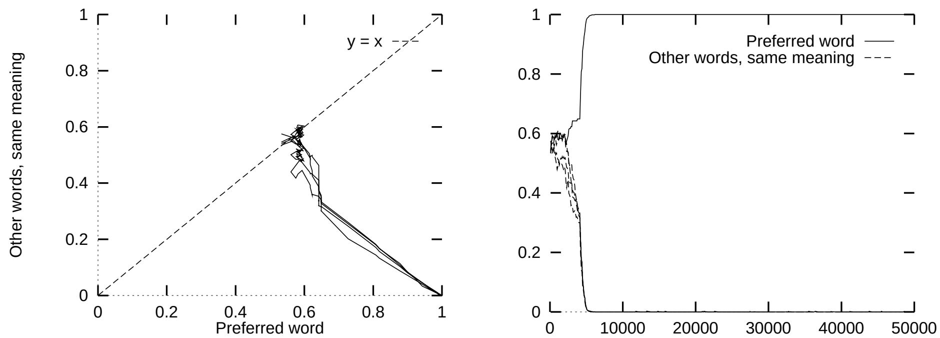


Figure 4.15: Left: type 1 phase plot. The diagram shows combinations (superimposed) of association strengths that occurred during the experiment, where one word (horizontal) ends up to be a preferred word, and the other word is any word associated with the same meaning. Right: time series of the same data.

formation yields conceptual systems of which both distinctiveness and parsimony are one. This is the case in all of the following experiments. The term *preferred word* here refers to a word that becomes most strongly associated during an experiment; in each of the experiments reported here, there is a single word for each referent (and therefore meaning) for which this applies.

The type of a phase plot is not determined by its construction; all depict the combinations of two association strengths that occurred during an experiment. Rather, it is determined by the role that the particular associations that are represented take in the course of the experiment.

Figure 4.15 (left) shows a phase plot of type 1. For each of the curves, the horizontal dimension corresponds to the preferred word of one of the meanings, and the vertical dimension corresponds to another word that is associated with the same meaning. Technically, each curve is a projection of the complete state of the system onto two dimensions, and the plot shows a superposition of four of these projections. The phase plot reveals that the behavior of the system in these four different projections is similar once the preferred word has become strongly associated (for horizontal coordinate > 0.65). This implies that the association strengths of the non-preferred words decrease with increasing association strength of the preferred word in a similar manner.

The dotted line shows the line $y = x$, where the association strength of the horizontal and vertical word are equal. The largest amount of variation

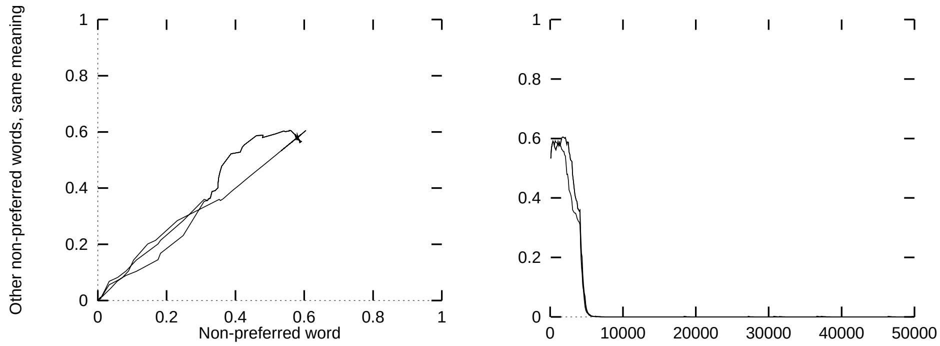
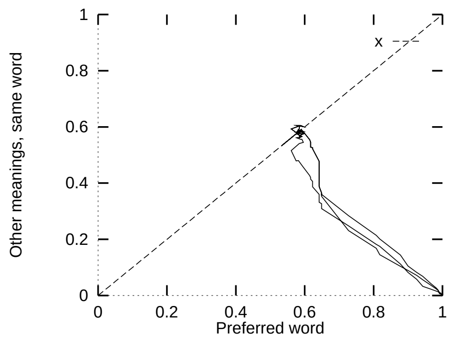


Figure 4.16: Left: type 1 phase plot. The diagram shows combinations (superimposed) of the association strength of a non-preferred word for a particular meaning and the association strengths between other non-preferred words and the same meaning. Right: time series of the same data.

in productive behavior is to be expected in the vicinity of this line, since it represents states where the different words are equiprobable and will therefore both be selected regularly. The graph shows that once the distance to this line has reached a certain threshold (around x = 0.65), mostly due to the decrease of the strength of the vertical association, the horizontal association increases until it reaches one, while the association strengths of the other words simultaneously decrease.

The right side of the figure shows the corresponding time series for one of the meanings; the other meanings have been left out, in order to show the competition between the different words. This type of diagram is called a *form-meaning competition diagram.* In all of the following phase plots, each data point in the left graph corresponds to two data points at a particular time step in the right graph. The vertical position of one of these points determines the horizontal position of the point in the phase plot, while that of the other determines the vertical position. Thus, the time dependence of the data is not visible in a phase plot, apart from the relation between successive points, which is represented by connection lines. What is gained is among other things that similar phenomena, determined as transitions from combinations of association strengths to other combinations of association strengths, are all represented at the same region in a phase plot, and allow one to see what transitions occur and which do not. Also, when the set of initial conditions of the system is constrained, it may be observed that particular regions of the phase space, e.g. here those where several words are strongly associated with a particular meaning, or



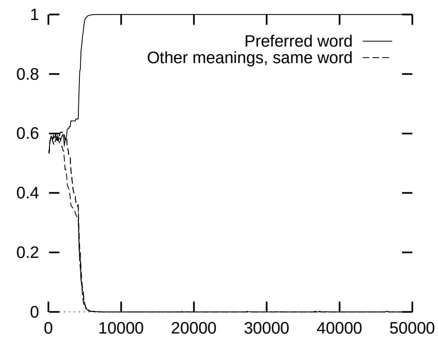


Figure 4.17: Left: type 2 phase plot. The diagram shows combinations (superimposed) of association strengths that occurred during the experiment, where one word (horizontal) ends up to be a preferred word, and the associations concern the same word but another meaning of the same agent. Right: time series of the same data.

one word with several meanings, are not visited in practice. It is important to realize however that this is only a result of a limited set of initial conditions, since any point in phase space lies on *some* trajectory. The behavior in phase space of a system does not yield information that is not present in the corresponding time series, but rather gives a different view of the same processes, which can sometimes be enlightening.

Figure 4.16 (left) shows all combinations (superimposed) of the association strength of a non-preferred word for a particular meaning and the association strengths between other non-preferred words and the same meaning. There appears to be a relation between the different association strengths depicted, since the curves only occupy a restricted region of the phase plot (around the line $y = x$). However, this relation is not a direct effect of the interaction between the associations that are depicted. Rather, it is a result of the relation between each non-preferred word and the preferred word for the same meaning; as the association strength of the preferred word increases, the strengths of all other associations decrease, as was seen in the previous type 1 plot. These simultaneous decreases cause the apparent relation between the non-preferred associations.

