Autonomous Concept Formation

Edwin D. de Jong

Vrije Universiteit Brussel

Artificial Intelligence Lab

Pleinlaan 2, B-1050 Brussels, Belgium

edwin@arti.vub.ac.be

http://arti.vub.ac.be

Abstract

A model for the formation of *situation concepts* is described. A characteristic of this form of concept formation is that it does not require instructive feedback. This renders it suitable for concept formation by autonomous agents. It is experimentally demonstrated that situation concepts constructed independently by several agents can convey useful information between agents through a learned system of communication. A relation was found between the development of the learned system of communication and the duration of the situations.

1 Introduction

The ability to communicate with others is a manifestation of our intelligence. An understanding of communication therefore contributes to the goals of artificial intelligence. A requirement for higher forms of communication, such as human language, is the development of concepts that are to be communicated. In this paper, a model is proposed that describes the formation of a particular type of concepts.

A *situation concept* is a part of the state of an environment that determines how the environment will present itself to a certain agent in response to (possibly absent) actions of that agent. Therefore, a situation concept is an agent specific aspect of the complete state of the environment. In many environments, the state of an environment can not be completely determined from the current sensor inputs. Thus, knowledge about the interaction history with the environment yields extra information about the current state. Situation concepts can be formed by agents by observing patterns in the sequence of inputs from the environment, actions of the agent, and subsequent evaluative feedback. They allow an agent to predict some aspect of the future, e.g. the evaluative feedback that will follow a certain action, or the next input from the environment given the action.

An important characteristic of situation concepts is that they can be developed by autonomous agents, since only evaluative feedback is assumed to be available from the environment. This distinguishes the method from traditional concept learning methods such as decision tree learning (see e.g. [Quinlan, 1990]), which require instructive feedback and thus are a form of supervised learning. [1] Although unsupervised learning methods, such as clustering, can also be viewed as methods for concept formation, these are less suitable as a basis for an agent that has to adapt itself to an environment, since feedback on how well the agent is performing can by definition not be taken into account.

It is assumed here that an autonomous agent should learn to produce successful behavior based on evaluative feedback. This feedback may be provided directly by the environment, as is usually assumed in reinforcement learning, or it may be determined by the agent itself based on its internal state and its interaction with the environment, which seems to correspond better to how humans and animals function. When such an agent adapts its behavior to an environment, its choice of actions will come to depend on its interaction history with that environment. However, the number of possible interaction histories may be large, even if histories of small length are considered, so that the exact same history is unlikely to recur frequently. In the interest of learning, it is therefore necessary to generalize.

Some generalization methods for reinforcement learning can be used as a basis for situation concept formation. A prerequisite is that the representation of the states of the environment is adapted to the learning problem, and provides a division of the interaction space into regions, such that an agent only needs to consider a region within that space, and not the specific interaction history (a point within that region), in order to decide which action to take. A good example of such an algorithm is the U-Tree method [McCallum, 1996], which uses the Kolmogorov-Smirnov test to determine whether the distribution of long-term expected rewards within a region of the state-action space, determined by features on the interaction history, varies or not. However, most generalization methods for reinforcement learning can not be used as a basis, either because the representation of the interaction space is not adapted to the learn-

---

[1] For an exposition of the difference between evaluative and instructive feedback, see e.g. [Sutton and Barto, 1998], pp. 31-33.

ing problem (e.g. plain discretization, CMAC [Albus, 1981]), or because no crisp division of the state-action space is used (e.g. neural networks).

In cognitive science literature, many models for concepts are discussed, see e.g. [Lakoff, 1987; van Orman Quine, 1975; Putnam, 1988]. When viewed as a cognitive model for a specific type of concepts, situation concepts distinguish themselves from other models by the level of detail at which they are specified. Since this level allows direct implementation, their validity can be tested in computational experiments, as testified by this paper. It should be emphasized that situation concepts are an idealized model of a particular class of concepts, and are not claimed to be a general model for the formation of concepts. Nonetheless, the possible forms of situation concepts are diverse, as may be judged from the examples in the paper.

Situation concepts are especially suited to serve as a basis for communication. In the model of language evolution investigated here, concept formation interacts with a process linking concepts to words or signals, see [Steels, 1997; Steels and Vogt, 1997]. In literature on the evolution of language, communication often corresponds directly to actions, which limits communication to the instruction of other agents. Examples include [Werner and Dyer, 1991], where sounds tell agents in a. simulation how to move to the emitter of the sound, [Yanco and Stein, 1993], where a leader robot instructs a following robot what way to move, and [Oliphant, 1997], where concepts are abstract and are modeled as having a one to one correspondance with situations. An exception is [MacLennan, 1991], where the term situation is also used to describe what the symbols that are communicated represent. The meaning of those situations is quite different from situation concepts though, since they are equal to the input of the agent, hence no concept formation is involved. In [Billard and Hayes, 1999], an interesting experiment is described where a robot develops concepts as regularities in its own behavior. This is possible because actions are selected by a process independent of concept formation, which explains why the problem mentioned above with relation to unsupervised learning plays no role here. Since the meaning of concepts in that work (objects in the environment of the robots) is fixed in advance for one of the robots (the teacher) though, it does not deal with the initial creation of the concepts, with which we will be concerned here.

Situation concepts are constructed individually by each agent. Since the concepts are based on experience with the same environment, there should be strong similarities between the conceptual systems of different agents, given that they are of the same type or species. This provides a basis for the development of communication. When agents link signals they receive and produce to their current situation, a system of communication may result where the individual situation concepts of agents are associated with shared signals. This principle is demonstrated in a simulated environment. The question that will be investigated here is whether the communication that results from this process is useful.

The structure of the paper is as follows. Section 2 formally defines situation concepts in general and describes how agents form a specific type of situation concepts in the experiments of this paper. Section 3 describes how agents adapt associations between concepts and signals in order to develop a system of communication. Section 4 explains how agents may utilize situation concepts once a system for communicating them has been learned. The setup of the experiments is described in section 5. In section 6, the benefit of communicating situation concepts is measured. Finally, section 7 presents conclusions drawn from the experiments.

2 Formation of Situation Concepts

In the most general formulation, a situation concept is a subset of the possible histories of an agent's interaction with 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. As an example, let's 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. Another example 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.

To formally describe interaction histories, time will be discretized here, which is a simplification of situation concepts in general. At time T the complete interaction history H_{max} is defined as the following set of symbols:

$$\{X_1, X_2, ... X_T, A_1, A_2, ... A_{T-1}, R_1, R_2, ... R_{T-1}\}$$

where for $0 \le t \le T$, X_t, A_t and R_t are symbols representing the input, action and reward, an evaluative feedback, at time t. Situation concepts are defined for a subset H of this complete history:

$$H \subseteq H_{max}$$

A situation concept S_p is a membership function that accepts a value for each element of an interaction history H within the corresponding domain (I^m for X_t, A^n for A_t and R for R_t) and yields a boolean value. If and only if this value is *true* at time T, the situation concept S_p applies; equivalently, it may be said in this case that the agent is in situation S_p

In the rest of the paper, the interaction history of situation concepts will consist of the current input from the environment, so that $H = \{X_T\}$. They are chosen such as to allow to predict the subsequent reward given an action the agent may choose. In the experiments, the formation of situation concepts is based on the Adaptive Subspace algorithm of [de Jong and Vogt, 1998], which recursively splits a space into two halves in a selected dimension, based on some split criterion. The split criterion here is whether the distribution of rewards over the sensor space differs between the two halves, which yields an algorithm similar in function to the previously mentioned U-tree algorithm and the continuous state generalization algorithms in [Uther and Veloso, 1998]. Initially, the complete sensor space is a single region. Regions have a one to one correspondance with situation concepts, and hence there is a single situation concept. When the distribution of rewards varies substantially within a region of the sensor space, this region is split in half, thus replacing it with two half-sized regions. This principle is applied recursively, and terminates when each situation corresponds to a region of the sensor space within which rewards are distributed homogeneously.

The result of concept formation is a tree which divides the sensor space into situations, represented by internal nodes, and for each situation contains a subtree representing possible actions in that situation, where each leaf stores an estimation of the reward following the selection of the corresponding action in that situation. An example of sue!) a tree is shown in figure 1. The actions that are distinguished depend on the situation and are constructed based on the same principle as the situations.

Action selection depends solely on the situation, not on the specific input determining the situation. The tree of situations and actions is traversed from left to right, following the conditions in the nodes that apply to the current input. This yields the current situation, represented as an internal node of the tree. The possible actions are now represented in the leaves of this node's subtree. Greedy action selection would select an action determined by the leaf with the highest estimation of the subsequent reward. For learning though, exploration is necessary. In the experiments here, the choice of explorative actions is based on the estimation error when the action was last selected and the time since it was last selected, and hence combines error based and recency based properties. For an overview of other exploration policies, see [Thrun, 1992 .