Figure 4.17 (left) shows a phase plot of type 2. For each of the curves, the horizontal dimension corresponds to the preferred word of one of the meanings, and the vertical dimension corresponds to associations of the word with the other meanings. The right side of the figure shows the corresponding time series. Again, the line $y = x$ is drawn. Once the

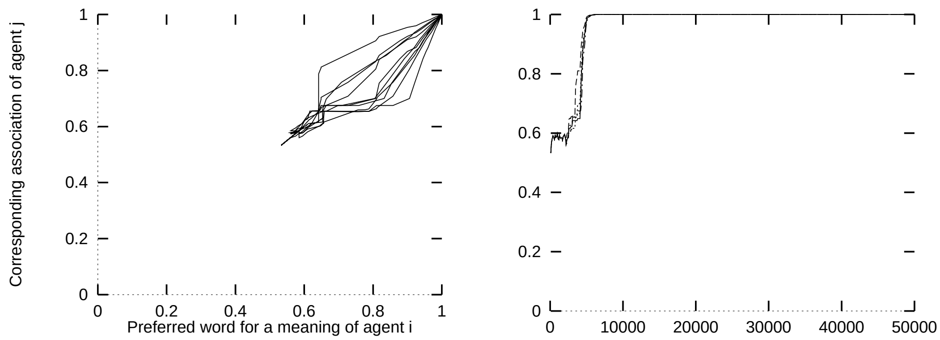


Figure 4.18: Left: type 3 phase plot. The diagram shows combinations (superimposed) of the association strength of the preferred word for a particular meaning for an agent i and the association strength between the same word and corresponding meaning for an agent j, for all combinations of i and j. Right: time series of the same data.

distance to the line $y = x$ grows, the word comes to be the preferred word for the horizontal meaning, and the association with other meanings decrease.

Figure 4.18 (left) shows a preferred association, i.e. an association between one of the meanings and the word that comes to be its preferred word, for all combinations of two different agents. Since there are $n_a = 5$ agents, there are $({n_a}^2 - n_a)/2 = 10$ such combinations. The corresponding time series (right) shows how the associations develop over time for each agent.

4.5.3 The Influence of Temperature on Phase Space Behavior

Now that the behavior of the communication system in phase space under normal conditions has been visualized, the effect of changes to the normal conditions can be examined. Because of its strong influence on the behavior of communication systems, the parameter whose phase space behavior will be investigated is the temperature. A high temperature means that signals with a high association strength will be produced more frequently, but signals with lower association strengths will also be produced, albeit less frequently. For increasingly low temperatures, the strongly associated signals are increasingly preferred, up to a point where, for the lowest temperatures, signals other than the most strongly associated ones are hardly produced anymore. The decrease in variation associated with a decreasing temperature relates to a technique known as *simulated annealing*, where a system initially has a high temperature and 'cools down' over time. Here however, the temperature of each single experiment is constant over time; the subject of investigation is the effect of temperature as a given control parameter of the system, rather than as a time-varying quantity.

In each phase plot, all combinations between the first five words have been plotted for all four meanings. This gives $\binom{5}{2} = 10$ word combinations times 4 meanings = 40 trajectories, all of which are superimposed. If communication develops successfully, some of the combinations will concern one preferred word and one non-preferred word; these combinations have data points in the lower right or upper left corner, where one dimension is high (near one) and the other one low (near zero). In the upper right corner no points should be expected. Furthermore, since there should only be one word per meaning (out of a total of five words) that is strongly associated, the majority of the combinations will concern two non-preferred words, which show in the phase plot as points near the origin.

Figures 4.19 through 4.22 show type 1 phase plots for temperatures varying from 0.02 to 0.4. For the lowest temperature of 0.02, communication did not develop successfully. This can also be seen from graph 4.19 (left), since neither the lower right corner nor the upper left corner contain any points. For the slightly higher temperature of 0.05, the situation is different; here, a good system of communication develops, and indeed figure 4.19 (right) contains trajectories in both the lower right and upper left corners. The development of communication occurs for temperatures up to 0.3. In plots 4.19 through 4.21, corresponding to temperatures 0.05 to 0.3, a gradual process is visible: the trajectories that extend towards the corners of the figure gradually retract. When the temperature increases even more, corresponding to less greedy and more random word production, this process of retraction controls the system to such an extent that it remains within a restricted area between the origin and the middle of the plot; the trajectories to the upper left and lower right corners disappear, which shows that no words become strongly associated anymore. Indeed, the fidelity of communication was low for these temperatures (0.39 and 0.26 respectively, after 50,000 games).

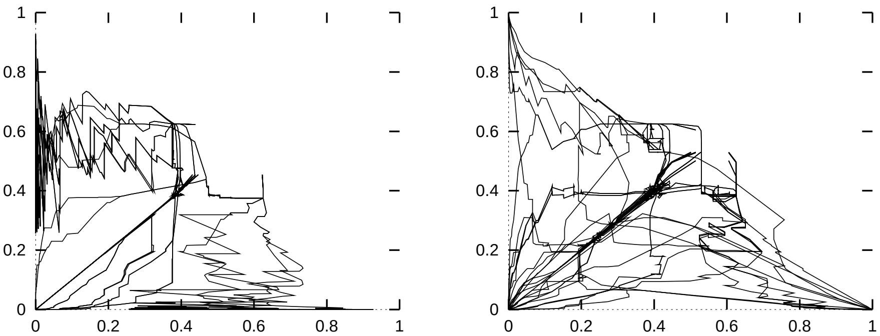


Figure 4.19: Phase space behavior for a temperature of 0.02 (left) and 0.05 (right).

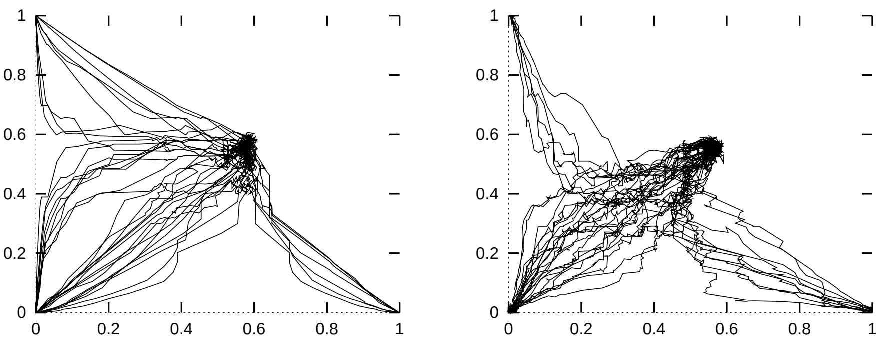


Figure 4.20: Phase space behavior for a temperature of 0.1 (left) and 0.2 (right).

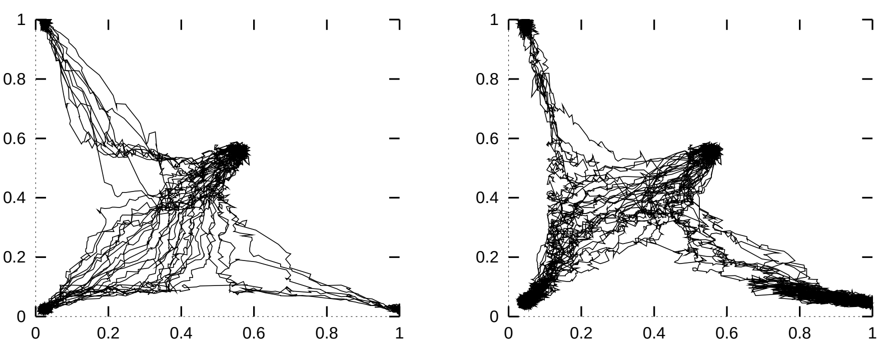


Figure 4.21: Phase space behavior for a temperature of 0.25 (left) and 0.3 (right).