3 Development of Communication

Situation concepts organize the possible inputs an agent may receive into groups such that experience gained with a certain input influences the reward estimation of similar inputs. Apart from speeding up learning, this form of generalization provides a basis for the development of communication. Although the specific concepts agents create may differ, they result from a search for patterns
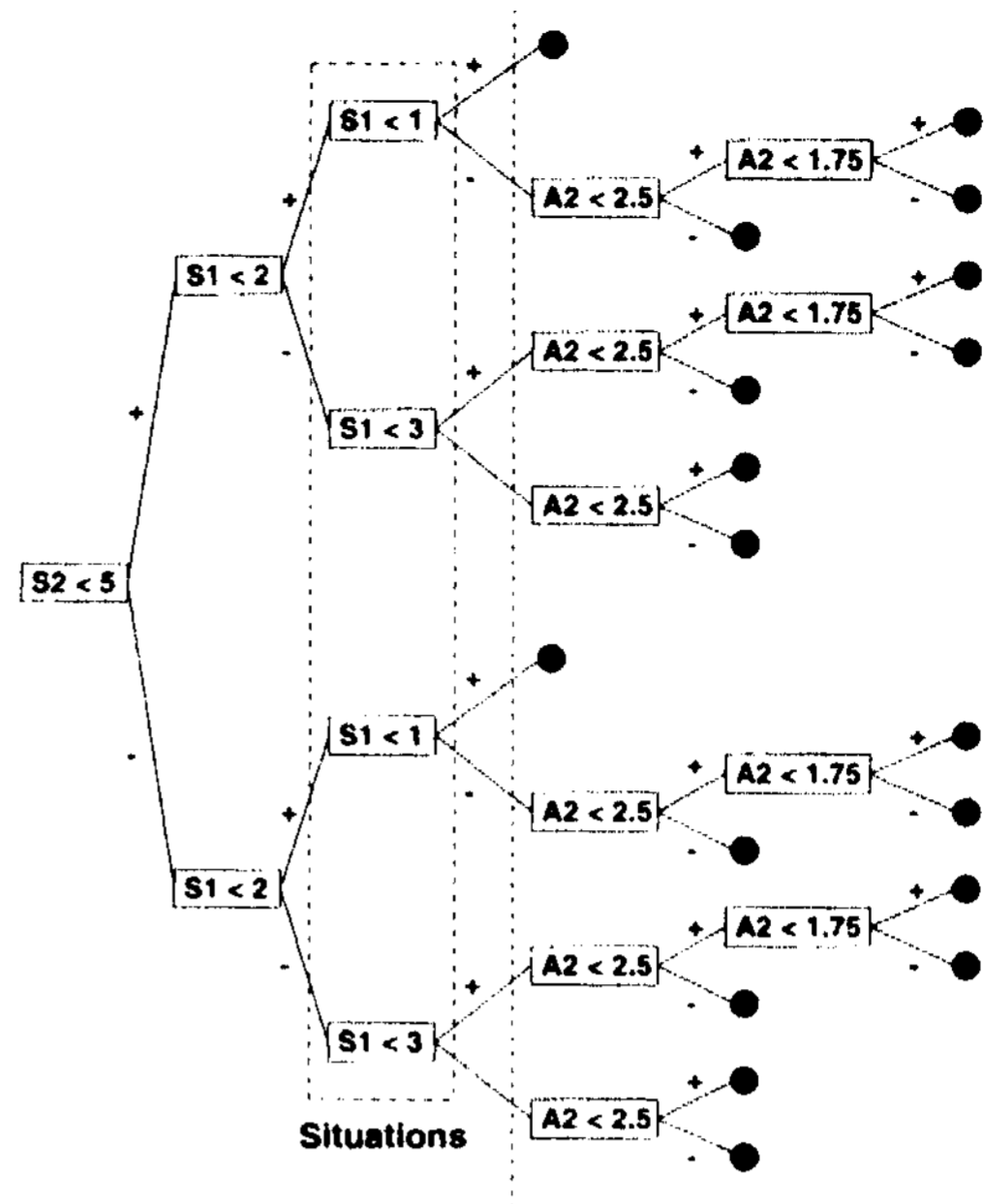

Figure 1: Example of a tree that defines situation concepts. Nodes to the left of the dotted line divide up the sensor space by constraining the range of a sensor (f and - represent the outcome of the inequality), nodes to the right divide up the action space. The dotted rectangle contains four situation concepts.

in interactions with the same environment. If a relation between the individual concepts of the agents and a shared set of signals can be found, this would allow the agents to 'speak a common language'. To this end, agents maintain a set of signals for each situation. An association between a signal and a situation has a *use* score and a *success* score. When an agent is in situation 5, it selects a signal associated with S and emits it. After every agent has produced the signal corresponding to its situation, every agent receives the collection of signals produced. Upon receiving these signals, each agent increases its use scores for the associations between these signals and the current situation. Since the conceptual systems of agents may differ, the signals associated with a certain state of the environment may differ from agent to agent.

4 Benefitting from Communication

The ability to communicate enables a capacity of gaining knowledge or information that cannot be attained through use of the ordinary perceptual devices. This advantage explains why so many animals, including humans, have retained the faculty to develop communication in the course of evolution once it evolved. An inspiring example is the alarm call system of vervet monkeys. Ingenious experiments involving playing back the calls produced by these monkeys show that these animals

have a warning system with specific calls for different kinds of predators [Seyfarth et al., 1980]. This example demonstrates the principle of how situation concepts can be useful when communicated. When a monkey did not detect the threat of an approaching predator, successful interpretation of the alarm calls produced by other community members may induce awareness 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.

It should be understood that the ultimate meaning of a situation concept is determined by its predictive value, and not by the pattern in interaction history that has been observed to be correlated with this aspect of the future. In terms of the alarm calls example, the situation corresponds to the presence of a predator, rather than to the observation of the predator by a monkey, even though situation concepts are by necessity initially constructed as correlations between interaction histories and consequences. Therefore, when a significant aspect of an agent's environment cannot be perceived directly through perception, communication may be the only way to determine the actual situation.

The benefit of communication may surface when sensor information is incomplete. The resulting uncertainty may be partially resolved by means of communication. Once a coherent mapping between situations and signals exists, the likelihood of being in a certain situation given the signals an agent receives can be determined using Hayes' formula:

$$P(\mu \mid \sigma) = \frac{P(\mu \wedge \sigma)}{P(\sigma)} = \frac{P(\mu) \cdot P(\sigma \mid \mu)}{P(\sigma)}$$

where σ is a signal that was perceived, and u is a situation. If the probability of a certain situation given the signals is high enough, the agent may decide that it is in that situation, and not in the situation indicated by its sensors (or, in general, its interaction history).

The use of Hayes' formula assumes that a coherent mapping between sit nations and signals is already available. However, since this is initially not the case, agents need to adapt their private associations between situations and signals. Depending on the process of adaptation, a coherent mapping may or may not emerge. Two sources of information are available as input to this adaptation process. Firstly, an agent may use its sensors to determine its situation and update the use scores of the associations between that situation and the signals it receives from the other agents. Secondly, the situation may be determined from the signals emitted by the other agents. To calculate the probability of being in a situation given the signals using Bayes' formula, a linear combination of the use and success scores is filled in for $P(\sigma \mid \mu)$ in the above formula; the remaining two values are obtained from counts of the occurence of situations and signals. This second source of information indicates more directly whether the link between a signal and a situation can increase the performance of the agent. Concretely, the estimated value of the action and

the actual reward following the action are compared to decide whether the determination of the situation was correct or not. If the magnitude of the difference is small, the success scores of the associations between the signals and the situation should be increased. Conversely, if the absolute difference exceeds a threshold, they should be decreased. Sending signals is not followed by evaluation, and hence does not influence scores.

Figure 2 shows an outline of the algorithm specifying when the situation is determined based on signals, and how association use and success are updated.

```
x := receive-input()
sensor-situation := determine-situation-from-sensors(x)
produce-signal(sensor-situation)
signals := receive-signals()
signal-situation := determine-situation-from-signals(signals)
if (P(signal-situation | signals) > random(1.0))
    action := choose-action(signal-situation)
    act(action)
    R := receive-reward()
    if (|R - value(action)| < threshold)
        increase-association-success(signal-situation, signals)
        for ((s in situations) and (not (s = signal-situation)))
            decrease-association-success(s, signals)
    else
        decrease-association-success(signal-situation, signals)
    update-value(action)
else
    all-signals := signals(sensor-situation)
    decrease-association-use(sensor-situation, all-signals)
    increase-association-use(sensor-situation, signals)
    action := choose-action(sensor-situation)
    act(action)
    R := receive-reward()
    update-value(action)
```

Figure 2: An outline of the algorithm. The main choice is whether the agent uses signals or sensors to determine its situation. This choice determines whether to adapt the use or the success of the association between the situation and the signals received from other agents.