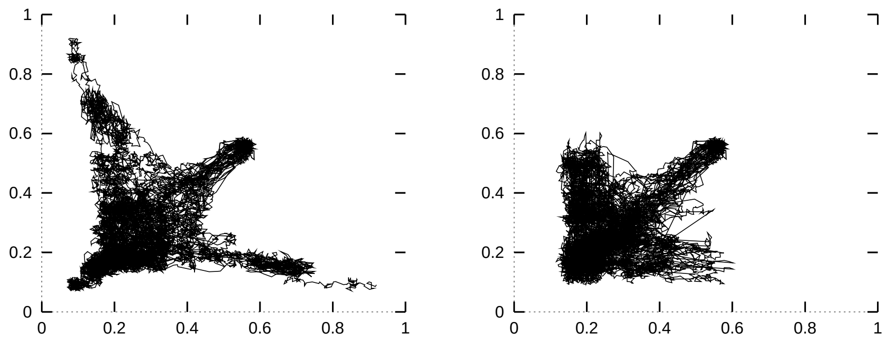


Figure 4.22: Phase space behavior for a temperature of 0.35 (left) and 0.4 (right)

4.6 Conclusions

The investigation has shown that a deterministic version of the communication system contains point attractors that correspond to ideal communication. This property has been shown both mathematically, by showing that all association updates decrease the distance to the attractors and that attractors are fixed points, and experimentally, by demonstrating that the system moves monotonically towards the attractor and that it remains in the attractor when placed there.

In itself, the deterministic system is of limited interest; the communicative behavior of its agents is constant near the attractor. The value of the deterministic system is in demonstrating the presence of attractors in a system otherwise completely analogous to the stochastic system. Although the stochastic system does not have attractors in the mathematical sense, it was demonstrated that it does contain points that play a similar role, called pseudo-attractors. Experiments analogous to those performed to prove the existence of attractors in the deterministic system have been performed for the stochastic case, and confirmed this similarity. The increased fraction of instances on which the system spontaneously finds a pseudo-attractor suggests that stochasticity may have a positive effect on language evolution. An investigation of the influence of the *temperature* parameter showed that low temperature values, corresponding to little exploration, impede the development of communication. The greedy selection mechanism of the deterministic system can be viewed as selection with zero temperature. This view explains the lower effectiveness in developing communication of the deterministic system; the rigid selection mechanism lacks exploration, which can lead to a deadlock situation. The superiority of the stochastic system was confirmed in an investigation of the **max-last-error-for-signals** parameter, for which it developed communication systems of higher quality across a wide range of parameter settings. Furthermore, a typology of phase plots was described, and the influence of temperature in phase space has been visualized.

The demonstration of the presence of attractors in a communication system should be viewed as an important first step towards a better understanding of the dynamics in such systems, and may hopefully inspire further research. It would be of particular interest to gain a better understanding of the attractor basins, i.e. the regions from which the system converges to an attractor. Since the stochastic system is the more productive version of the two systems that have been described in this chapter, the standard notion of attractor basin can not be applied directly, as this notion assumes determinism; due to the stochastic nature of the system, convergence to an attractor can in general not be guaranteed because stochasticity may always cause the state of the system to follow an unlikely course. This points to the more general issue of the stochasticity of the system. Since the basic notions of dynamical systems theory were not designed for stochastic systems, a possible step forward would be to model the system using mathematics that describe such systems more naturally. A concept called *convergence of measures* would be the proper technique for this. However, it is questionable whether this would really yield more insight into the system than has been gained from the operational criterion that was used to determine the presence of pseudo-attractors. In general, care should be taken here, as the analysis of artificial models is not a goal in itself, and aspects of the behavior of such systems are not of inherent interest. Rather, analysis should be used to gain a better understanding of the systems under study so that it may guide the development of more advanced ones.

Chapter 5

Conclusions

The research in this thesis has addressed the question of how autonomous agents can develop concepts about their environment and develop a system of communication that allows them to exchange information about this environment based on those concepts. An central tenet has been the idea that since agents construct concepts based on experience, agents may not be assumed to possess the same conceptual systems, even if they speak the same language.

In chapter 2, the construction of concepts by autonomous agents has been considered. It was argued that such agents may be assumed to receive evaluative feedback about their behavior. It is clear that such feedback can be used by the agent to adapt its behavior. However, once it is realized that the concepts an agent possesses influence its behavior, a principle for the construction of concepts suggests itself, viz. the use of evaluative feedback as a criterion for concept formation. This idea can be operationalized by determining whether the introduction of a new concept improves the ability of the agent to select appropriate actions.

Several existing methods qualify to be used as concept formation methods following this scheme. One such method has been described in detail and tested in experiments. An interesting aspect of the method is that it simultaneously performs state and action generalization in continuous sensor-action spaces. The effect of noise on evaluative feedback and on sensor information on concept formation has been investigated for various split thresholds. The method was consistently found to construct concepts that capture the different situations of the experimental environment, and hence provides a useful basis for the development of a communication system.

Following related work by Steels (Steels, 1999), the concepts developed by agents are called *meanings*, whereas the objective concepts given by the environment are termed *referents*. An important remark was that this distinction does not imply that referents can always be identified; the observable entities in a population of communicating agents are words, and since some mental representation of their meanings must be present, the meanings must also play a role in an account of communication. The third element, referent, is hypothetical however, and could only play such a substantial role in the analysis because of the nature of the concepts. Since these correspond to situations in the environment, a criterion for ideal concepts is available. In combination with the ease with which simulation experiments can be analyzed, this allowed for determining the referents in the experiments.

The result of concept formation can be expressed as a conditional probability matrix relating meanings to referents. The entropy in the rows and columns of this matrix can be used to calculate measures of the quality of the conceptual system. The *distinctiveness* measure expresses the degree to which a meaning identifies a referent, while the *parsimony* measure indicates to what extent a referent gives rise to a unique meaning (and implicitly whether superfluous meanings have been generated, hence its name).

Given a group of agents that have developed concepts capturing the different situations in their environment, the question facing them is how they can develop a system of communication to share their knowledge. This question has been explored by investigating an algorithm used by each individual agent for updating associations between the situation concepts it has developed and an open set of words. A first important result is that even under the strong commitments that were made, including the inability of agents to pass meanings to other agents and the possibility that the meanings of different agents may be different, coherent systems of communication are formed that allow the agents to share information about their environment.

The algorithm has been described in detail, and its different components have been tested for utility. The tests demonstrated that all components of the algorithm are useful, in the sense that removing any one component decreases the accuracy of communication with statistical significance. The success component is crucial for the development of communication. However, further analysis showed that it is not the success information itself that is crucial, but the lateral inhibition between competing associations. The algorithm achieves this specificity without imposing a one-to-one relation between meanings and words. This property is important, as it was seen to play a role in the development of communication by agents with differing conceptual systems. The feasibility of developing shared communication under those circumstances was demonstrated in an experiment. In this experiment, the algorithm led to a situation where multiple meanings were stably associated with the same word and different words were associated with a single meaning. This took place in such a way that the differences in the conceptual systems were compensated for, and successful communication was highly probable.