5 Experimental Setup

In the experiments, five agents can move horizontally and vertically on a grid, see figure 8. Input consists of the agent's own horizontal and vertical coordinates, and an input indicating the type of a predator that is present or the absence of predators. Actions consist, of moving one step) left or right or staying, and selecting a vertical position.

A predator of random type is created in 10% of the timesteps at a random horizontal position, provided no predator is present yet. Three different types of predators exist. The vertical position of an agent determines whether it is safe from the predator or not, and since the number of vertical positions is three, each position corresponds to a single type of predator. The horizontal position of an agent determines whether it can see the
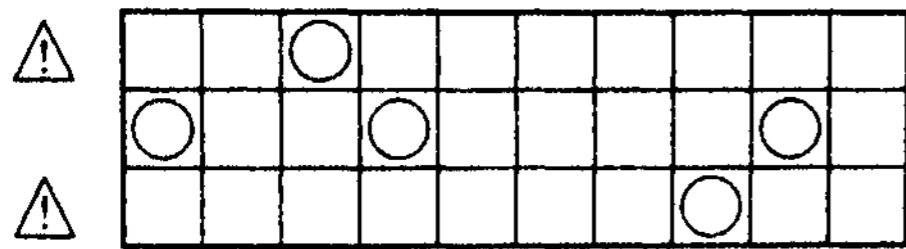
Figure 3: Visualization of the experimental environment. Horizontal positions determine the visibility of the predator, vertical positions function as abstractions of hideaways (only the middle row is safe here).
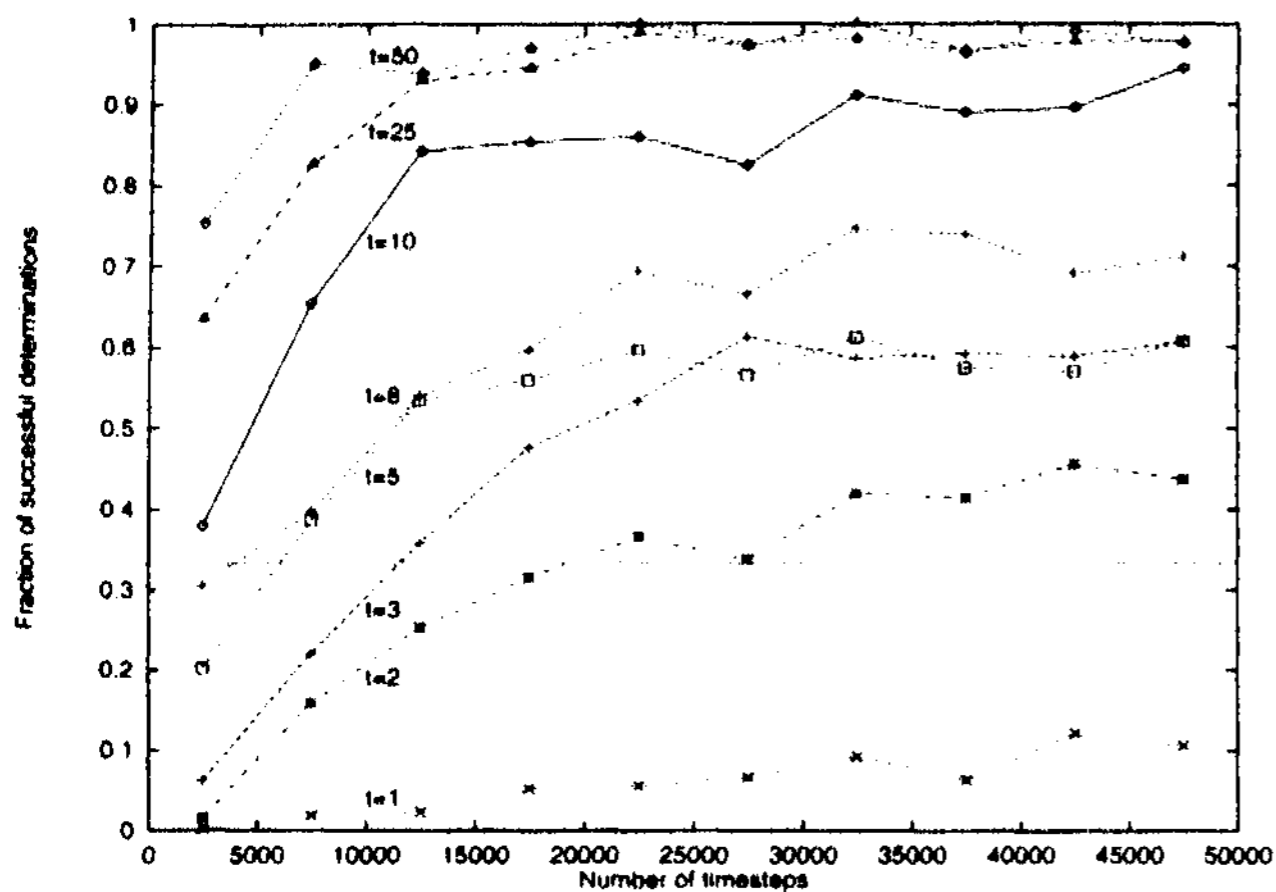