In analogy with the treatment of conceptual systems, a communication system can be represented by conditional probability matrices. Based on the entropy in these matrices, principled measures for the quality of the communication system have been derived. The *specificity* measure indicates to what extent a word identifies a referent, while *consistency* measures the consistency with which agents produce a single word for each referent. The measures were found to be useful in analyzing the communication systems that form in experiments, and overcome a fundamental problem with a related measure. Apart from these agent-based measures, the *coherence* and *fidelity* measures of the communicative behavior of a population of agents were described.

The development of communication can be viewed as the behavior of a dynamical system. In this system, each dimension or variable represents the strength of an association between a word and a referent of some agent. Since various entities of each type are involved, this system will usually have a high number of dimensions. Nonetheless, its behavior can be analyzed. A central question is whether such a system has attractors that correspond to perfect communication. Since standard dynamical systems theory deals with deterministic systems, elements of the system that introduced stochasticity were removed for this analysis.

Based on the association update formulas, a mathematical proof of the existence of point attractors has been provided. This finding has been confirmed in perturbation experiments where the system was taken out of the attractor; in all cases, the distance between the state of the system and the nearest attractor decreased monotonically after such a perturbation. Furthermore, when the system is placed exactly in the point attractor, it remains there, thus satisfying the additional criterion for attractors of the definition used.

In itself, the presence of attractors in the deterministic system is of limited interest; once inside the defined neighborhood, the communication behavior of the system is constant. However, the analysis does provide a basis for a better understanding of the stochastic system. This system is more interesting as it develops communication under a wider variety of conditions than the deterministic system and with a higher average accuracy. In order to examine the idea that this system has points playing a similar role, called pseudo-attractors, an operational criterion was used that allows the system to remain within a *neighborhood* of a pseudo-attractor instead of converging completely towards it. The same experimental procedure showed that the behavior of the stochastic system with respect to these points was analogous to that of the attractors in the deterministic system.

Finally, the effect of an important control parameter on communication, the *temperature* governing the exploration factor in word production, has been investigated. The experiment showed that a low temperature resulted in communication systems with low fidelity. This provides an explanation for the advantage of the stochastic system in developing communication. The greedy selection as it is used in the deterministic system corresponds to low temperatures. For low temperatures, exploration is lacking, which can result in a deadlock situation where several referents are associated with the same word. Such deadlock situations were experimentally observed, and the role of exploration in overcoming such situations explains the more stable development of communication in the stochastic system. Furthermore, the temperature has been related to the phase space behavior of the system. Although the communication measures are sufficient to determine the effect of temperature on communication, the phase plots show a complementary view of this effect that improves understanding of these findings. These analyses provide evidence for the idea that it may be useful to view the development of communication as the behavior of a dynamical system.

5.1 Perspective

These final remarks concern possible extensions of this research. Given that the idea of using situation concepts as a basis for communication has been found to work well, more complex variants of the idea may now be considered. A strong limitation in the current work is the form of the concepts and regions representing action-values. These are both solid regions, the latter associated with a single approximation of the value function. One interesting development that would be important for more continuous environments is to use the experiences stored in the action-value nodes to give more local estimates of the value function by using only the points nearest to the query; this would have the additional effect that the resulting function would be smooth instead rather than crisp. Other promising extensions would be to make better use of the interaction history by considering not merely the most recent sensor values, but also sensor values, actions, and rewards of earlier time steps. In this case it is probably a good idea not to consider all of this information as additional features, but to develop features that capture relevant aspects of an interaction history, perhaps based on interaction with the environment. Finally, it is clear that to arrive at more complex concepts, the possible shapes of the concepts must be extended beyond the current hyperrectangles. However, this has the effect of increasing the search space. Once this is realized, it becomes clear that the real question is what *bias* may be usefully employed to search in such large spaces. Although this broad observation does not directly point to particular mechanisms that are promising to investigate, it does suggest a guiding principle for future research.

In accordance with the bottom up principle, the focus with respect to communication has also been on the most basic requirement, being a mechanism that leads to a shared system relating concepts to words. One possible extensions would be to allow communication to influence the concepts; as has been argued, this principle can be very powerful as it allows concepts to spread through a population, allowing agents to benefit from the experience of others. Furthermore, in order to arrive at more sophisticated communication systems, the development of coding schemes or grammars will have to be addressed. This form of communication, where the meaning of an utterance depends on the order of its elements, necessarily addresses an important open issue in adaptive behavior research: the development of internal state. Work in this direction is beginning to emerge, and appears a promising avenue for research.

Summary

The research in this thesis addresses the question of how autonomous agents may develop concepts about their environment and develop a system of communication that allows them to exchange information about this environment based on those concepts. An autonomous agent is a system, in software or in hardware, that receives sensor input from the environment, selects actions, and may receive evaluative feedback reflecting the appropriateness of its actions. Communication is viewed as the transfer of information, in the sense that when a sender sends a message to a receiver, the amount of uncertainty in the receiver's knowledge about its environment decreases as a result of receiving the message. When agents have incomplete knowledge about their environment, communication can be valuable as a means to reduce this uncertainty by sharing information, and can be used to coordinate the actions of agents. Communication is learned during the life time of the agents, and the research concerns the question of how agents may cooperate to arrive at a shared system of communication.

Features of the information available to an agent through its sensors can be used to construct concepts, also called meanings. Constructing concepts based on the requirements posed by the environment is a more flexible approach than fixing the concepts of agents at design time, and may be necessary when the agents are to function in unknown or changing environments.

In this thesis, a particular type of concepts is described, called *situation concepts*. Situation concepts consist of features in the history of interaction between the agent and its environment, which consists of sensor data, actions, and subsequent evaluative feedback. A defining criterion of a situation concept is that it predicts some aspect of the future evolution of the state of the environment, possibly conditioned on the actions the agent may take. Several existing methods, particularly from the field of reinforcement learning, can be viewed as constructing a form of situation concepts. A particular method for constructing a specific type of situation concepts, called

the *adaptive subspace method*, is described. The method uses the current sensor values of an agent as features and develops concepts that specify an interval for each sensor. These concept predict the value of actions the agent can take when its current sensor values are within the specified intervals. The meanings thus formed represent situations, and are especially appropriate for use in communication, since they convey information about the environment.

The development of communication is viewed as the formation of associations between words and the meanings formed by the agents in a population, in such a way that agents tend to use the same word in the same situation. When agents autonomously construct concepts, a consequence is that they may not possess identical concepts. Additional constraints that are respected, such as the commitment that agents have no direct access to the meanings formed by other agents (they can not 'look inside each other's head'), and that no single agent may decide on the system of communication, further complicate the problem of how such a system of associations may come about.

Rather than viewing communication as fixed, it is viewed as a dynamical system. A dynamical system is a mathematical model of a system that changes over time. The variables of this system are the strengths of the associations between words and the meanings of the agents in a population. An algorithm is described in detail that, when used by each individual agent to adapt its associations between words and the situation concepts it has formed, leads to a shared system of communication.

The necessity of different components of the algorithm is shown with statistical significance. Associations are linear combinations of use (the frequency with which a word is observed in a situation) and success (the degree to which the word correctly indicates that its associated situation is the current situation in the environment). Analysis of the success component of the algorithm showed that not the success information itself, but the lateral inhibition between competing associations is crucial for the development of communication. It is experimentally demonstrated how the development of communication can compensate for differences in conceptual systems.

Systematic measures have been introduced to determine the quality of conceptual systems and communication systems. The measures require knowledge of the ideal concepts, called *referents*; although such knowledge is not available in general, simulation experiments often do provide the opportunity for such referents to be determined.