Figure 4: Histograms of the fraction of successful determinations of the situation based on communication when a predator first arrives and is invisible to the agent. Each line is the average of five repetitions of the same experiment. The interval during which a predator is present was varied between t = I and t = 50. The graph shows that when the interval is long enough, agents benefit from communication and determine the right situation in substantially more than the fraction of 1/3 (dotted line) that would be expected without communication.

predator. The scope of the agents' perception amounts to 90% of the field; hence, for each agent, 10% of the predators are expected to be invisible. When a predator is invisible to an agent at creation time it will remain so until it is removed again, i.e. until the end of the situation.

6 Measuring the Benefit of Communication

The previous section demonstrated how communication may be of benefit to an agent. Our current purpose will be to investigate whether this benefit does indeed arise. If such a benefit can be measured convincingly, this would indicate that an effective system of communication has emerged.

If the predator is invisible to an agent, there is a chance of 1 in 3 that the agent can randomly select the right vertical position. But if the agent has the right vertical position in significantly more than a third of all cases where a predator is present and invisible to the agent,

this is an indication that the agent benefits from the signals it receives from other agents. However, this way of measurement has a possible problem. If a predator arrives and is invisible to an agent, the agent will receive a low success. Since the processes of learning and adaptation are active continuously, the agent's estimation of the position's attractiveness will slowly but surely decrease. When the value of the action corresponding to the position has decreased below another action's value, and the agent will start to choose that other action, and move to another position. The speed of this process may be increased by the exploration mechanism, which will at times select the action of moving to another position upon which the value of that action increases, since its reward is higher than the original estimation. Because the number of positions is limited, the agent will eventually hit the position where it's safe from the predator.

The problem can be circumvented in the following way. At the first timestep after the creation of a predator, the only information available to an agent for which the predator is invisible consists of the signals that are produced by other agents that can sec the predator. Thus, if the agent moves to the correct position at this timestep more often than in a third of the cases, it must have extracted information from communication. Since the event that a predator is created and is invisible to an agent that uses the signals to determine its situation at that same moment is rather infrequent, the number of measurements is scarce. Therefore, this information is measured over a period of 50.000 timesteps.

There are numerous factors that influence the course of the experiment, including the increase and decrease of association strengths, removal of infrequent signals, and selection of the signal an agent produces. However, a single factor has been found to be very strongly related to the development of successful communication. This factor is the duration of a situation, and the relationship can, besides the feasibility of learned communication of situation concepts, be seen as a general result of this research.

Experiments have been performed where the duration t of the intervals during which a predator is present varied from a single timestep up to 50 timesteps. To visualise the information, a histogram has been made where the fraction of correct situation determinations is calculated over bins of 5.000 timesteps. In order to get a more reliable estimate, the experiment has been repeated a number of times for each parameter setting. Because of the duration of the experiments, the number of runs per parameter setting was limited to 5. In the graph in figure 4, each line represents the average of five runs with different random seeds for a chosen duration of situations.

The graph shows how the benefit of communication depends on the duration of the interval. When predators stay for only a single timestep, the fraction of successful determinations of the situation based on signals in case of an invisible predator stays well under one third, which means that the agent is doing worse than when it would

randomly choose its position. However, as the duration of the intervals increases, so do performance and speed of convergence. For intervals of 3 or more timesteps, the fraction of successful guesses is already higher than a third, and for intervals of 10 or more steps the agents reach perfection in determining which predator is present based on communication.

7 Conclusions

A model for the formation of a particular type of concepts, called *situation concepts,* has been described. Situation concepts capture information about how the environment of an agent will respond to its actions. They can be constructed by analysing patterns in the history of interaction between the agent and its environment. Apart from providing a particularly detailed model of concept formation, situation concepts are especially suited to serve as a basis for communication. Since agents interact with the same environment, successful communication of the situation allows agents to be better informed about their environment than their sensors alone would permit, which improves the ability to select appropriate actions.

Both the formation of situation concepts and the subsequent development of learned communication have been demonstrated in a simulation experiment. Moreover, the possible benefit of communication as an extra information source has been observed in the experiment by monitoring the actions of agents at moments when the sensors provided incomplete information. Finally, a strong relationship was found between the duration of situations and the development of successful communication.

References

[Albus, 1981] J.S. Albus. *Drains, Behavior, and Robotics.* Byte Books, Peterborough, NH, 1981.

[Billard and Hayes, 1999] Aude Billard and Gillian Hayes. Drama, a connectionist architecture for control and learning in autonomous robots. *Adaptive Behavior,* 7(1), 1999.

[de Jong and Vogt, 1998] Edwin D. de Jong and Paul Vogt. How Should a Robot Discriminate Between Objects? A comparison between two methods. In *Proceedings of the Fifth Inter-national Conference of The Society for Adaptive Behavior SAB'98,* volume 5, Cambridge, MA, 1998. The MIT Press.

[Drescher, 1991] Gary L. Drescher. *Made-up minds. A constructivist approach to artificial intelligence.* The MIT Press, Cambridge, MA, 1991.

[Lakoff, 1987] G. Lakoff. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* University of Chicago Press, Chicago, 1987.

[MacLennan, 1991] Bruce MacLennan. Synthetic ethology: An approach to the study of communication. In Chris G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II,* volume X. Addison-Wesley, 1991.

[McCallum, 1996] Andrew K. McCallum. Reinforcement learning with selective perception and hidden state. *Ph.D. Thesis,* 1996.

[Oliphant, 1997] Michael Oliphant. *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language.* PhD thesis, University of California, San Diego, CA, 1997.

[Putnam, 1988] Hilary Putnam. *Representation and reality.* The MIT Press (A Bradford Book), Cambridge, MA, 1988.

[Quinlan, 1990] J. Ross Quinlan. Induction of decision trees. In Jude W. Shavlik and Thomas G. Dietterich, editors, *Readings in Machine Learning.* Morgan Kaufmann, 1990. Originally published in *Machine Learning* 1:81 106, 1986.

[Seyfarth *et al,* 1980] R.M. Seyfarth, D.L. Cheney, and P. Marler. Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science,* 210:801-803, 1980.

[Steels and Vogt, 1997] Luc Steels and Paul Vogt. Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life,* Cambridge MA and London, 1997. The MIT Press.

[Steels, 1997] Luc Steels. The synthetic modeling of language origins. *Evolution of Communication,* 1 (1):1 — 34, 1997.

[Sutton and Bar to, 1998] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: an introduction.* The MIT Press (A Bradford Book), Cambridge, MA, 1998.

[Thrun, 1992] Sebastian Thrun. *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches,* chapter The role of exploration in learning control. Van Nostrand Reinhold, Florence, Kentucky, 1992.

[Uther and Veloso, 1998] Willam T.B. Uther and Manuela M. Veloso. Tree based discretization for continuous state space reinforcement learning. In *Proceedings of AAAI-98,* Madison, WI, 1998.

[van Orman Quine, 1975] Willard van Orman Quine. *Word and Object.* MIT Press, Cambridge, Mass., 1975.

[Werner and Dyer, 1991] Gregory M. Werner and Michael G. Dyer. Evolution of communication in artificial organisms. In Chris G. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II,* volume X. Addison-Wesley, 1991.

[Yanco and Stein, 1993] H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots. In H.L. Roitblat Meyer, J-A. and S. Wilson, editors, *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior,* pages 478-485, Cambridge, MA, 1993. The MIT Press.