The *specificity* measure for communication is based on the principle that knowledge of a word should yield information (i.e. reduce uncertainty) about the referent (the current situation in the environment), and *vice versa* for the *consistency* measure. In cases of maximal specificity, the information a word yields is complete, and thus *identifies* the situation, whereas in the worst case, the word does not yield any information at all. The measure quantifies these and all intermediate cases. The consistency measure is computed as the extent to which a referent identifies a word, and thus expresses whether for each a referent the agent consistently uses the same word. If both specificity and consistency are high, each agent consistently uses a unique word for each referent. A population measure called *coherence* is used to determine whether different agents use the same words. In combination with the specificity and consistency measures, the experimenter can determine to what degree a perfect system of communication, consistently linking each referent to a unique word, is approximated.

Interestingly, the same principle can be used to evaluate the quality of a conceptual system. Ideally, each concept an agent has formed identifies a single referent in the environment; this is expressed by the *distinctiveness* measure, calculated as the degree to which the meanings an agent possesses distinguish between the different referents. Conversely, *parsimony* expresses the degree to which a referent identifies a meaning. It thus reflects whether the agent has not generated more meanings than necessary. Together, high distinctiveness and high parsimony imply that a conceptual system is ideal in the sense that it approximates a one-to-one relation between meanings and referents.

A contribution is made to the viewpoint of communication as a dynamical system by considering the attractors in the communication system that has been described. A deterministic version of the system is proved mathematically and demonstrated experimentally to have point attractors that correspond to perfect communication. An operational definition of pseudo-attractors is used to demonstrate that the stochastic system has points that play a similar role. Stochasticity is found to be a useful ingredient in the development of communication, in that it avoids deadlocks and results in communication more consistently and under a wider variety of parameter settings. This finding is confirmed by a systematic investigation of the effect of different amounts of stochasticity, regulated by the *temperature* parameter that governs word production. The analysis provides evidence that a dynamical systems perspective on the development of communication is valuable.

Samenvatting

Het onderzoek in dit proefschrift stelt zich de vraag hoe *autonomous agents* concepten kunnen ontwikkelen over hun omgeving en een communicatiesysteem kunnen ontwikkelen dat ze toestaat informatie over hun omgeving uit te wisselen op basis van deze concepten. Een autonomous agent is een systeem, in software of in hardware, dat informatie uit zijn omgeving ontvangt via sensoren, acties kan kiezen, en numerieke beoordelingen van de kwaliteit van zijn gedrag ontvangt. Communicatie wordt gezien als het overbrengen van informatie, in de zin dat wanneer een zender een boodschap naar een ontvanger stuurt, de onzekerheid in de kennis van de ontvanger over zijn omgeving afneemt als gevolg van het ontvangen van de boodschap. Als agents onvolledige informatie hebben over hun omgeving kan communicatie waardevol zijn als middel om deze onzekerheid te verminderen door informatie te delen, en kan het gebruikt worden om de acties van de agents te coördineren. Communicatie wordt geleerd tijdens het bestaan van de agents, en het onderzoek betreft de vraag hoe agents samen kunnen werken om een gemeenschappelijk communicatiesysteem te bewerkstelligen.

Kenmerken van de informatie waarover een agent beschikt via zijn sensoren kunnen gebruikt worden om concepten te construeren, ook betekenissen genoemd. Concepten construeren op basis van de eisen die gesteld worden door de omgeving is een flexibelere benadering dan de concepten vastleggen op het moment van ontwerpen, en kan noodzakelijk zijn indien de agents dienen te functioneren in een onbekende of veranderende omgeving.

In dit proefschrift wordt een bepaald soort concepten, genaamd *situatie concepten*, beschreven. Situatie concepten bestaan uit kenmerken in de geschiedenis van de interactie tussen de agent en zijn omgeving, die bestaat uit sensor gegevens, acties, en daaropvolgende beoordelingen. Kenmerkend voor een situatie concept is dat het een aspect van de toekomstige ontwikkeling van de toestand van de omgeving voorspelt, mogelijk afhankelijk van de acties die de agent kan kiezen. Verscheidene bestaande methoden, in het

bijzonder uit het veld van de *reinforcement learning*, construeren concepten die gezien kunnen worden als situatie concepten. Een specifieke methode om een bepaald soort situatie concepten te construeren, genaamd de *adaptive subspace* methode, wordt beschreven. De methode gebruikt de huidige sensor waarden van de agent als kenmerken en ontwikkelt concepten die een interval voor elke sensor vastleggen. Deze concepten voorspellen de waarde van de acties die de agent kan kiezen wanneer zijn huidige sensor waarden binnen die intervallen liggen. De betekenissen die zo worden gevormd stellen situaties voor, en zijn speciaal geschikt om in communicatie te worden gebruikt, aangezien ze informatie over de omgeving voorstellen.

De ontwikkeling van communicatie wordt gezien als de vorming van associaties tussen woorden en betekenissen die gevormd zijn door de agents in een populatie, zodanig dat de agents hetzelfde woord plegen te gebruiken in dezelfde situatie. Een gevolg van het feit dat agents zelfstandig concepten vormen is dat ze niet noodzakelijkerwijs dezelfde concepten bezitten.

Bijkomende beperkingen die in acht worden genomen, zoals de beperking dat agents geen directe toegangen hebben tot de betekenissen die andere agents gevormd hebben (de agents kunnen niet 'in elkaars hoofd kijken'), en dat het communicatie systeem niet bepaald mag worden door één enkele agent, compliceren de vraag hoe een dergelijk systeem van associaties kan ontstaan.

Communicatie wordt niet gezien als iets wat vast ligt, maar als een dynamisch systeem. Een dynamisch systeem is een wiskundig model van een systeem dat verandert in de loop van de tijd. De variabelen van dit systeem zijn de sterkten van de associaties tussen de woorden en de betekenissen van de agents in een populatie. Er wordt een algorithme in detail beschreven dat, wanneer het door elke individuele agent wordt gebruikt om zijn associaties tussen woorden en de sitatie concepten die het gevormd heeft aan te passen, leidt tot een gemeenschappelijk communicatiesysteem.

De noodzakelijkheid van de verschillende componenten van het algorithme wordt statistisch significant aangetoond. Associaties zijn lineaire combinaties van gebruik (de frequentie waarmee een woord geöbserveerd wordt in een sitatie) en succes (de mate waarin het woord correct aangeeft dat de geassocieerde situatie de huidige situatie in de omgeving is). Een analyse van de succes component van het algorithme toonde aan dat niet de succes informatie zelf, maar de laterale inhibitie tussen associaties cruciaal is voor de ontwikkeling van communicatie. Er wordt experimenteel aangetoond hoe de ontwikkeling van communicatie verschillen tussen conceptuele systemen kan compenseren.

Er worden systematische maten voor het meten van de kwaliteit van conceptuele systemen en communicatiesystemen geintroduceerd. De maten vereisen kennis van de ideale concepten, de zogenaamde *referenten*; hoewel deze kennis in het algemeen niet beschikbaar is, kunnen dergelijke referenten in simulatie-experimenten vaak juist wel worden bepaald.

De *specificity* maat is gebaseerd op het principe dat kennis over een woord informatie op moet leveren (d.w.z. de onzekerheid inperken) over de referent (de huidige situatie in de omgeving), en *vice versa* voor de *consistency* maat. Bij maximale specificiteit geeft een woord complete informatie, en legt het dus de situatie vast, terwijl in het in het slechtste geval helemaal geen informatie geeft. De maat quantificeert deze en alle tussenliggende gevallen. De consistency maat wordt berekend als de mate waarin een referent een woord bepaalt, en drukt dus uit in hoeverre een agent voor iedere referent consistent hetzelfde woord gebruikt. Indien zowel specificity als consistency hoog zijn, gebruikt elke agent een uniek woord voor elke referent. Een populatie-maat genaamd *coherence* wordt gebruikt om te bepalen in hoeverre verschillende agents wel dezelfde woorden gebruiken. In combinatie met de specificity en consistency maten kan de onderzoeker bepalen in welke mate een perfect communicatie-systeem, dat elke referent consistent aan een uniek woord koppelt, wordt benaderd.

Verrassenderwijs kan hetzelfde principe gebruikt worden om de kwaliteit van een conceptueel systeem te bepalen. In het ideale geval zou elk concept dat een agent gevormd heeft een enkele referent in de omgeving moeten aanduiden; dit wordt uitgedrukt door de *distinctiveness* maat, die berekend wordt als de mate waarin de betekenissen die een agent gevormd heeft onderscheid maken tussen de verschillende referenten. *Parsimony* daarentegen geeft aan in hoeverre een referent een betekenis identificeert. Deze maat geeft dus aan of de agent niet meer betekenissen dan nodig heeft gegenereerd. Samen geven hoge distinctiveness en parsimony maten aan dat een ideaal conceptueel systeem wordt benaderd in de zin van een een-op-een verhouding tussen betekenissen en referenten.

Er wordt een bijdrage geleverd aan het gezichtspunt waarbij communicatie als dynamisch systeem wordt gezien door attractoren te beschouwen in het beschreven communicatiesysteem. Van een deterministische versie van het systeem wordt wiskundig bewezen en experimenteel aangetoond dat het punt-attractoren heeft die corresponderen met perfecte communicatie. Er wordt experimenteel aangetoond dat het stochastische systeem pseudo-attractoren heeft die een vergelijkbare rol spelen. Stochasticiteit wordt nuttig bevonden als ingrediënt bij de ontwikkeling van communicatie, aangezien het stagnatie vermijdt en consistenter en onder een grotere verscheidenheid aan parameter instellingen resulteert in communicatie. Deze bevinding wordt bevestigd door een systematisch onderzoek naar de invloed van verschillende hoeveelheden stochasticiteit die geregeld kan worden met behulp van de *temperatuur* parameter, die de productie van woorden reguleert. De analyse verschaft bewijsmateriaal voor het idee dat een dynamisch systemen perspectief op communicatie waardevol is.

References

Ackley, D. H., & Littman, M. L. (1994). Altruism in the evolution of communication. In R. A. Brooks & P. Maes (Eds.), *Proceedings of the fourth international workshop on the synthesis and simulation of living systems.* Cambridge, MA: The MIT Press (A Bradford Book).

Agre, P., & Chapman, D. (1987). Pengi: An implementation of a theory of activity. In *Proceedings aaai-87.* San Mateo, CA: Morgan Kaufmann.

Albus, J. (1981). *Brains, behavior, and robotics.* Peterborough, NH: Byte Books.

Bala, J., De Jong, K., Huang, J., Vafaie, H., & Wechsler, H. (1996). Using learning to facilitate the evolution of features for recognizing visual concepts. *Evolutionary Computation*, *4*(3).

Beer, R. D., Chiel, H. J., & Sterling, L. S. (1990). A biological perspective on autonomous agent design. In P. Maes (Ed.), *Designing autonomous agents* (pp. 169–185). The MIT Press: Cambridge, MA.

Beer, R. D., & Gallagher, J. C. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior*, *1*(1).

Bellman, R. (1957). *Dynamic programming.* Princeton, N.J.: Princeton University Press.

Belpaeme, T. (1999). Evolution of visual feature detectors. In *Proceedings of the first european workshop on evolutionary computation in image analysis and signal processing (evoiasp99, göteborg, sweden).* UK.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509–517.

Billard, A., & Hayes, G. (1997). Learning to communicate through imitation in autonomous robots. In *Proceedings of the 7th international conference on artificial neural networks icann 97.* Springer-Verlag.

Billard, A., & Hayes, G. (1999). Drama, a connectionist architecture for control and learning in autonomous robots. *Adaptive Behavior*, *7*(1).

Blair, A. D., & Pollack, J. B. (1997). Analysis of dynamical recognizers. *Neural Computation*, *9*(5), 1127–1142.

Bonabeau, E., Theraulaz, G., Deneubourg, J.-L., Aron, S., & Camazine, S. (1997). Self-organization in social insects. *Trends in Ecology and Evolution*, *12*(5), 188-193.

Cannon, W. B. (1932). *The wisdom of the body.* New York: Norton.

Changeux, J.-P. (1985). *Neuronal man.* New York: Pantheon Books.

Chapman, D., & Kaelbling, L. P. (1991). Input generalization in delayed reinforcement learning: An algorithm and performance comparisons. In J. Mylopoulos & R. Reiter. (Eds.), *Proceedings of the twelfth international joint conference on artificial intelligence (ijcai-91)* (pp. 726–731). San Mateo, Ca.: Morgan Kaufmann.

Clancey, W. J. (1997). *Situated cognition: on human knowledge and computer representations.* Cambridge, UK: Cambridge University Press.

Crites, R. H., & Barto, A. G. (1996). Improving elevator performance using reinforcement learning. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 1017–1023). The MIT Press.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*, 203,204.

De Boer, B. (1999). *Self-organisation in vowel systems.* Unpublished doctoral dissertation, Vrije Universiteit Brussel, Brussels, Belgium.

De Jong, E. D. (1997a). An accumulative exploration method for reinforcement learning. In S. Sen (Ed.), *Proceedings of the aaai'97 workshop on multiagent learning, available as aaai technical report ws-97-03* (p. 19-24). Menlo Park, California: AAAI Press.

De Jong, E. D. (1997b). Multi-agent coordination by communication of evaluations. In M. Boman & W. V. de Velde (Eds.), *Proceedings of the 8th european workshop on modelling autonomous agents in a multi-agent world maamaw'97.* Berlin: Springer-Verlag.

De Jong, E. D. (1999). Autonomous concept formation. In *Proceedings of the sixteenth international joint conference on artificial intelligence ijcai'99* (p. 344-349). San Francisco, CA: Morgan Kaufmann.

De Jong, E. D., & Steels, L. (1999). Generation and selection of sensory channels. In *Evolutionary image analysis, signal processing and telecommunications first european workshops, evoiasp'99 and euroectel'99 joint proceedings* (p. 90-100). Berlin: Springer-Verlag.

De Jong, E. D., & Vogt, P. (1998). How Should a Robot Discriminate Between Objects? A comparison between two methods. In *Proceedings of the fifth international conference of the society for adaptive behavior sab'98* (Vol. 5, p. 86-91). Cambridge, MA: The MIT Press.

Di Paolo, E. A. (1997). Social coordination and spatial organization: Steps towards the evolution of communication. In P. Husbands & I. Harvey (Eds.), *Proceedings of the 4th european conference on artificial life ecal'97.* Cambridge, MA: The MIT Press (A Bradford Book).

Drescher, G. L. (1991). *Made-up minds. a constructivist approach to artificial intelligence.* Cambridge, MA: The MIT Press.

Duda, R., & Hart, P. (1973). *Pattern recognition and scene analysis.* John Wiley and Sons.

Ebner, M., & Zell, A. (1999). Evolving a task specific image operator. In R. Poli, H.-M. Voigt, S. Cagnoni, D. Corne, G. D. Smith, & T. C. Fogarty (Eds.), *Evolutionary image analysis, signal processing and telecommunications: First european workshop, evoiasp'99 and euroectel'99* (Vol. 1596, pp. 74–89). Goteborg, Sweden: Springer-Verlag.

Eco, U. (1976). *A theory of semiotics.* Bloomington: Indiana University Press.

Edelman, G. (1987). *Neural darwinism.* New York: Basic Books Inc.

Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. Van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (p. 195-225). MIT Press.

Finin, T., Fritzson, R., McKay, D., & McEntire, R. (1994). *Kqml - a language and protocol for knowledge and information exchange* (Tech. Rep. No. CS-94-02). Computer Science Department, University of Maryland, UMBC Baltimore MD 21228: Computer Science Department, University of Maryland and Valley Forge Engineering Center, Unisys Corporation.

Finkel, R. A., & Bentley, J. L. (1974). Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, *4*(1), 1–9.

Fipa content language library. (1999). Geneva, Switzerland. (URL: http://www.fipa.org)

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

Foley, W. A. (1997). *Anthropological linguistics: An introduction.* Blackwell publishers.

Friedman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Transactions on Mathematical Software*, *3*(3), 209–226.

Fritzke, B. (1997, april). *Some competitive learning methods.* (Institute for Neural Computation,Ruhr-Universität Bochum)

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*(1), 1–58.

Gennari, J., Langley, P., & Fisher, D. (1989). Models of incremental concept formation. In J. Carbonell (Ed.), *Machine learning: Paradigms and methods.* MIT / Elsevier.

Gordon, D., & desJardins, M. (1995). Evaluations and selection of biases in machine learning. *Machine Learning*, *20*(1/2), 5-22.

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

Hauser, M. D. (1997). *The evolution of communication.* Cambridge, MA: The MIT Press.

Holland, J. H. (1983). Escaping brittleness. In *Proceedings Second International Workshop on Machine Learning* (p. 92-95).

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). Models of information processing in the basal ganglia. In J. L. D. J. C. Houk & D. G. Beiser (Eds.), (p. 249-270). Cambridge, MA: MIT Press.

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

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia*, *28*, 517-527.

Jackins, C. L., & Tanimoto, S. L. (1980). Oct-trees and their use in representing three-dimensional objects,. *Computer Graphics and Image Processing, 1980*, *14*, 249–270.

Kaelbling, L. P. (1993). *Learning in embedded systems.* Cambridge, MA: MIT-Press.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*.

Katz, J. J. (1981). *Language and other abstract objects.* Oxford, UK: Basil Blackwell.

Kohonen, T. (1995). *Self-organizing maps* (Vol. 30). Berlin: Springer.

Kripke, S. (1972). *Naming and necessity.* Oxford, UK: Blackwell.

Lambrecht, K. (1994). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents.* Cambridge University Press.

Landelius, T. (1997). *Reinforcement learning and distributed local model synthesis.* Unpublished doctoral dissertation, Linköping University, Linköping, Sweden.

Lehmann, E., & D'Abrera, H. (1975). *Nonparametrics. statistical methods based on ranks.* San Francisco: Holden-Day Inc.

Littman, M., & Boyan, J. (1993). A distributed reinforcement learning scheme for network routing. In J. Alspector, R. Goodman, & T. Brown (Eds.), *Proceedings of the 1993 international workshop on applications of neural networks to telecommunications* (p. 45-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacDorman, K. F. (1999). Partition nets: An efficient on-line learning algorithm. In *Icar 99: Ninth international conference on advanced robotics.* Tokyo.

MacDorman, K. F. (2000). Proto-symbol emergence. In *Proceedings of iros-2000: International conference on intelligent robots and systems.* Takamatsu.

MacLennan, B. (1991). Synthetic ethology: An approach to the study of communication. In C. G. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life ii* (Vol. X). Addison-Wesley.

Maes, P. (1989). How to do the right thing. *Connection Science journal, 1*(3).

Marconi, D. (1997). *Lexical competence.* Cambridge, MA: The MIT Press (A Bradford Book).

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

Maynard Smith, J., & Harper, D. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology, 177*, 305-311.

McCallum, A. K. (1996). Reinforcement learning with selective perception and hidden state. *Ph.D. Thesis.*

Michalski, R., Carbonell, J., & Mitchell, T. (1986). *Machine learning: An artificial intelligence approach* (Vol. 2). Los Altos, CA: Morgan Kaufmann Publishers.

Michie, D., & Chambers, R. A. (1968). BOXES: An experiment in adaptive control. In D. E & M. D. (Eds.), *Machine intelligence 2* (pp. 137–152). Edinburgh: Oliver and Boyd.

Mitchell, T. M. (1980). *The need for biases in learning generalizations* (Technical Report No. CBM-TR-117). New Brunswick, New Jersey: Department of Computer Science, Rutgers University.

Mitchell, T. M. (1997). *Machine learning.* McGraw-Hill.

Moore, A. (1991). Variable resolution dynamic programming: Efficiently learning action maps in multivariate real-valued state-spaces. In L. Birnbaum & G. Collins (Eds.), *Proceedings of the eighth international conference on machine learning.* Morgan Kaufman.

Moore, A. W., & Atkeson, C. G. (1995). The parti-game algorithm for variable resolution reinforcement learning in multidimensional state-spaces. *Machine Learning, 21*(3), 199-233.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435–450.

Noble, J. (1998). *The evolution of animal communication systems: questions of functions examined through simulation.* Unpublished doctoral dissertation, University of Sussex, Sussex, UK.

Oates, T., Eyler-Walker, Z., & Cohen, P. R. (1999). *Using syntax to learn semantics: An experiment in language acquisition with a mobile robot* (Tech. Rep. No. Technical Report 99-35). University of Massachusetts, Amherst.

Oates, T., Schmill, M. D., & Cohen, P. R. (1999). Identifying qualitatively different outcomes of actions: Experiments with a mobile robot. In *Working notes of the ijcai-99 workshop on robot action planning.* Stockholm, Sweden.

Oja, E. (1983). *Subspace methods of pattern recognition.* Letchworth: Research Studies Press Ltd.

Oliphant, M. (1997). *Formal approaches to innate and learned communication: Laying the foundation for language.* Unpublished doctoral dissertation, University of California, San Diego, CA.

Pollack, J. B. (1991). The induction of dynamical recognizers. *Machine Learning, 7*, 227–252.

Pulliam, H., & Dunford, C. (1980). *Programmed to learn.* New York: Columbi University Press.

Pullum, G. K. (1991). *The great eskimo vocabulary hoax and other irreverent essays on the study of language.* Chicago: The university of chicago press.

Putnam, H. (1975). *Mind, language and reality* (Vol. 2). Cambridge: Cambridge University Press.

Putnam, H. (1988). *Representation and reality.* Cambridge, MA: The MIT Press (A Bradford Book).

Quinlan, J. R. (1990). Induction of decision trees. In J. W. Shavlik & T. G. Dietterich (Eds.), *Readings in machine learning.* Morgan Kaufmann. (Originally published in *Machine Learning* 1:81–106, 1986)

Randløv, J., & Alstrøm, P. (1998). Learning to drive a bicycle using reinforcement learning and shaping. In *Proc. 15th international conf. on machine learning* (pp. 463–471). Morgan Kaufmann, San Francisco, CA.

Rummery, G., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems.* (Tech. Rep. No. CUED/F-INFENG/TR 166). Engineering Department, Cambridge University.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(2), 210–229. (Reprinted in E. A. Feigenbaum and J. Feldman (Eds.) 1963, *Computers and Thought*, McGraw-Hill, New York)

Sankoff, D., & Kruskal, J. B. (1983). Time warps, string edits and macromolecules; the theory and practice of sequence compari son. In (pp. 23–29). Addison-Wesley.

Santamaría, J. C., Sutton, R. S., & Ram, A. (1998). Experiments with reinforcement learning in problems with continuous state and action spaces. *Adaptive Behavior*, *6*(2), 163-217.

Sapir, E. (1949). *Selected writings.* Berkeley: University of California Press.

Schmidhuber, J. (1991). *Adaptive confidence and adaptive curiosity* (Tech. Rep. No. FKI-149-91). München: Technische Universität München.

Schultz, W., Dayan, P., & Read Montague, P. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593-1599.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *In Press.*

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

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656.

Shavlik, J. W., & Dietterich, T. G. (1990). *Readings in machine learning.* Morgan Kaufmann Publishers.

Steels, L. (Ed.). (1993). *The biology and technology of intelligent autonomous agents.* Berlin: Springer Verlag.

Steels, L. (1996a). Emergent adaptive lexicons. In P. Maes (Ed.), *From animals to animats 4: Proceedings of the fourth international conference on simulating adaptive behavior.* Cambridge, MA: The MIT Press.

Steels, L. (1996b). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the international conference on multi-agent systems icmas'96.* Kyoto, Japan.

Steels, L. (1996c). Self-organizing vocabularies. In C. Langton (Ed.), *Proceedings of alife v, nara, japan.*

Steels, L. (1996d). The spontaneous self-organization of an adaptive language. In S. Muggleton (Ed.), *Machine intelligence 15.* Oxford, UK: Oxford University Press.

Steels, L. (1997a). The synthetic modeling of language origins. *Evolution of Communication*, *1*(1), 1-34.

Steels, L. (1997b). The origins of syntax in visually grounded robotic agents. In M. Pollack (Ed.), *Proceedings of the 15th international joint conference on artificial intelligence.* Los Angeles, CA: Morgan Kauffman Publishers.

Steels, L. (1998). The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, *1*, 169-194.

Steels, L. (1999). *The talking heads experiment.* VUB Artificial Intelligence Laboratory.

Steels, L., & Brooks, R. (Eds.). (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Steels, L., & Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Proceedings of artificial life vi.* Los Angeles: MIT Press.

Steels, L., & Kaplan, F. (1999). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models.* Cambridge University Press.

Steels, L., & Vogt, P. (1997). Grounding adaptive language games in robotic agents. In C. Husbands & I. Harvey (Eds.), *Proceedings of the fourth european conference on artificial life.* Cambridge MA and London: The MIT Press.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos.* Reading, MA: Addison-Wesley Publishing Company.

Sutton, R. S. (1988). *Learning to predict by the methods of temporal differences* (Tech. Rep. No. TR87-509.1). GTE Laboratories.

Sutton, R. S. (1990). Reinforcement learning architectures for animats. In J.-A. Meyer & S. W. Wilson (Eds.), *From Animals to Animats 1. Proceedings of the First International Conference on Simulation of Adaptive Behavior (SAB90)* (pp. 288–296). Cambridge, MA: A Bradford Book. MIT Press.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. *Advances in Neural Information Processing Systems*, *8*, 1038-1044.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction.* Cambridge, MA: The MIT Press (A Bradford Book).

Tani, J., & Nolfi, S. (1998). Learning to perceive the world as articulated. In R. Pfeifer, B. Blumberg, J. Meyer, & S. Wilson (Eds.), *Proceedings of the fifth international conference on simulation of adaptive behavior (from animals to animats)* (Vol. 5, p. 270-279). Cambridge, MA: the MIT Press.

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM*, *38*(3), 58–68.

Thorndike, E. (1911). *Animal intelligence: Experimental studies.* New York: MacMillan.

Thrun, S. (1992). The role of exploration in learning control. In *Handbook of intelligent control: Neural, fuzzy and adaptive approaches.* Florence, Kentucky: Van Nostrand Reinhold.

Tomita, M. (1982). Dynamic construction of finite-state automata from examples using hill-climbing. In *Proceedings of the fourth annual cognitive science conference* (pp. 105–108). Ann Arbor, MI.

Van de Velde, W. (1995). The biology and technology of intelligent autonomous agents. In L. Steels (Ed.), (Vol. 144, p. 197-221). Berlin: Springer.

van Orman Quine, W. (1975). *Word and object.* Cambridge, Mass.: MIT Press.

Vogt, P. (1998). Perceptual grounding in robots. In A. Birk & J. Demiris (Eds.), *Learning robots, proceedings of the ewlr-6, lecture notes on artificial intelligence 1545.* Springer.

Watkins, C. (1989). *Learning from delayed rewards.* Unpublished doctoral dissertation, King's College, Cambridge, UK.

Weiss, G. (1999). *Multiagent systems: A modern approach to distributed artificial intelligence.* Cambridge, MA: MIT Press.

Werner, G. M., & Dyer, M. G. (1991). Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial life ii* (Vol. X). Addison-Wesley.

Wheeler, M., & de Bourcier, P. (1995). How not to murder your neighbor: using synthetic behavioral ecology to study aggressive signaling. *Adaptive behavior*, *3*(3), 273-309.

Williams, R. (1990). Adaptive state representations and estimation using recurrent connectionist networks. In M. W. T, S. R. S, & W. P. J. (Eds.), *Neural networks for control* (pp. 97–114). Cambridge, MA: MIT Press.

Willshaw, D., Buneman, O., & Longuet-Higgins, H. (1969). Non-holographic associative memory. *Nature*, *222*, 960–962.

Wilson, S. W. (1990). The Animat Path to AI. In J.-A. Meyer & S. W. Wilson (Eds.), *From Animals to Animats 1. Proceedings of the First International Conference on Simulation of Adaptive Behavior (SAB90)* (pp. 15–21). Cambridge, MA: A Bradford Book. MIT Press. (http://prediction-dynamics.com/)

Wise, R., & Rompre, P.-P. (1989). Brain dopamine and reward. *Annual Review of Psychology*, *40*, 191-225.

Wolpert, D. H., & Macready, W. G. (1995). *No free lunch theorems for search* (Tech. Rep. No. SFI-TR-95-02-010). Santa Fe Institute.

Wrobel, S. (1991). Towards a model of grounded concept formation. In J. Mylopoulos & R. Reiter (Eds.), *Proceedings of the twelft international joint conference on artificial intelligence.* San Mateo, CA: Morgan Kaufmann Publishers.

Yanco, H., & Stein, L. (1993). An adaptive communication protocol for cooperating mobile robots. In H. R. Meyer, J-A. & S. Wilson (Eds.), *From animals to animats 2. proceedings of the second international conference on simulation of adaptive behavior* (p. 478-485). Cambridge, MA: The MIT Press.

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

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

Zhang, W., & Dietterich, T. G. (1996). High-performance job-shop scheduling with A time-delay TDλ network. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 1024–1030). The MIT Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185.

Index